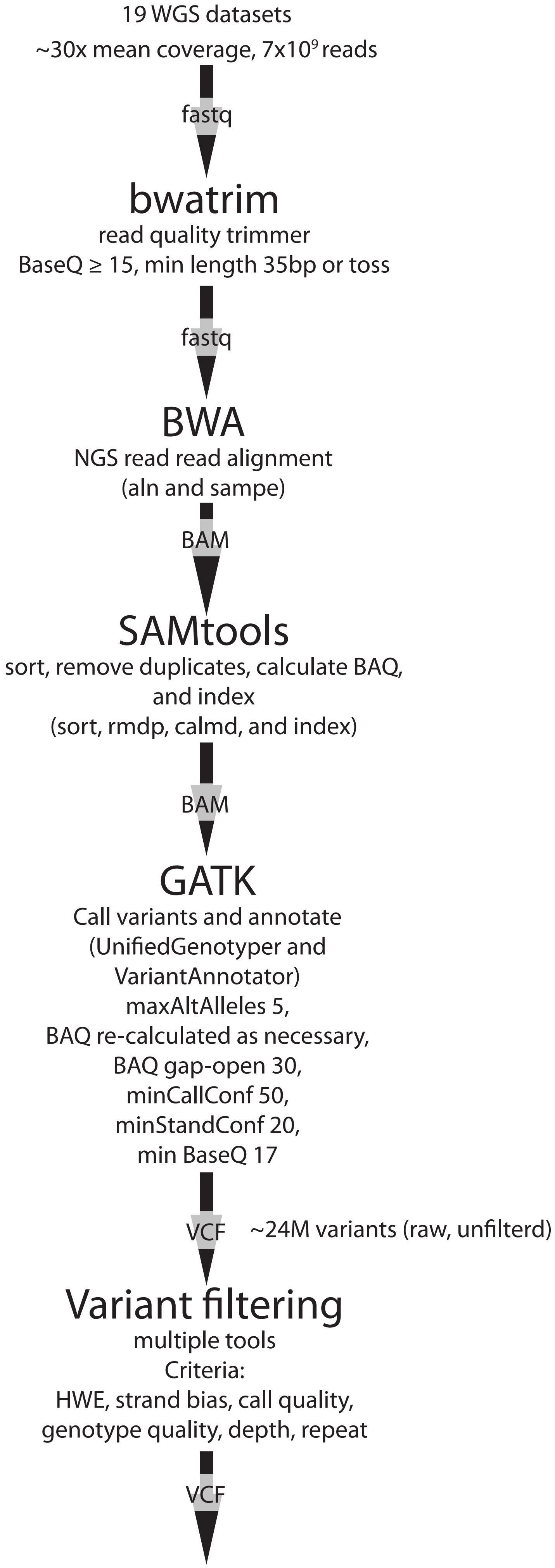
19 WGS datasets
~30x mean coverage, 7x10⁹ reads
fastq
bwatrim
read quality trimmer
BaseQ ≥ 15, min length 35bp or toss
fastq
BWA
NGS read read alignment
(aln and sampe)
BAM
SAMtools
sort, remove duplicates, calculate BAQ,
and index
(sort, rmdp, calmd, and index)
BAM
GATK
Call variants and annotate
(UnifiedGenotyper and
VariantAnnotator)
maxAltAlleles 5,
BAQ re-calculated as necessary,
BAQ gap-open 30,
minCallConf 50,
minStandConf 20,
min BaseQ 17
VCF
~24M variants (raw, unfilterd)
Variant filtering
multiple tools
Criteria:
HWE, strand bias, call quality,
genotype quality, depth, repeat
VCF