ustacks

The unique stacks program will take as input a set of short-read sequences and align them into exactly-matching stacks (or putative alleles). Comparing the stacks it will form a set of putative loci and detect SNPs at each locus using a maximum likelihood framework[1].

[1]Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. *Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags.* **PLoS Genetics**, 6(2):e1000862.

Program Options

```
ustacks -f file_path  -i id -o path [-m min_cov] [-M max_dist] [-p num_threads] [-d] [-t file_
```

f — input file path.

i — a unique integer ID to identify this sample.

o — output path to write results.

m — Minimum depth of coverage required to create a stack (default 2).

M — Maximum distance (in nucleotides) allowed between stacks (default 2).

N — Maximum distance allowed to align secondary reads to primary stacks (default: M + 2).

t — input file Type. Supported types: fasta, fastq, gzfasta, or gzfastq (default: guess).

R — retain unused reads.

H — disable calling haplotypes from secondary reads.

p — enable parallel execution with num_threads threads.

h — display this help messsage.

Stack assembly options:

d — enable the Deleveraging algorithm, used for resolving over merged tags.

--keep_high_cov — disable the algorithm that removes highly-repetitive stacks and nearby errors.

--max_locus_stacks [num] — maximum number of stacks at a single de novo locus (default 3).

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

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).

Model options:

--model_type [type] — either 'snp' (default), 'bounded', or 'fixed'

For the SNP or Bounded SNP model:

--alpha [num] — chi square significance level required to call a heterozygote or homozygote, either 0.1, 0.05 (default), 0.01, or 0.001.

For the Bounded SNP model:

--bound_low [num] — lower bound for epsilon, the error rate, between 0 and 1.0 (default 0).

--bound_high [num] — upper bound for epsilon, the error rate, between 0 and 1.0 (default 1).

For the Fixed model:

--bc_err_freq [num] — specify the barcode error frequency, between 0 and 1.0.

Example Usage

Here we run **ustacks** against four samples from a genetic cross, two parents and two progeny. We assign each sample its own unique integer ID (`-i`) and we specify the parameters for creating putative alleles (`-m`) and mergin alleles into putative loci (`-M`). We speed up the matching process by specifying 15 parallel threads.

```
% ustacks -f ./samples/f0_male.fq    -o ./stacks -i 1 -m 3 -M 4 -p 15
% ustacks -f ./samples/f0_female.fq  -o ./stacks -i 2 -m 3 -M 4 -p 15
% ustacks -f ./samples/progeny_01.fq -o ./stacks -i 3 -m 3 -M 4 -p 15
% ustacks -f ./samples/progeny_02.fq -o ./stacks -i 4 -m 3 -M 4 -p 15
```

Here we run **ustacks** against three samples from a population and we are allowing gapped alignments between alleles when forming putative loci.

```
% ustacks -f ./samples/sample_39-1.fq.gz -o ./stacks -i 1 -M 6 --gapped -p 15
% ustacks -f ./samples/sample_40-2.fq.gz -o ./stacks -i 2 -M 6 --gapped -p 15
% ustacks -f ./samples/sample_41-1.fq.gz -o ./stacks -i 3 -M 6 --gapped -p 15
```

Other Pipeline Programs

Raw Reads

- process_radtags
- process_shortreads
- clone_filter
- kmer_filter

Core

- ustacks
- pstacks
- cstacks
- sstacks
- genotypes
- populations
- rxstacks

Execution control

- denovo_map.pl
- ref_map.pl
- load_radtags.pl

Utilities

- index_radtags.pl
- export_sql.pl
- sort_read_pairs.pl
- exec_velvet.pl