Simon Schliesky

by.png

1 Main Findings

What did I do so far?
programming and data

- Custom Analysis pipelines
- Evaluation of *de novo* Assembly Software with 454 data
- Evaluation of Assembly Software with Illumina data
- Chimeric contig detection
- Comparison of public datasets from different sources by complexity reduction

metabolism

- *Megathyrsus maximus* PEP-CK type C_4
- *Megathyrsus* blueprint for engineering C_4-cycle

transport

- *Megathyrsus* intercellular transport requirement
- *Megathyrsus* modular intracellular transport machinery

1.0.1 Custom analysis pipelines

this one is very generic so I will fit it in somewhere in between

1.0.2 Evaluation of assembly softwares

For de novo assembly of 454 pyrosequencing reads we tested six different assembly algorithms with simulated reads. These reads were extracted from the *Arabidopsis* genome and were therefore considered as perfect reads. To get a more realistic picture of assembly we additionally modified the reads with 1%/3%/5% *in silico* base changes. The six assembly algorithms, mira **?**, velvet **?**, SOAP **?**, CAP3 **?**, TGICL **?**, and CLC *de novo* assembly **?**, qualitatively performed similar, with contig numbers in the 10^5s and N50s between 476 and 732.

Critical Assessment of Assembly Strategies Brautigam et al. (2011)

In this study, we tested six assembly algorithms[1] for quality in *de novo* assembly of 454 data. We could show that CAP3 and TGICL are more robust against point mutations which simulated sequencing error, as well as biological variance. Furthermore, we showed that the tested graph-based algorithms have difficulties assembling full-length transcripts, even when a high number of reads is available. In contrast, the OLC-based assemblers and the proprietary algorithm by CLCbio produced mostly full-length transcripts read number above 100.

[1]Graph-based: SOAP, Velvet, MIRA; OLC-based: CAP3, TGICL; proprietary: CLC

2 Introduction

3 Conclusion

4 Appendix

The citations here will be replaced by hard-copies of the publications in the final, non-public version of this thesis, for I do not have the rights to publish them under CC-BY 4.0

Brautigam et al. (2011)
Schulze et al. (2012)
Hamisch et al. (2012)
Schliesky et al. (2012)
Bhide et al. (2014)
Bräutigam et al. (2014)

Bibliography

Amey Bhide, Simon Schliesky, Marlis Reich, Andreas Weber, and Annette Becker. Analysis of the floral transcriptome of tarenaya hassleriana (cleomaceae), a member of the sister group to the brassicaceae: towards understanding the base of morphological diversity in brassicales. *BMC Genomics*, 15(1):140, 2014.

A. Brautigam, T. Mullick, S. Schliesky, and A. P. Weber. Critical assessment of assembly strategies for non-model species mrna-seq data and application of next-generation sequencing to the comparison of c(3) and c(4) species. *J Exp Bot*, 62(9):3093–102, 2011.

Andrea Bräutigam, Simon Schliesky, Canan Külahoglu, Colin P. Osborne, and Andreas P.M. Weber. Towards an integrative model of c4 photosynthetic subtypes: insights from comparative transcriptome analysis of nad-me, nadp-me, and pep-ck c4 species. *J Exp Bot*, 2014.

D. Hamisch, D. Randewig, S. Schliesky, A. Brautigam, A. P. Weber, R. Geffers, C. Herschbach, H. Rennenberg, R. R. Mendel, and R. Hansch. Impact of so(2) on arabidopsis thaliana transcriptome in wildtype and sulfite oxidase knockout plants analyzed by rna deep sequencing. *New Phytol*, 196(4):1074–85, 2012.

S. Schliesky, U. Gowik, A. P. M. Weber, and A. Bräutigam. Rna-seq assembly – are we there yet? *Front Pla Sci*, 3, 2012.

W. X. Schulze, K. W. Sanggaard, I. Kreuzer, A. D. Knudsen, F. Bemm, I. B. Thogersen, A. Brautigam, L. R. Thomsen, S. Schliesky, T. F. Dyrlund, M. Escalante-Perez, D. Becker, J. Schultz, H. Karring, A. Weber, P. Hojrup, R. Hedrich, and J. J. Enghild. The protein composition of the digestive fluid from the venus flytrap sheds light on prey digestion mechanisms. *Mol Cell Proteomics*, 11(11):1306–19, 2012.