ITSxpress Documentation

Release 1.7.0

Adam R. Rivers

Feb 05, 2020

Contents:

CHAPTER 1

ITSxpress: Software to rapidly trim the Internally transcribed spacer (ITS) region of FASTQ files

1.1 Author

- Adam R. Rivers, US Department of Agriculture, Agricultural Research Service

1.2 Citation

Rivers AR, Weber KC, Gardner TG et al. ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis [version 1; referees: awaiting peer review]. F1000Research 2018, 7:1418 (doi: 10.12688/f1000research.15704.1)

1.3 Introduction

The internally transcribed spacer region is a region between highly conserved the small subunit (SSU) of rRNA and the large subunit (LSU) of the rRNA. In Eukaryotes it contains the 5.8s genes and two variable length spacer regions. In amplicon sequencing studies it is common practice to trim off the conserved (SSU, 5,8S or LSU) regions. Bengtsson-Palme et al. (2013) published software the software package ITSx to do this.

ITSxpress is designed to support the calling of exact sequence variants rather than OTUs. This newer method of sequence error-correction requires quality score data from each sequence, so each input sequence must be trimmed. ITSXpress makes this possible by taking FASTQ data, de-replicating the sequences then identifying the start and stop sites using HMMSearch. Results are parsed and the trimmed files are returned. The ITS 1, ITS2 or the entire ITS region including the 5.8s rRNA gene can be selected. ITSxpress uses the hmm model from ITSx so results are comparable.

ITSxpress is also available as a QIIME2 Plugin

1.4 Installation

ITSxpress can be installed from:

1. Bioconda: (preferred method because it handles dependencies):

```
conda install itsxpress
```

2. Pip: https://pypi.org/project/itsxpress/:

```
pip install itsxpress
```

3. The Github repository: https://github.com/USDA-ARS-GBRU/itsxpress

```
git clone https://github.com/USDA-ARS-GBRU/itsxpress.git
```

1.5 Dependencies

The software requires Vsearch, BBtools, Hmmer >= 3.1b and Biopython. Bioconda takes care of this for you so it is the preferred installation method.

1.6 Usage

-h, --help	Show this help message and exit.
--fastq	A `.fastq`, `.fq`, `.fastq.gz` or `.fq.gz` file. Interleaved or not. Required.
--single_end	A flag to specify that the fastq file is single-ended (not paired). single-ended (not paired). Default is false.
--fastq2	A `.fastq`, `.fq`, `.fastq.gz` or `.fq.gz` file representing read 2 if present, optional.
--outfile	The trimmed FASTQ file, if it ends in `gz` it will be gzipped.
--outfile2	The trimmed FASTQ read 2 file, if it ends in `gz` it will be gzipped. If used, reads will be retuned as unmerged pairs rather than than merged.
--tempdir	Specify the temp file directory. Default is None.
--keeptemp	Should intermediate files be kept? Default is false.
--region	Options : {ITS2, ITS1, ALL}

--taxa	Select the taxonomic group sequenced: {Alveolata, Bryophyta, Bacillariophyta, Amoebozoa, Euglenozoa, Fungi, Chlorophyta, Rhodophyta, Phaeophyceae, Marchantiophyta, Metazoa, Oomycota, Haptophyceae, Raphidophyceae, Rhizaria, Synurophyceae, Tracheophyta, Eustigmatophyceae, All}. Default Fungi.
--cluster_id	The percent identity for clustering reads range [0.99-1.0], set to 1 for exact dereplication. Default 1.0.
--log	Log file. Default is ITSxpress.log.
--threads	Number of processor threads to use. Default is 1.
--reversed_primers	Primers are in reverse orientation as in Taylor et al. 2016, DOI:10.1128/AEM.02576-16. If selected ITSxpress returns trimmed reads flipped to the forward orientation

1.7 Examples

Use case 1: Trimming the ITS2 region from a fungal amplicon sequencing dataset with forward and reverse gzipped FASTQ files using two cpu threads. Return a single merged file for use in Deblur.

```
itsxpress --fastq r1.fastq.gz --fastq2 r2.fastq.gz --region ITS2 \
--taxa Fungi --log logfile.txt --outfile trimmed_reads.fastq.gz --threads 2
```

ITSxpress can take gzipped or un-gzipped FASTQ files and it can write gzipped or un-gzipped FASTQ files. It expects FASTQ files to end in: .fq, .fastq, .fq.gz or fastq.gz.

Use case 2: Trimming the ITS2 region from a fungal amplicon sequencing dataset with forward and reverse gzipped FASTQ files using two cpu threads. Return a forward and reverse read files for use in Dada2.

```
itsxpress --fastq r1.fastq.gz --fastq2 r2.fastq.gz --region ITS2 \
--taxa Fungi --log logfile.txt --outfile trimmed_reads.fastq.gz --threads 2
```

ITSxpress can take gzipped or un-gzipped FASTQ files and it can write gzipped or un-gzipped FASTQ files. It expects FASTQ files to end in: .fq, .fastq, .fq.gz or fastq.gz.

Use case 3: Trimming the ITS2 region from a fungal amplicon sequencing dataset with an interleaved gzipped FASTQ files using two cpu threads. Return a single merged file for use in Deblur.

```
itsxpress --fastq interleaved.fastq.gz  --region ITS2 --taxa Fungi \
--log logfile.txt --outfile trimmed_reads.fastq.gz --threads 2
```

Use case 4: Trimming the ITS2 region from a fungal amplicon sequencing dataset with an single-ended gzipped FASTQ files using two cpu threads.

```
itsxpress --fastq single-end.fastq.gz --single_end --region ITS2 --taxa Fungi \
--log logfile.txt --outfile trimmed_reads.fastq.gz --threads 2
```

Single ended data is less common and may come from a dataset where the reads have already been merged.

Use case 5: Trimming the ITS1 region from a Alveolata amplicon sequencing dataset with an interleaved gzipped FASTQ files using 8 cpu threads.

```
itsxpress --fastq interleaved.fastq.gz --region ITS1 --taxa Alveolata \
--log logfile.txt --outfile trimmed_reads.fastq.gz --threads 8
```

1.8 License information

CHAPTER 2

itsxpress

2.1 itsxpress package

2.1.1 Submodules

2.1.2 itsxpress.definitions module

2.1.3 itsxpress.main module

2.1.4 Module contents

CHAPTER 3

Indices and tables

- genindex
- modindex
- search