SURPI

How to run SURPI

June, 2014

The test dataset used here (SRR1106548) is derived from plasma samples spiked with known titers of HIV (10^4, 10^3, 10^2 copies/ml) as described in Supplemental Methods Naccache et al 2014 and is found here.

The test dataset SRR1106548.sra includes paired-end data as well as orphan single-end read 1 (R1) and read 2 (R2) data from 3 separately barcoded samples corresponding to spiked HIV titers of 10^4, 10^3, 10^2 copies/ml. To restore the original FASTQ files (with the human reads removed), execute the following steps:

1) Install SURPI locally, or start up an AWS instance (see *Readme_AWS* for details).

2) Install the SRA Toolkit (required for fastq-extractBarcodedSRA.sh). Installation instructions can be found here.

3) Extract barcoded reads from SRR1106548.sra.

```
$ fastq-extractBarcodedSRA.sh SRR1106548.sra
```

4) Combine extracted FASTQ files into one input file SRR1106548.fastq.

```
$ cat bc*.fastq > SRR1106548.fastq
```

5) Create a config file

```
$ SURPI.sh -z SRR1106548.fastq
```

6) Modify resulting SRR1106548.config file as needed. The below parameters should be set as shown:

```
quality= "Illumina"
adapter_set="Truseq"
```

7) Start SURPI

```
$ ./go_SRR1106548 &
```

We've found it useful to monitor a SURPI run by using htop, and monitoring the logfile (`SURPI.SRR1106548.log` for the above example)

When complete, SURPI (when running in comprehensive mode) generates the following folders within the original run folder:

```
DATASETS_SRR1106548
deNovoASSEMBLY_SRR1106548
LOG_SRR1106548
OUTPUT_SRR1106548
TRASH_SRR1106548
```

The most useful data can be found in the `OUTPUT_SRR1106548` folder. For more information on this data, refer to the *Output_Interpretation.pdf* document.