Dereplication options lotus/sdm

General info

- Dereplication refers to finding exact 100% identity matches of amplicon reads; these are subsequently collapsed to save memory and computational effort for subsequent programs

- One important application of dereplicated reads is to filter unique reads, reads that have no other matches, out of the dataset, as they are likely to represent sequencing errors

- With LotuS the type of unique can be very specifically set. Use the –derepMin flag or for sdm directly the -min_derep_copies flag

Derep_min 5 (5 times total in any number of samples)

Accepted:

- 5x in S1

- 2x in S1 and 3x in S4

- 1000x in S1

Rejected:

- 2x in S1 and 2x in S4

- 1x in S1 and 1x in S2 and 1x in S3

Derep_min 3:2 (3 times total in at least 2 samples)

Accepted:

- 2x in S1 and 3x in S4
- 2x in S1 and 2x in S4
- 1000x in S1

Rejected:

- 5x in S1
- 1x in S1 and 1x in S2 and 1x in S3

Derep_min 3:2,5:1 (3 times total in >= 2 samples or 5 times in >= 1 samples)

Accepted:

- 2x in S1 and 3x in S4
- 2x in S1 and 2x in S4
- 5x in S1
- 1000x in S1

Rejected:

- 1x in S1 and 1x in S2 and 1x in S3

Derep_min 3:3,5:2,10:1 (>=3 times total in >= 3 samples or >=5 times total in >= 3 samples or >=10 times in >= 1 samples)

Accepted:

- 2x in S1 and 3x in S4
- 1x in S1 and 1x in S2 and 1x in S3
- 5x in S1
- 1000x in S1

Rejected:

- 2x in S1 and 2x in S4

Derep_min 3:3,5:2 (>=3 times total in >= 3 samples or >=5 times total in >= 2 samples)

Accepted:
- 2x in S1 and 3x in S4
- 1x in S1 and 1x in S2 and 1x in S3

Rejected:
- 2x in S1 and 2x in S4
- 1000x in S1
- 5x in S1

Derep_min 3:3,5:2,10:1 & specific mapping column

Will override "10:1" term to specific singletons