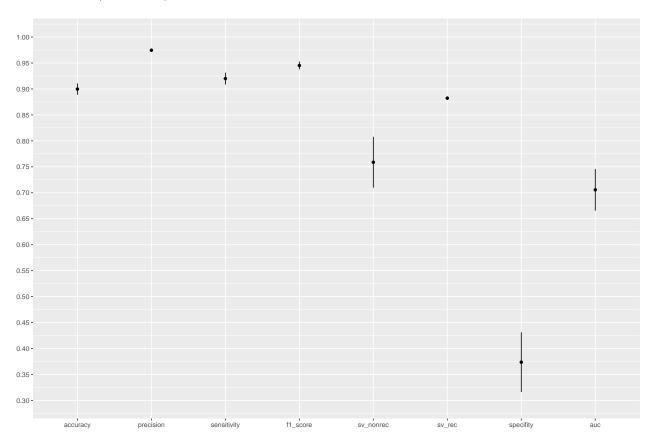
simulation results analysis for mix recombinants

Qian Feng 22/02/2020

This file is for analyzing the simulation results for mix recombinants. The combination of recombinants per replicate was sampled from Ghana real data.

Classification measures: accuracy, precision, sensitivity, specifity, $f1_score$, sv_nonrec , sv_rec , auc



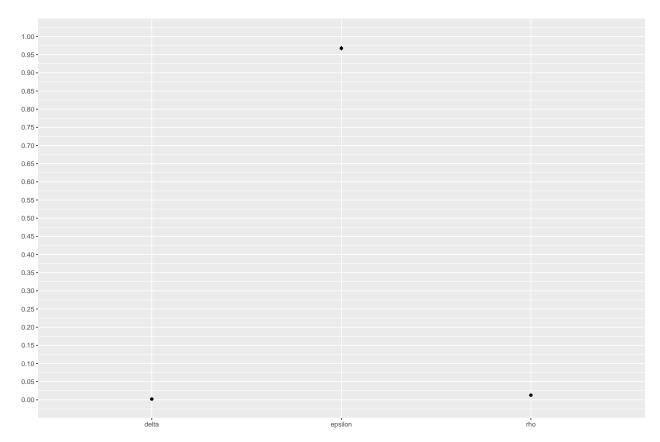
Mosaic three pars

```
## mosaic_pars value CI.L CI.R

## delta 1 0.0021 0.001904004 0.002295996

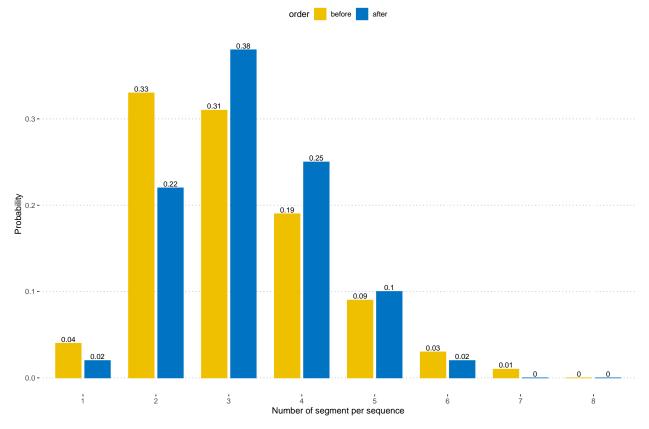
## epsilon 2 0.9677 0.961898507 0.973501493

## rho 3 0.0129 0.012723603 0.013076397
```



Let's see the segment count distribuition from the simulated datasets.

Yellow bars show the distribuition about segment count per sequence only from simulated dataset, without running mosaic step. Blue bars illustrate the distribuition about segment count per target sequence after running mosaic step.



Above original distribuition before mosaic step seems to move towards to right. Conclusion is when I generate recombinants from real data mosaic output, the simulated dataset tend to generate more segments per target sequence after mosaic implementation.

I think we are trying to solve the low specificity issue. How to control the number of recombinants in multiple_source partial alignment case?

