

VirDiG: a de novo transcriptome assembler for coronavirusMinghao Li¹, Xiaoyu Guo¹, and Jin Zhao^{1*}¹ College of Computer Science and Technology, Qingdao University, China*Corresponding author. zhaojin@qdu.edu.cn**1. Detailed evaluation results**

We employed BLAT to compare the transcripts assembled by the algorithms with de novo assemblers. Detailed results are provided in Tables S1-S23

Table S1. Detailed evaluation results for SARS-CoV-1 dataset.

Accession No.	Assembler	Precision	Sensitivity	F1-scores
SRR1942954	Trinity	25.00%	18.75%	21.43%
	IDBA	0.00%	0.00%	0.00%
	SOAP	11.11%	6.25%	8.00%
	Bridger	66.67%	12.50%	21.05%
	BinPacker	42.86%	18.75%	26.09%
	TransLiG	8.33%	6.25%	7.14%
	rnaSPAdes	21.43%	18.75%	20.00%
	VirDiG	100.00%	68.75%	81.48%
SRR1942956	Trinity	16.13%	62.50%	25.64%
	IDBA	0.00%	0.00%	0.00%
	SOAP	2.63%	12.50%	4.35%
	Bridger	66.67%	12.50%	21.05%
	BinPacker	8.82%	18.75%	12.00%
	TransLiG	14.29%	18.75%	16.22%
	rnaSPAdes	15.38%	25.00%	19.05%
	VirDiG	68.42%	81.25%	74.29%
SRR1942957	Trinity	32.14%	56.25%	40.91%
	IDBA	0.00%	0.00%	0.00%
	SOAP	2.08%	6.25%	3.13%
	Bridger	25.00%	6.25%	10.00%
	BinPacker	42.86%	18.75%	26.09%
	TransLiG	18.75%	18.75%	18.75%
	rnaSPAdes	30.43%	43.75%	35.90%
	VirDiG	60.00%	75.00%	66.67%

Table S2. Detailed evaluation results for MERS-CoV dataset.

Accession No.	Assembler	Precision	Sensitivity	F1-scores
SRR10357372	Trinity	50.00%	16.67%	25.00%
	IDBA	0.00%	0.00%	0.00%
	SOAP	0.53%	8.33%	0.99%
	Bridger	100.00%	8.33%	15.38%
	BinPacker	50.00%	8.33%	14.29%
	TransLiG	6.00%	25.00%	9.68%
	rnaSPAdes	1.27%	33.33%	2.45%
	VirDiG	36.84%	58.33%	45.16%
SRR10357373	Trinity	100.00%	16.67%	28.57%
	IDBA	0.00%	0.00%	0.00%
	SOAP	0.61%	8.33%	1.13%
	Bridger	50.00%	8.33%	14.29%
	BinPacker	50.00%	8.33%	14.29%
	TransLiG	2.38%	8.33%	3.70%
	rnaSPAdes	1.94%	41.67%	3.70%
	VirDiG	46.67%	58.33%	51.85%
SRR10357374	Trinity	50.00%	16.67%	25.00%
	IDBA	0.00%	0.00%	0.00%
	SOAP	0.40%	8.33%	0.76%
	Bridger	16.67%	8.33%	11.11%
	BinPacker	50.00%	8.33%	14.29%
	TransLiG	3.03%	25.00%	5.41%
	ranSPAdes	1.24%	41.67%	2.40%
	VirDiG	36.84%	58.33%	45.16%

Table S3. Detailed evaluation results for SARS-CoV-2 dataset.

Accession No.	Assembler	Precision	Sensitivity	F1-scores
SRR12789544	Trinity	16.67%	38.46%	23.26%
	IDBA	37.50%	23.08%	28.57%
	SOAP	33.33%	7.69%	12.50%
	Bridger	100.00%	23.08%	37.50%
	BinPacker	14.29%	23.08%	17.65%
	TransLiG	33.33%	7.69%	12.50%
	rnaSPAdes	100.00%	15.38%	26.67%
	VirDiG	100.00%	69.23%	81.82%
SRR12789557	Trinity	12.96%	53.85%	20.90%
	IDBA	9.09%	7.69%	8.33%
	SOAP	25.00%	7.69%	11.76%
	Bridger	75.00%	23.08%	35.29%
	BinPacker	62.50%	38.46%	47.62%

	TransLiG	66.67%	15.38%	25.00%
	rnaSPAdes	66.67%	15.38%	25.00%
	VirDiG	100.00%	69.23%	81.82%
SRR12789558	Trinity	12.20%	38.46%	18.52%
	IDBA	16.67%	15.38%	16.00%
	SOAP	50.00%	7.69%	13.33%
	Bridger	75.00%	23.08%	35.29%
	BinPacker	54.55%	46.15%	50.00%
	TransLiG	40.00%	15.38%	22.22%
	ranSPAdes	66.67%	15.38%	25.00%
	VirDiG	100.00%	69.23%	81.82%

Table S4. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 20x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	38.46%	43.48%
	IDBA	100.00%	53.85%	70.00%
	SOAP	75.00%	23.08%	35.29%
	Bridger	25.00%	7.69%	11.76%
	BinPacker	57.14%	30.77%	40.00%
	TransLiG	33.33%	7.69%	12.50%
	rnaSPAdes	66.67%	15.38%	25.00%
	VirDiG	92.86%	100.00%	96.30%

Table S5. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 30x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	40.00%	30.77%	34.78%
	IDBA	100.00%	30.77%	47.06%
	SOAP	87.50%	53.85%	66.67%
	Bridger	40.00%	30.77%	34.78%
	BinPacker	66.67%	46.15%	54.55%
	TransLiG	50.00%	30.77%	38.10%
	rnaSPAdes	100.00%	15.38%	26.67%
	VirDiG	80.00%	92.31%	85.71%

Table S6. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 40x

Virus	Assembler	Precision	Sensitivity	F1-scores
	Trinity	37.50%	23.08%	28.57%
	IDBA	100.00%	46.15%	63.16%

SARS-CoV-2	SOAP	100.00%	7.69%	14.29%
	Bridger	60.00%	23.08%	33.33%
	BinPacker	62.50%	38.46%	47.62%
	TransLiG	80.00%	30.77%	44.44%
	rnaSPAdes	50.00%	7.69%	13.33%
	VirDiG	92.86%	100.00%	96.30%

Table S7. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 50x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	46.15%	48.00%
	IDBA	100.00%	38.46%	55.56%
	SOAP	100.00%	15.38%	26.67%
	Bridger	50.00%	23.08%	31.58%
	BinPacker	63.64%	53.85%	58.33%
	TransLiG	66.67%	15.38%	25.00%
	rnaSPAdes	66.67%	15.38%	25.00%
	VirDiG	80.00%	92.31%	85.71%

Table S8. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 60x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	46.15%	48.00%
	IDBA	100.00%	38.46%	55.56%
	SOAP	100.00%	7.69%	14.29%
	Bridger	40.00%	15.38%	22.22%
	BinPacker	57.14%	30.77%	40.00%
	TransLiG	66.67%	15.38%	25.00%
	rnaSPAdes	100.00%	15.38%	26.67%
	VirDiG	92.86%	100.00%	96.30%

Table S9. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 70x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	46.15%	48.00%
	IDBA	66.67%	15.38%	25.00%
	SOAP	100.00%	7.69%	14.29%
	Bridger	42.86%	23.08%	30.00%
	BinPacker	54.55%	46.15%	50.00%
	TransLiG	40.00%	15.38%	22.22%
	rnaSPAdes	50.00%	7.69%	13.33%

VirDiG	92.86%	100.00%	96.30%
--------	--------	----------------	---------------

Table S10. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 80x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	58.33%	53.85%	56.00%
	IDBA	100.00%	23.08%	37.50%
	SOAP	100.00%	7.69%	14.29%
	Bridger	60.00%	23.08%	33.33%
	BinPacker	62.50%	38.46%	47.62%
	TransLiG	60.00%	23.08%	33.33%
	rnaSPAdes	60.00%	23.08%	33.33%
	VirDiG	92.86%	100.00%	96.30%

Table S11. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 90x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	46.15%	48.00%
	IDBA	100.00%	30.77%	47.06%
	SOAP	100.00%	7.69%	14.29%
	Bridger	40.00%	15.38%	22.22%
	BinPacker	50.00%	23.08%	31.58%
	TransLiG	50.00%	23.08%	31.58%
	rnaSPAdes	50.00%	7.69%	13.33%
	VirDiG	92.86%	100.00%	96.30%

Table S12. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 100x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	50.00%	46.15%	48.00%
	IDBA	100.00%	15.38%	26.67%
	SOAP	100.00%	7.69%	14.29%
	Bridger	33.33%	15.38%	21.05%
	BinPacker	44.44%	30.77%	36.36%
	TransLiG	40.00%	15.38%	22.22%
	rnaSPAdes	50.00%	7.69%	13.33%
	VirDiG	92.86%	100.00%	96.30%

Table S13. Detailed evaluation results of simulated data close to the ideal sequencing depth of 20x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	75.00%	85.71%
	BinPacker	78.57%	91.67%	84.62%
	TransLiG	90.00%	75.00%	81.82%
	rnaSPAdes	100.00%	75.00%	85.71%
	VirDiG	81.82%	75.00%	78.26%

Table S14. Detailed evaluation results of simulated data close to the ideal sequencing depth of 30x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	75.00%	85.71%
	BinPacker	66.67%	83.33%	74.07%
	TransLiG	100.00%	75.00%	85.71%
	rnaSPAdes	100.00%	75.00%	85.71%
	VirDiG	100.00%	100.00%	100.00%

Table S15. Detailed evaluation results of simulated data close to the ideal sequencing depth of 40x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	66.67%	80.00%
	BinPacker	100.00%	83.33%	90.91%
	TransLiG	100.00%	66.67%	80.00%
	rnaSPAdes	100.00%	75.00%	85.71%
	VirDiG	91.67%	91.67%	91.67%

Table S16. Detailed evaluation results of simulated data close to the ideal sequencing depth of 50x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%

SARS-CoV-2	Bridger	100.00%	66.67%	80.00%
	BinPacker	78.57%	91.67%	84.62%
	TransLiG	100.00%	58.33%	73.68%
	rnaSPAdes	88.89%	66.67%	76.19%
	VirDiG	91.67%	91.67%	91.67%

Table S17. Detailed evaluation results of simulated data close to the ideal sequencing depth of 60x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	91.67%	95.65%
	BinPacker	78.57%	91.67%	84.62%
	TransLiG	100.00%	50.00%	66.67%
	rnaSPAdes	100.00%	83.33%	90.91%
	VirDiG	100.00%	100.00%	100.00%

Table S18. Detailed evaluation results of simulated data close to the ideal sequencing depth of 70x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	58.33%	73.68%
	BinPacker	85.71%	100.00%	92.31%
	TransLiG	100.00%	58.33%	73.68%
	rnaSPAdes	100.00%	83.33%	90.91%
	VirDiG	100.00%	100.00%	100.00%

Table S19. Detailed evaluation results of simulated data close to the ideal sequencing depth of 80x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	83.33%	90.91%
	BinPacker	84.62%	91.67%	88.00%
	TransLiG	90.00%	75.00%	81.82%
	rnaSPAdes	100.00%	75.00%	85.71%

VirDiG	100.00%	100.00%	100.00%
--------	---------	---------	---------

Table S20. Detailed evaluation results of simulated data close to the ideal sequencing depth of 90x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	90.00%	75.00%	81.82%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	50.00%	66.67%
	BinPacker	78.57%	91.67%	84.62%
	TransLiG	100.00%	50.00%	66.67%
	rnaSPAdes	100.00%	75.00%	85.71%
	VirDiG	100.00%	100.00%	100.00%

Table S21. Detailed evaluation results of simulated data close to the ideal sequencing depth of 100x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-2	Trinity	100.00%	91.67%	95.65%
	IDBA	100.00%	75.00%	85.71%
	SOAP	100.00%	91.67%	95.65%
	Bridger	100.00%	50.00%	66.67%
	BinPacker	92.31%	100.00%	96.00%
	TransLiG	100.00%	91.67%	95.65%
	rnaSPAdes	90.00%	75.00%	81.82%
	VirDiG	100.00%	100.00%	100.00%

Table S22. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 50x

Virus	Assembler	Precision	Sensitivity	F1-scores
SARS-CoV-1	Trinity	50.00%	37.50%	42.86%
	IDBA	100.00%	18.75%	31.58%
	SOAP	100.00%	18.75%	31.58%
	Bridger	60.00%	18.75%	28.57%
	BinPacker	44.44%	25.00%	32.00%
	TransLiG	80.00%	25.00%	38.10%
	rnaSPAdes	100.00%	12.50%	22.22%
	VirDiG	100.00%	93.75%	96.77%

Table S23. Detailed evaluation results of simulated data close to the real situation with sequencing depth of 50x

Virus	Assembler	Precision	Sensitivity	F1-scores
MERS-CoV	Trinity	60.00%	25.00%	35.29%
	IDBA	100.00%	16.67%	28.57%
	SOAP	50.00%	8.33%	14.29%
	Bridger	80.00%	33.33%	47.06%
	BinPacker	71.43%	41.67%	52.63%
	TransLiG	50.00%	8.33%	14.29%
	rnaSPAdes	50.00%	8.33%	14.29%
	VirDiG	100.00%	91.67%	95.65%

2. Command Lines used for benchmarking

For benchmarking, we used several existing tools, which are listed below. All tools were tested with their default settings unless stated otherwise.

- **BLAT:** version 2.14.1

```
blastn -query transcript.fasta -db reftranscript -out <output> -evaluate 1e-20 -dust no
-task megablast -num_threads 2 -outfmt 6 -max_target_seqs 3
```

- **Trinity:** version 2.15.1

```
Trinity --seqType fq --left reads1.fastq --right reads2.fastq --CPU 1 --max_memory
40G --output <output>
```

- **Bridger:** version 2014-12-01

```
Bridger.pl --seqType fq --left reads1.fastq --right reads2.fastq --CPU 6 -o <output>
```

- **TransLiG:** version 1.3

```
TransLiG -s fq -p pair -l reads1.fastq -r reads2.fastq -o <output>
```

- **rnaSPAdes:** version 3.15.4

```
rnaspades.py -1 reads1.fastq -2 reads2.fastq -o <output>
```

- **BinPacker:** version 1.0

```
BinPacker -s fq -p pair -l reads1.fastq -r reads2.fastq -o <output> -m FR
```

- **IDBA-tran:** version 1.1.3

`idba_tran -r read.fa -o <output_dir>`

- **SOAPdenovo-trans :** version 1.0.5

`SOAPdenovo-Trans all -s config_file -o <output>`