Prepare WSSD input

Only keep duplication calls >10 kbp.
Example: example_data/test.wssd.bed

Prepare WGAC input

Step 1. Choose an appropriate alignment identity for filtering, e.g., 0.98
Graphing the distribution of WGAC call number against the identity is a good way to do this.

Step 2. According the filtering identity, join calls to regions
Command: perl wgac_call2region.pl test.wgac.clean lt 0.98
Output file: test.wgac.clean.region.lt-0.98

Join WSSD and WGAC

Step 1. Merge WSSD and low-identity WGAC results
Command: perl getStat.pl test.wssd.bed test.wgac.clean.region.lt-0.98 test.sd.region.wssd+lt-0.98
Output file: test.sd.region.wssd+lt-0.98

Step 2. Get high-identity WGAC hits
Command: perl filter.pl test.wgac.clean 0.98
Output file: test.wgac.clean.ge-0.98

Step 3. Filter high-identity WGAC hits using WSSD results
Different criteria are used for different lengths of alignments (see lines 133-148 of the program wgac_vs_wssd.pl).
Command: perl wgac_vs_wssd.pl test.wgac.clean.ge-0.98 test.wssd.bed test.wgac.clean.ge-0.98.fltr
Output file: test.wgac.clean.ge-0.98.fltr

Step 4. Merge filtered high-identity WGAC results, WSSD and low-identity WGAC results
Command: perl getStat2.pl test.wgac.clean.ge-0.98.fltr test.sd.region.wssd+lt-0.98 test.sd.region.final
Output file: test.sd.region.final

Final dataset

test.sd.region.final

Please contact Jicai Jiang (jicai.jiang@gmail.com) if you have any questions or concerns.