

Supplementary Information

# zDUR: Reference-free FASTQ compressor with high compression ratio and speed

Artem Ershov<sup>1</sup>, Renpeng Ding<sup>1</sup>, Qian Fu<sup>1</sup>, Ivan Kozlov<sup>1</sup>, Ekaterina Fadeeva<sup>1</sup>, Evgeniy Mozheiko<sup>1</sup>, Ming Ni<sup>1</sup>, Yong Hou<sup>1</sup>, and Yan Zhou<sup>\*2</sup>

<sup>1</sup>MGI Tech, BGI, Shenzhen 518083, China

<sup>2</sup>State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China

---

\*Corresponding author

# Benchmark data

Table S1 summarizes the datasets used in the benchmark.

Table S1: Description of benchmark datasets. Sizes are reported in gigabytes.

| Dataset                     | Size  | Technology  | Origin      | Instrument  | Quality levels |
|-----------------------------|-------|-------------|-------------|-------------|----------------|
| NovaSeq_HG002_30x           | 206.8 | WGS         | human       | NovaSeq     | Q4             |
| NovaSeq_HG003_30x           | 206.8 |             | human       | NovaSeq     | Q4             |
| NovaSeq_HG004_30x           | 206.8 |             | human       | NovaSeq     | Q4             |
| ERR5654413                  | 27.6  | Metagenome  | human gut   | NovaSeq     | Q4             |
| ERR1539426                  | 17.1  |             | human gut   | HiSeq       | Q40            |
| ERR589597                   | 17.1  |             | human mouth | HiSeq       | Q40            |
| SRR15463463                 | 25.8  |             | human mouth | NovaSeq     | Q4             |
| SRR17548781                 | 23.1  |             | human skin  | NovaSeq     | Q4             |
| SRR3187092                  | 19.0  |             | human skin  | NovaSeq     | Q40            |
| V350217806.L01_2            | 95.0  | Single-Cell | mouse       | MGISEQ-2000 | Q40            |
| V300067629.L03              | 130.4 |             | zebrafish   | MGISEQ-2000 | Q40            |
| V300076043.L01              | 126.7 |             | zebrafish   | MGISEQ-2000 | Q40            |
| V350044419.KF21-22.L01.read | 49.5  | Stereo-seq  | mouse       | MGISEQ-2000 | Q40            |
| E100018389.L01.17           | 48.9  |             | zebrafish   | DNBSEQ-T7   | Q40            |
| DP8400019516TL.L01.49       | 47.4  |             | zebrafish   | DNBSEQ-T1   | Q40            |

## Download Links

- **NovaSeq\_HG002\_30x<sup>1</sup>**
  - <https://cgl.gi.ucsc.edu/data/giraffe/mapping/reads/real/HG002/HG002.NovaSeq.pcr-free.35x.R1.fastq.gz>
  - <https://cgl.gi.ucsc.edu/data/giraffe/mapping/reads/real/HG002/HG002.NovaSeq.pcr-free.35x.R2.fastq.gz>
- **NovaSeq\_HG003\_30x<sup>1</sup>**
  - <https://cgl.gi.ucsc.edu/data/giraffe/mapping/reads/real/HG003/HG003.NovaSeq.pcr-free.35x.R1.fastq.gz>
  - <https://cgl.gi.ucsc.edu/data/giraffe/mapping/reads/real/HG003/HG003.NovaSeq.pcr-free.35x.R2.fastq.gz>
- **NovaSeq\_HG004\_30x<sup>1</sup>**
  - <ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR147/019/SRR14724519>
- **ERR5654413**
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR565/003/ERR5654413/ERR5654413\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR565/003/ERR5654413/ERR5654413_1.fastq.gz)
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR565/003/ERR5654413/ERR5654413\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR565/003/ERR5654413/ERR5654413_2.fastq.gz)
- **ERR1539426**
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR153/006/ERR1539426/ERR1539426\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR153/006/ERR1539426/ERR1539426_1.fastq.gz)
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR153/006/ERR1539426/ERR1539426\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR153/006/ERR1539426/ERR1539426_2.fastq.gz)
- **ERR589597**
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR589/ERR589597/ERR589597\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR589/ERR589597/ERR589597_1.fastq.gz)
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR589/ERR589597/ERR589597\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR589/ERR589597/ERR589597_2.fastq.gz)
- **SRR15463463**

<sup>1</sup>Files from these datasets were downsampled with `seqtk sample -s1000 ORIGINAL.fq 300000000 > DOWNSAMPLED.fq`

- [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR154/063/SRR15463463/SRR15463463\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR154/063/SRR15463463/SRR15463463_1.fastq.gz)
- [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR154/063/SRR15463463/SRR15463463\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR154/063/SRR15463463/SRR15463463_2.fastq.gz)
- **SRR17548781**
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR175/081/SRR17548781/SRR17548781\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR175/081/SRR17548781/SRR17548781_1.fastq.gz)
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR175/081/SRR17548781/SRR17548781\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR175/081/SRR17548781/SRR17548781_2.fastq.gz)
- **SRR3187092**
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR318/002/SRR3187092/SRR3187092\\_1.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR318/002/SRR3187092/SRR3187092_1.fastq.gz)
  - [ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR318/002/SRR3187092/SRR3187092\\_2.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR318/002/SRR3187092/SRR3187092_2.fastq.gz)
- **V350044419\_KF21-22\_L01\_read**
  - <https://db.cngb.org/search/project/CNP0001543/>
- **E100018389\_L01\_17**
  - [https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461254/CNX0378098/CNR0459120/E100018389\\_L01\\_17\\_1.fq.gz](https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461254/CNX0378098/CNR0459120/E100018389_L01_17_1.fq.gz)
  - [https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461254/CNX0378098/CNR0459120/E100018389\\_L01\\_17\\_2.fq.gz](https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461254/CNX0378098/CNR0459120/E100018389_L01_17_2.fq.gz)
- **DP8400019516TL\_L01\_49**
  - [https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461255/CNX0378106/CNR0459128/DP8400019516TL\\_L01\\_49\\_1.fq.gz](https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461255/CNX0378106/CNR0459128/DP8400019516TL_L01_49_1.fq.gz)
  - [https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461255/CNX0378106/CNR0459128/DP8400019516TL\\_L01\\_49\\_2.fq.gz](https://ftp.cngb.org/pub/CNSA/data3/CNP0002220/CNS0461255/CNX0378106/CNR0459128/DP8400019516TL_L01_49_2.fq.gz)
- **V350217806\_L01\_2**
  - [https://zenodo.org/records/15149973/files/V350217806\\_L01\\_2\\_1.fq.gz?download=1&preview=1](https://zenodo.org/records/15149973/files/V350217806_L01_2_1.fq.gz?download=1&preview=1)
  - [https://zenodo.org/records/15149973/files/V350217806\\_L01\\_2\\_2.fq.gz?download=1&preview=1](https://zenodo.org/records/15149973/files/V350217806_L01_2_2.fq.gz?download=1&preview=1)
- **V300067629\_L03**
  - [https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521540/CNX0425253/CNR0509168/V300067629\\_L03\\_read\\_1.fq.gz](https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521540/CNX0425253/CNR0509168/V300067629_L03_read_1.fq.gz)
  - [https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521540/CNX0425253/CNR0509168/V300067629\\_L03\\_read\\_2.fq.gz](https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521540/CNX0425253/CNR0509168/V300067629_L03_read_2.fq.gz)
- **V300076043\_L01**
  - [https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521535/CNX0425248/CNR0509163/V300076043\\_L01\\_read\\_1.fq.gz](https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521535/CNX0425248/CNR0509163/V300076043_L01_read_1.fq.gz)
  - [https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521535/CNX0425248/CNR0509163/V300076043\\_L01\\_read\\_2.fq.gz](https://ftp.cngb.org/pub/CNSA/data2/CNP0002712/CNS0521535/CNX0425248/CNR0509163/V300076043_L01_read_2.fq.gz)

## 2 Complete benchmark results

Tables S2, S3, and S4 present compression ratios, as well as compression and decompression times, for all tested tools except FaStore, which is discussed separately in Section 3. Tables S5 and S6 present the memory consumption of compression and decompression for all tested tools except FaStore.

All tools were run with 32 threads, using settings for lossless compression and verbose output, which are provided in Section 5. Leon failed to process several files, and in some cases, its decompressed files were much smaller than the originals. We reported this issue on Leon’s Github page.

Execution times were measured using GNU time.

Table S2: Compression ratios for various sequencing datasets, expressed as a fraction of the original file size. A dash (-) indicates that the tool could not successfully compress the dataset. Sizes are reported in gigabytes.

| Dataset                 | Technology  | Size  | Fraction of the input file |        |          |        |        |        |        |
|-------------------------|-------------|-------|----------------------------|--------|----------|--------|--------|--------|--------|
|                         |             |       | pigz                       | DSRC   | fqzcomp5 | Leon   | repaq  | SPRING | zDUR   |
| NovaSeq_HG002.30x_sub1  | WGS         | 206.8 | 19.3%                      | 14.01% | 11.3%    | -      | 11.01% | 4.02%  | 6.81%  |
| NovaSeq_HG003.30x_sub1  |             | 206.8 | 19.37%                     | 14.06% | 11.35%   | -      | 11.1%  | 4.07%  | 6.88%  |
| NovaSeq_HG004.30x_sub1  |             | 206.8 | 19.26%                     | 13.97% | 11.29%   | -      | 10.99% | 3.96%  | 6.73%  |
| ERR1539426              | Metagenome  | 17.1  | 36.54%                     | 26.38% | 24.38%   | 24.42% | 31.15% | 17.26% | 19.56% |
| ERR5654413              |             | 27.6  | 21.13%                     | 14.02% | 13.21%   | 9.29%  | 12.59% | 4.9%   | 6.75%  |
| ERR589597               |             | 17.1  | 35.05%                     | 25.25% | 22.91%   | 22.41% | 29.37% | 15.7%  | 17.16% |
| SRR15463463             |             | 25.8  | 18.74%                     | 12.9%  | 9.59%    | 7.45%  | 9.3%   | 5.08%  | 6.26%  |
| SRR17548781             |             | 23.1  | 18.73%                     | 16.47% | 10.33%   | 15.37% | 11.42% | 10.24% | 11.1%  |
| SRR3187092              |             | 19.0  | 38.67%                     | 28.36% | 24.59%   | 30.3%  | 33.09% | 23.68% | 24.63% |
| V300067629.L03          | Single-Cell | 130.4 | 33.36%                     | 27.02% | 17.44%   | -      | 20.89% | 18.03% | 16.53% |
| V300076043.L01          |             | 126.7 | 33.43%                     | 27.03% | 18.3%    | -      | 21.65% | 18.11% | 16.6%  |
| V350217806.L01.2.1      |             | 95.0  | 35.21%                     | 28.22% | 21.38%   | -      | 27.46% | 20.76% | 17.09% |
| DP8400019516TL.L01.49.1 | Stereo-seq  | 47.4  | 28.44%                     | 21.85% | 15.94%   | -      | 20.62% | 18.06% | 16.23% |
| E100018389.L01.17.1     |             | 48.9  | 26.03%                     | 19.17% | 13.56%   | -      | 15.95% | 15.4%  | 14.06% |
| V350044419_KF21-22.L01  |             | 49.5  | 31.24%                     | 24.93% | 19.06%   | 23.61% | 22.95% | 20.0%  | 18.77% |

Table S3: Compression times for various sequencing datasets. A dash (-) indicates that the tool could not successfully compress the dataset. Sizes are reported in gigabytes.

| Dataset                 | Technology  | Size  | Compression time |        |          |         |          |            |         |
|-------------------------|-------------|-------|------------------|--------|----------|---------|----------|------------|---------|
|                         |             |       | pigz             | DSRC   | fqzcomp5 | Leon    | repaq    | SPRING     | zDUR    |
| NovaSeq_HG002.30x_sub1  | WGS         | 206.8 | 9m 36s           | 2m 9s  | 6m 53s   | -       | 44m 28s  | 1h 31m 25s | 6m 41s  |
| NovaSeq_HG003.30x_sub1  |             | 206.8 | 9m 36s           | 2m 9s  | 7m 5s    | -       | 43m 58s  | 1h 39m 52s | 10m 45s |
| NovaSeq_HG004.30x_sub1  |             | 206.8 | 9m 55s           | 2m 4s  | 7m 53s   | -       | 49m 1s   | 1h 35m 21s | 6m 57s  |
| ERR1539426              | Metagenome  | 17.1  | 1m 4s            | 18s    | 2m 40s   | 6m 34s  | 12m 6s   | 9m 12s     | 1m 47s  |
| ERR5654413              |             | 27.6  | 1m 26s           | 19s    | 2m 10s   | 11m 31s | 5m 48s   | 10m 22s    | 2m 3s   |
| ERR589597               |             | 17.1  | 1m 1s            | 19s    | 2m 37s   | 6m 29s  | 11m 42s  | 8m 25s     | 1m 40s  |
| SRR15463463             |             | 25.8  | 1m 6s            | 17s    | 2m 13s   | 8m 25s  | 5m 55s   | 25m 9s     | 2m 4s   |
| SRR17548781             |             | 23.1  | 1m 4s            | 18s    | 2m 22s   | 13m 19s | 8m 44s   | 37m 15s    | 2m 49s  |
| SRR3187092              |             | 19.0  | 1m 10s           | 21s    | 2m 32s   | 11m 48s | 11m 35s  | 16m 54s    | 2m 11s  |
| V300067629.L03          | Single-Cell | 130.4 | 7m 3s            | 1m 49s | 6m 34s   | -       | 52m 53s  | 1h 21m 14s | 7m 42s  |
| V300076043.L01          |             | 126.7 | 7m 1s            | 1m 46s | 5m 59s   | -       | 52m 18s  | 1h 19m 33s | 7m 41s  |
| V350217806.L01.2.1      |             | 95.0  | 5m 8s            | 1m 12s | 4m 47s   | -       | 1h 7m 8s | 1h 23m 45s | 5m 9s   |
| DP8400019516TL.L01.49.1 | Stereo-seq  | 47.4  | 2m 11s           | 42s    | 4m 26s   | -       | 27m 38s  | 1h 47m 49s | 2m 8s   |
| E100018389.L01.17.1     |             | 48.9  | 2m 18s           | 39s    | 4m 27s   | -       | 20m 34s  | 1h 43m 8s  | 2m 4s   |
| V350044419_KF21-22.L01  |             | 49.5  | 2m 30s           | 45s    | 4m 28s   | 23m 1s  | 22m 20s  | 39m 31s    | 2m 18s  |

Table S4: Decompression times for various sequencing datasets. A dash (-) indicates that the tool could not successfully compress the dataset. Sizes are reported in gigabytes.

| Dataset                | Technology  | Size  | Decompression time |        |          |        |            |         |        |
|------------------------|-------------|-------|--------------------|--------|----------|--------|------------|---------|--------|
|                        |             |       | pigz               | DSRC   | fqzcomp5 | Leon   | repaq      | SPRING  | zDUR   |
| NovaSeq_HG002.30x_sub1 | WGS         | 206.8 | 13m 18s            | 4m 13s | 7m 37s   | -      | 1h 50m 44s | 23m 1s  | 8m 54s |
| NovaSeq_HG003.30x_sub1 |             | 206.8 | 13m 18s            | 4m 17s | 7m 18s   | -      | 1h 50m 57s | 24m 19s | 8m 59s |
| NovaSeq_HG004.30x_sub1 |             | 206.8 | 11m 11s            | 4m 7s  | 8m 42s   | -      | 1h 53m 44s | 23m 56s | 8m 34s |
| ERR1539426.1           | Metagenome  | 17.1  | 1m 35s             | 28s    | 2m 19s   | 3m 25s | 17m 20s    | 3m 11s  | 40s    |
| ERR5654413.1           |             | 27.6  | 1m 59s             | 46s    | 2m 15s   | 4m 25s | 12m 49s    | 3m 29s  | 57s    |
| ERR589597.1            |             | 17.1  | 1m 41s             | 28s    | 2m 24s   | 3m 16s | 16m 1s     | 3m 0s   | 40s    |
| SRR15463463.1          |             | 25.8  | 1m 29s             | 40s    | 2m 17s   | 3m 51s | 10m 26s    | 3m 37s  | 54s    |
| SRR17548781.1          |             | 23.1  | 1m 26s             | 39s    | 1m 59s   | 4m 7s  | 10m 43s    | 4m 10s  | 56s    |
| SRR3187092.1           |             | 19.0  | 1m 59s             | 32s    | 1m 52s   | 3m 47s | 19m 45s    | 4m 3s   | 46s    |
| V300067629.L03         | Single-Cell | 130.4 | 11m 54s            | 2m 58s | 8m 7s    | -      | 1h 46m 49s | 22m 35s | 5m 45s |
| V300076043.L01         |             | 126.7 | 11m 44s            | 2m 43s | 8m 2s    | -      | 1h 44m 21s | 21m 51s | 5m 27s |
| V350217806.L01.2.1     |             | 95.0  | 8m 35s             | 2m 15s | 6m 24s   | -      | 1h 34m 8s  | 17m 10s | 3m 52s |
| DP8400019516TLL01.49.1 | Stereo-seq  | 47.4  | 3m 24s             | 1m 9s  | 5m 6s    | -      | 44m 4s     | 10m 27s | 1m 46s |
| E100018389.L01.17.1    |             | 48.9  | 4m 2s              | 1m 12s | 5m 12s   | -      | 41m 55s    | 11m 27s | 1m 58s |
| V350044419.KF21-22.L01 |             | 49.5  | 4m 5s              | 1m 11s | 5m 23s   | -      | 44m 9s     | 9m 17s  | 1m 58s |

Table S5: Compression memory consumption for selected sequencing datasets. File size and memory usage is reported in gigabytes. A value of "<1" indicates memory consumption less than 1 GB.

| Dataset                | Technology | Size  | Compression memory (GB) |      |          |       |        |      |
|------------------------|------------|-------|-------------------------|------|----------|-------|--------|------|
|                        |            |       | pigz                    | DSRC | fqzcomp5 | repaq | SPRING | zDUR |
| NovaSeq_HG002.30x_sub1 | WGS        | 206.8 | <1                      | <1   | 22       | 15    | 33     | 113  |
| SRR15463463            | Metagenome | 25.8  | <1                      | <1   | 17       | 14    | 7      | 23   |

Table S6: Decompression memory consumption for selected sequencing datasets. Memory usage is reported in gigabytes. A value of "¡1" indicates memory consumption less than 1 GB.

| Dataset                | Technology | Size  | Decompression memory (GB) |      |          |       |        |      |
|------------------------|------------|-------|---------------------------|------|----------|-------|--------|------|
|                        |            |       | pigz                      | DSRC | fqzcomp5 | repaq | SPRING | zDUR |
| NovaSeq_HG002.30x_sub1 | WGS        | 206.8 | <1                        | <1   | 20       | 2     | 16     | 65   |
| SRR15463463            | Metagenome | 25.8  | <1                        | <1   | 11       | ¡1    | 7      | 11   |

### 3 Comparison of zDUR with FaStore

Out of all tested datasets, FaStore successfully processed only SRR3187092, despite several datasets having a constant read length (e.g., all reads in both files of SRR15463463 are 151 bases long).

Nonetheless, zDUR significantly outperforms FaStore (with the `--fast` option) on SRR3187092, achieving  $9.4\times$  faster compression and  $6.8\times$  faster decompression (see Table S7).

Moreover, despite omitting the contig assembly step from the original algorithm, zDUR achieves better compression due to more precise quality modeling, improved header handling, and the use of libbsc as a general-purpose codec.

Table S7: Comparison of zDUR and FaStore performance on SRR3187092 dataset.

| Tool    | Fraction of the input file | Compression time | Decompression time |
|---------|----------------------------|------------------|--------------------|
| FaStore | 25.43%                     | 20m 32s          | 5m 14s             |
| zDUR    | 24.63%                     | 2m 11s           | 46s                |

## 4 Ablation study

Tables S8 and S9 compare the performance of three zDUR compression modes: c-simtree, c-stereoseq, and c-fast.

Table S8: Compression ratio and runtime comparison of the three zDUR modes.

| Dataset          | Technology | Size | Fraction of the input file |             |        | Compression time (s) |             |        |
|------------------|------------|------|----------------------------|-------------|--------|----------------------|-------------|--------|
|                  |            |      | c-simtree                  | c-stereoseq | c-fast | c-simtree            | c-stereoseq | c-fast |
| ERR5654413       | Metagenome | 27.6 | 6.75%                      | /           | 8.62%  | 130                  | /           | 84     |
| SRR15463463      | Metagenome | 25.8 | 6.26%                      | /           | 10.90% | 139                  | /           | 82     |
| V350217806_L01.2 | scRNA      | 95.0 | 17.67%                     | 17.09%      | 20.04% | 538                  | 500         | 186    |

Table S9: Sequence compression ratio comparison of the three zDUR modes.

| Dataset          | Technology | Size | Fraction of the input file |             |        |
|------------------|------------|------|----------------------------|-------------|--------|
|                  |            |      | c-simtree                  | c-stereoseq | c-fast |
| ERR5654413       | Metagenome | 11.3 | 7.63%                      | /           | 24.87% |
| SRR15463463      | Metagenome | 10.6 | 8.33%                      | /           | 22.12% |
| V350217806_L01.2 | scRNA      | 37.1 | 8.55%                      | 7.24%       | 17.20% |

## 5 Tools options

All tools were run via a wrapper pipeline, available separately at <https://github.com/MGI-EU/fastq-bench-mark-pipeline>. The exact commands are listed below. INPUT1 and INPUT2 represent absolute paths of the two input files from each dataset.

```
# DSRC:
dsr c -v -t32 INPUT1 INPUT1.dsrc
dsr c -v -t32 INPUT2 INPUT2.dsrc
dsr d -v -t32 INPUT1.dsrc INPUT1.dsrc.decomp
dsr d -v -t32 INPUT2.dsrc INPUT2.dsrc.decomp

# FaStore:
# --fast was used since zDUR does not implement reclustering
# due to an unfavorable compression ratio vs. performance tradeoff
fastore_compress.sh --fast --lossless --in INPUT1 --pair INPUT2 --out INPUT1.fastore \
    --threads 32 --verbose
fastore_decompress.sh --in INPUT1.fastore --out INPUT1.fastore.decomp1 \
    --pair INPUT1.fastore.decomp2 --threads 32 --verbose

# Leon:
leon -file INPUT1 -c -lossless -nb-cores 32 -verbose 1
leon -file INPUT2 -c -lossless -nb-cores 32 -verbose 1
leon -file INPUT1.leon -d -nb-cores 32 -verbose 1
leon -file INPUT2.leon -d -nb-cores 32 -verbose 1

# SPRING:
# -r was used for comparison, because FaStore and zDUR always reorder reads;
# in our experience, -r does not affect speed and only provides better compression
spring -c -r -i INPUT1 INPUT2 -t32 -o INPUT1.spring
spring -d -i INPUT1.spring -o INPUT1.spring.decomp1 INPUT1.spring.decomp2 -t32

# fqzcomp5:
fqzcomp5 -t32 -v INPUT1 INPUT1.fqz5
fqzcomp5 -t32 -v INPUT2 INPUT2.fqz5
fqzcomp5 -d -t32 -v INPUT1.fqz5 INPUT1.fqz5.decomp
```

```
fqzcomp5 -d -t32 -v INPUT2.fqz5 INPUT2.fqz5.decomp
```

```
# pigz:
```

```
pigz --keep -f INPUT1 -p 32
```

```
pigz --keep -f INPUT2 -p 32
```

```
pigz -d --keep -f INPUT1_tmp.gz -p 32
```

```
pigz -d --keep -f INPUT2_tmp.gz -p 32
```

```
# repaq:
```

```
repag -c --in1 INPUT1 --in2 INPUT2 -o INPUT1.rqf.xz -t 32
```

```
repag -d -i INPUT1.rqf.xz --out1 INPUT1.decomp1 --out2 INPUT2.decomp2 -t 32
```

```
# zDUR (WGS and Metagenome datasets):
```

```
zDUR c-simtree --i1 INPUT1 --i2 INPUT2 -o INPUT1.zdur --force --threads 32
```

```
zDUR d -i INPUT1.zdur --o1 INPUT1.zdur.decomp1 --o2 INPUT1.zdur.decomp2 --force --threads 32
```

```
# zDUR (Single-Cell and Stereo-seq datasets):
```

```
zDUR c-stereoseq --i1 INPUT1 --i2 INPUT2 -o INPUT1.zdur --force --threads 32
```

```
zDUR d -i INPUT1.zdur --o1 INPUT1.zdur.decomp1 --o2 INPUT1.zdur.decomp2 --force --threads 32
```