

## mcorr

Using *Correlation Profile* of mutations to infer the recombination rate from large-scale sequencing data in bacteria.

## Software Requirments

- Install `git` from <https://git-scm.com>;
- Install `go` from <https://golang.org/doc/install>;
- Install `python` from <https://www.python.org/>;
- Install `pip3` from <https://pip.pypa.io/en/stable/installing/>.
- Use `pip3` to install required Python packages: `numpy`, `matplotlib`, `lmfit`, and `tqdm`

```
pip3 install --user numpy matplotlib lmfit tqdm
```

## Installation

1. Download and install `mcorr-xmfa` and `mcorr-bam` from your terminal:

```
go get -u github.com/kussell-lab/mcorr/cmd/mcorr-xmfa
go get -u github.com/kussell-lab/mcorr/cmd/mcorr-bam
```

2. Both programs are installed in `$HOME/go/bin`. Add `$HOME/go/bin` to your `$PATH` environment.

We have tested installation in Windows 10, Ubuntu 17.10, and MacOS High Sierra, using Python 3 and Go v1.9.2.

Typical installation time on an iMac is 10 minutes.

## Usage

The inference of recombination parameters requires two steps:

1. Calculate Correlation Profile

For whole-genome alignments (multiple gene alignments), use `mcorr-xmfa`:

```
mcorr-xmfa <input XMFA file> <output prefix>
```

The XMFA files should contain only *coding* sequences. The description of XMFA file can be found in <http://darlinglab.org/mauve/user-guide/files.html>.

For read alignments, use `mcorr-bam`:

```
mcorr-bam <GFF3 file> <sorted BAM file> <output prefix>
```

The GFF3 file is used for extracting the coding regions of the sorted BAM file.

Both programs will produce two files:

- a .csv file stores the calculated Correlation Profile, which will be used for fitting in the next step;
  - a .json file stores the (intermediate) Correlation Profile for each gene.
2. Fit the Correlation Profile using `FitP.py`, which can be found in `$HOME/go/src/github.com/kussell-lab/mcorr/cmd/fitting/`:

```
python3 $HOME/go/src/github.com/kussell-lab/mcorr/cmd/fitting/FitP.py <.csv file> <output prefix>
```

It will produce two files:

- `<output_prefix>_best_fit.svg` – the plots of the Correlation Profile, fitting, and residuals;
- `<output_prefix>_fit_results.csv` – the table of fitted parameters.

Example data can be found [https://github.com/kussell-lab/mcorr\\_examples](https://github.com/kussell-lab/mcorr_examples).