Supplementary Material

CSAR-web: a web server of contig scaffolding using algebraic rearrangements

Kun-Tze Chen[1]   Chin Lung Lu[1,*]

[1]Department of Computer Science, National Tsing Hua University, Hsinchu 30013, Taiwan

1 Performance comparison of r2cat and CSAR

In our previous study (1), we have used six real datasets (five with bacterial genomes and one with human chromosome 14) to demonstrate that CSAR exhibited the best average performance in terms of many metrics, such as sensitivity, precision, F-score, genome coverage, NGA50 and running time, as compared to Projector 2, OSLay and Mauve Aligner. In this supplementary material, we used the same datasets to test r2cat and compared its average performance with that of CSAR. Consequently, on the five bacterial datasets, r2cat is superior to CSAR in terms of running time and number of scaffolds, but r2cat is inferior to CSAR in terms of sensitivity, precision, F-score, genome coverage and NGA50 (refer to Table 1). On the dataset of human chromosome 14, r2cat is superior to CSAR in terms of only number of scaffolds, but r2cat is inferior to CSAR in terms of sensitivity, precision, F-score, genome coverage, NGA50 and running time (refer to Table 2).

Table 1: Average performance of r2cat and CSAR on the five bacterial datasets

Scaffolding tool	Sen.	Prec.	F-score	Cov.	NGA50 (bp)	# Scaf.	Time (sec)
CSAR (PROmer)	**65.5**	**86.5**	**74.0**	**72.0**	**329,088**	36	26.5
CSAR (NUCmer)	54.4	82.4	63.6	61.4	266,306	53	6.9
r2cat	46.2	52.1	48.8	53.5	196,227	**18**	**3.0**

The values of sensitivity (abbreviated as 'Sen.'), precision (abbreviated as 'Prec.'), F-score and genome coverage (abbreviated as 'Cov.') are shown in percentage (%). The column '# Scaf.' gives the number of scaffolds returned by each scaffolding tool and the column 'Time' displays the running time. The best result in each column is displayed in bold.

---

*To whom correspondence should be addressed. Email: cllu@cs.nthu.edu.tw

Table 2: Average performance of r2cat and CSAR on the dataset of human chromosome 14

Scaffolding tool	Sen.	Prec.	F-score	Cov.	NGA50 (bp)	# Scaf.	Time (min)
CSAR (PROmer)	76.8	83.7	79.9	79.5	185,257	1,702	162.1
CSAR (NUCmer)	**86.1**	**94.3**	**89.8**	**89.8**	**1,086,064**	1,764	**10.0**
r2cat	42.9	42.9	42.9	45.3	39,781	**31**	1364.6

References

[1] Chen, K.-T., Liu, C.-L., Huang, S.-H., Shen, H.-T., Shieh, Y.-K., Chiu, H.-T., and Lu, C. L. (2018) CSAR: a contig scaffolding tool using algebraic rearrangements. *Bioinformatics,* **34**, 109–111.