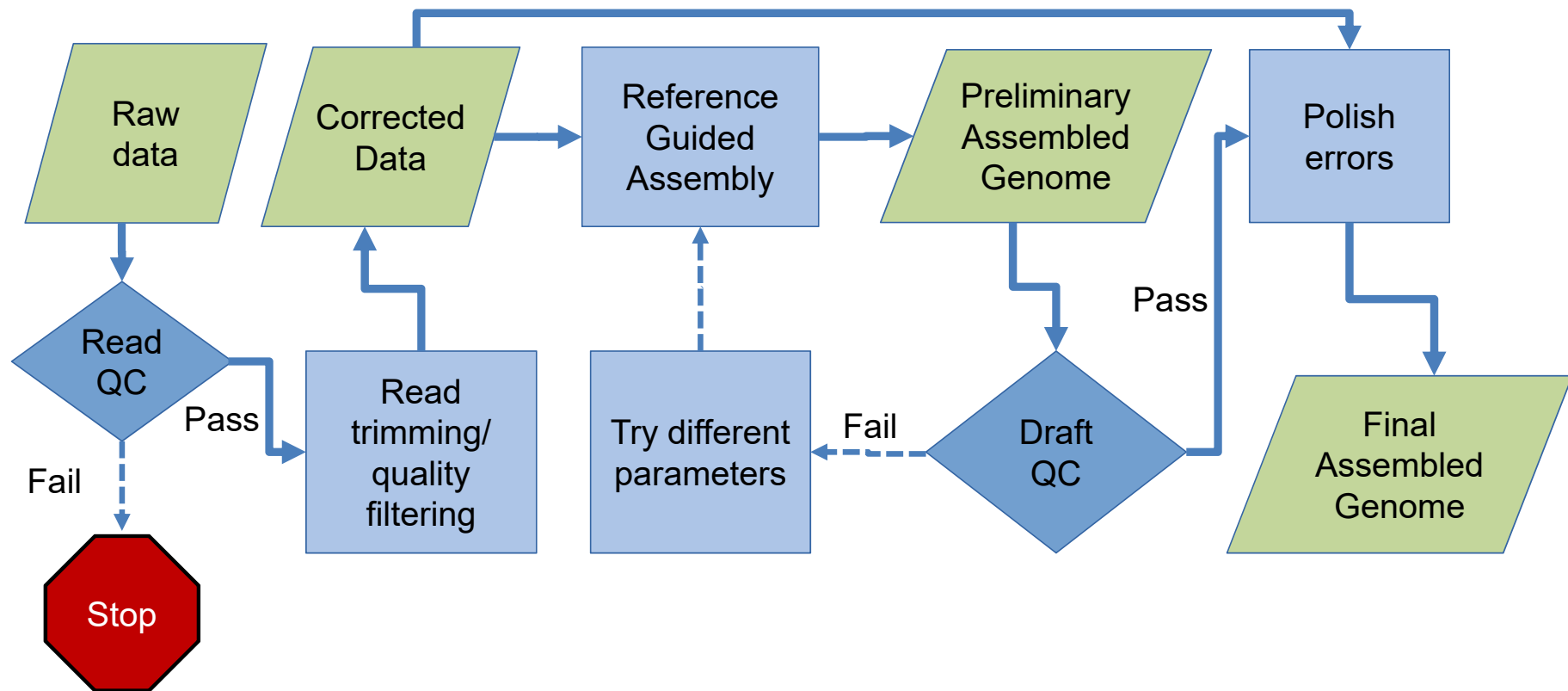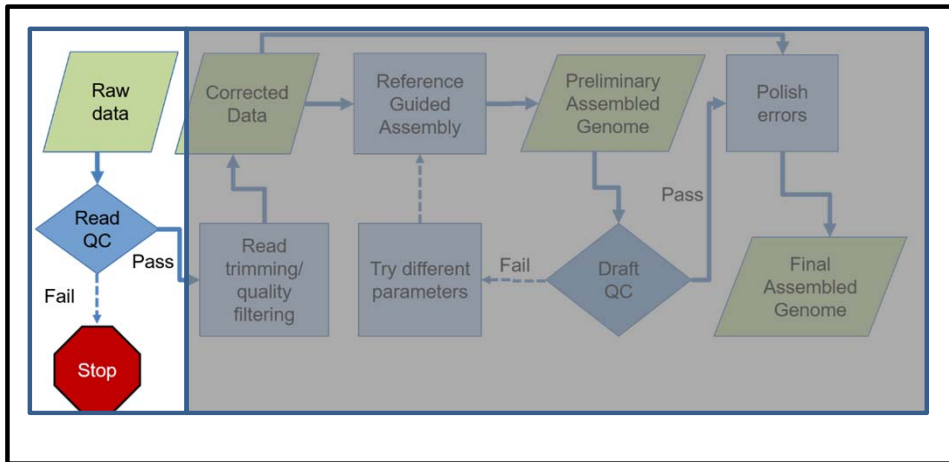Assembly pipeline: How we get from raw data to the final assembled genome.
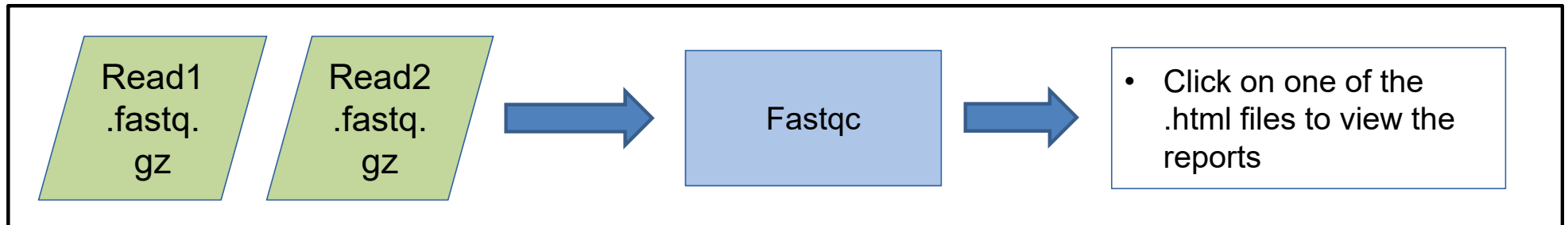
Location in the overall pipeline

| Raw data |
| Read QC |
Pass
Fail
Stop

Corrected Data → Reference Guided Assembly → Preliminary Assembled Genome → Polish errors

Read trimming/ quality filtering → Try different parameters → Draft QC

Fail
Pass

Final Assembled Genome

- **Start with raw sequencing data, in fastq format and zipped.**
- **Remember, there are two reads for each DNA fragment. The first read of each fragment is stored in one file, and the second read of each fragment is stored in another.**
- **Run Fastqc, a program that summarizes the quality of reads. Also outputs a number of useful metrics.**

Plain English description of the steps in the pipeline.

Read1 .fastq. gz   Read2 .fastq. gz   →   Fastqc   →   - Click on one of the .html files to view the reports

Inputs and outputs for the current step of the pipeline.