cstacks

A catalog can be built from any set of samples processed by the **ustacks** or **pstacks** programs. It will create a set of consensus loci, merging alleles together. In the case of a genetic cross, a catalog would be constructed from the parents of the cross to create a set of all possible alleles expected in the progeny of the cross.

Program Options

```
cstacks -P in_dir -M popmap [-n num_mismatches] [--gapped] [-p num_threads] [-b batch_id]
cstacks --aligned -P in_dir -M popmap [-p num_threads] [-b batch_id]
cstacks -s sample1_path [-s sample2_path ...] -o path [-n num_mismatches] [--gapped] [-p num_thre
cstacks --aligned -s sample1_path [-s sample2_path ...] -o path [-p num_threads] [-b batch_id]
```

b — database/batch ID for this catalog (default 1).

P — path to the directory containing Stacks files.

M — path to a population map file.

g,--aligned — base catalog construction on alignment position, not sequence identity.

n — number of mismatches allowed between sample loci when build the catalog (default 1).

p — enable parallel execution with num_threads threads.

s — sample prefix from which to load loci into the catalog.

o — output path to write results.

--catalog [path] — provide the path to an existing catalog. **cstacks** will add data to this existing catalog.

Gapped assembly options:

--gapped — preform gapped alignments between stacks.

--max_gaps — number of gaps allowed between stacks before merging (default: 2).

--min_aln_len — minimum length of aligned sequence in a gapped alignment (default: 0.80).

Advanced options:

m — include tags in the catalog that match to more than one entry (default false).

--k_len [len] — specify k-mer size for matching between between catalog loci (automatically calculated by default).

--report_mmatches — report query loci that match more than one catalog locus.

Example Usage

Here we specify the output directory to write the catalog, and then each sample we want to add to the catalog (one `-s` option for each sample). Here we are allowing up to four, fixed nucleotide differences between loci that we are trying to incorporate into the catalog and we are using 15 threads to speed up the sequence-matching process between loci:

```
% cstacks -b 1 -o ./stacks -s ./stacks/f0_male -s ./stacks/f0_female -n 4 -p 15
```

Here we specify the directory containing all our Stacks output files from **`ustacks`**, and then provide a population map, which contains all of the individual samples we want to add to the catalog within it (see here for information on population maps):

```
% cstacks -P ./stacks -M ./popmap -n 4 -p 15
```

This example is similar to the previous one, except in this case the data is aligned to the reference genome and the Stacks files were generated by **`pstacks`** (`-n` does not apply to aligned data):

```
% cstacks -P ./stacks -M ./popmap --aligned
```

In this example, we start with a pre-existing catalog to which we want to add additional samples, and we want to allow/incorporate gaps between the sample and catalog loci:

```
% cstacks -b 1 -o ./stacks --catalog ./stacks/batch_1 -s ./stacks/sample_37 -s ./stacks/sample_
```

Other Pipeline Programs