YaHS·····

·Hi-C BAM·YaHS scaffold·····AGP·FASTA·····

YaHS··Hi-C·····contig···scaffold·····BAM·····contig·····
·····read·····
·····

··· | ···· | ····

···· | FASTA···BAM··ID····· | ·····

BAM | quickcheck····· | ····read group·····

····· | ·····-q 10 | ·····

·· | ····AGP·FASTA···· | ·····N50

·····

··contig FASTA·····FAI·Hi-C BAM·····YaHS····YaHS·····
·····SORT_MEM·samtools sort·····

```
set -euo pipefail
: "${ASM:?contig FASTA····}" "${BAM:?Hi-C BAM····}"
: "${SORT_THREADS:?·····}" "${SORT_MEM:?··512M·····}"
: "${OUT:?·····}"
[[ "$SORT_THREADS" =~ ^[1-9][0-9]*$ ]] || exit 2
for f in "$ASM" "$ASM.fai" "$BAM"; do test -s "$f" || exit 2; done
test ! -e "$OUT" && test ! -L "$OUT" || exit 2
mkdir "$OUT"
yahs --version > "$OUT/version.txt" 2>&1
sha256sum "$ASM" "$ASM.fai" "$BAM" > "$OUT/inputs.sha256"
samtools quickcheck -v "$BAM"
samtools view -H "$BAM" > "$OUT/header.sam"
python check_reference.py "$ASM.fai" "$OUT/header.sam"
samtools sort -n -@ "$SORT_THREADS" -m "$SORT_MEM" \
  -T "$OUT/name-sort" -o "$OUT/name-sorted.bam" "$BAM"
samtools quickcheck -v "$OUT/name-sorted.bam"
/usr/bin/time -v -o "$OUT/yahs.time.txt" \
  yahs -q 10 -o "$OUT/yahs" "$ASM" "$OUT/name-sorted.bam" \
  > "$OUT/yahs.log" 2>&1
test -s "$OUT/yahs_scaffolds_final.agp"
test -s "$OUT/yahs_scaffolds_final.fa"
samtools faidx "$OUT/yahs_scaffolds_final.fa"
printf '%s\n' COMMAND_AND_FILES_OK > "$OUT/status.txt"
```

·····FASTA···BAM·@SQ·····ID/·····BAM·····read·····

```
import sys

def add(d, name, length):
    length = int(length)
    if not name or name in d or length <= 0:
        raise ValueError("··ID·····")
    d[name] = length

fai, sq = {}, {}
with open(sys.argv[1]) as f:
    for line in f:
        fields = line.rstrip("\n").split("\t")
        if len(fields) < 2:
            raise ValueError("FAI···")
        add(fai, fields[0], fields[1])
with open(sys.argv[2]) as f:
    for line in f:
        if line.startswith("@SQ\t"):
            tags = dict(x.split(":", 1)
                        for x in line.rstrip("\n").split("\t")[1:])
            add(sq, tags["SN"], tags["LN"])
if not fai or fai != sq:
    raise ValueError("FASTA·BAM·····ID/·····")
print("REFERENCE_DICTIONARY_OK", len(fai))
```

·····YaHS··

·samtools sort·····I/O·····BAM·
·····YaHS·····FASTA·BAM·····set
-e·····stderr·time·····

·····

·····scaffold·····AGP·scaffold·····contig·····gap·····Hi-C
·····
·····

·····

·····BAM·····RSS·····BAM
·····YaHS·····
·CPU·····

·····assembly·····
·····scaffold·····N50·····

·····

·····AGP·FASTA···AGP··contig·····Hi-C····YaH
S·····contig·····contig·····
·····N50·····

·····YaHS/samtools·····AGP·FASTA·····
·····YaHS·····

·····contig·····scaffold···AGP·····FASTA·····
·····scaffold·····

····

YaHS··README····· https://github.com/c-zhou/yahs

YaHS····· https://raw.githubusercontent.com/c-zhou/yahs/main/yahs.c

Samtools sort·· https://www.htslib.org/doc/samtools-sort.html