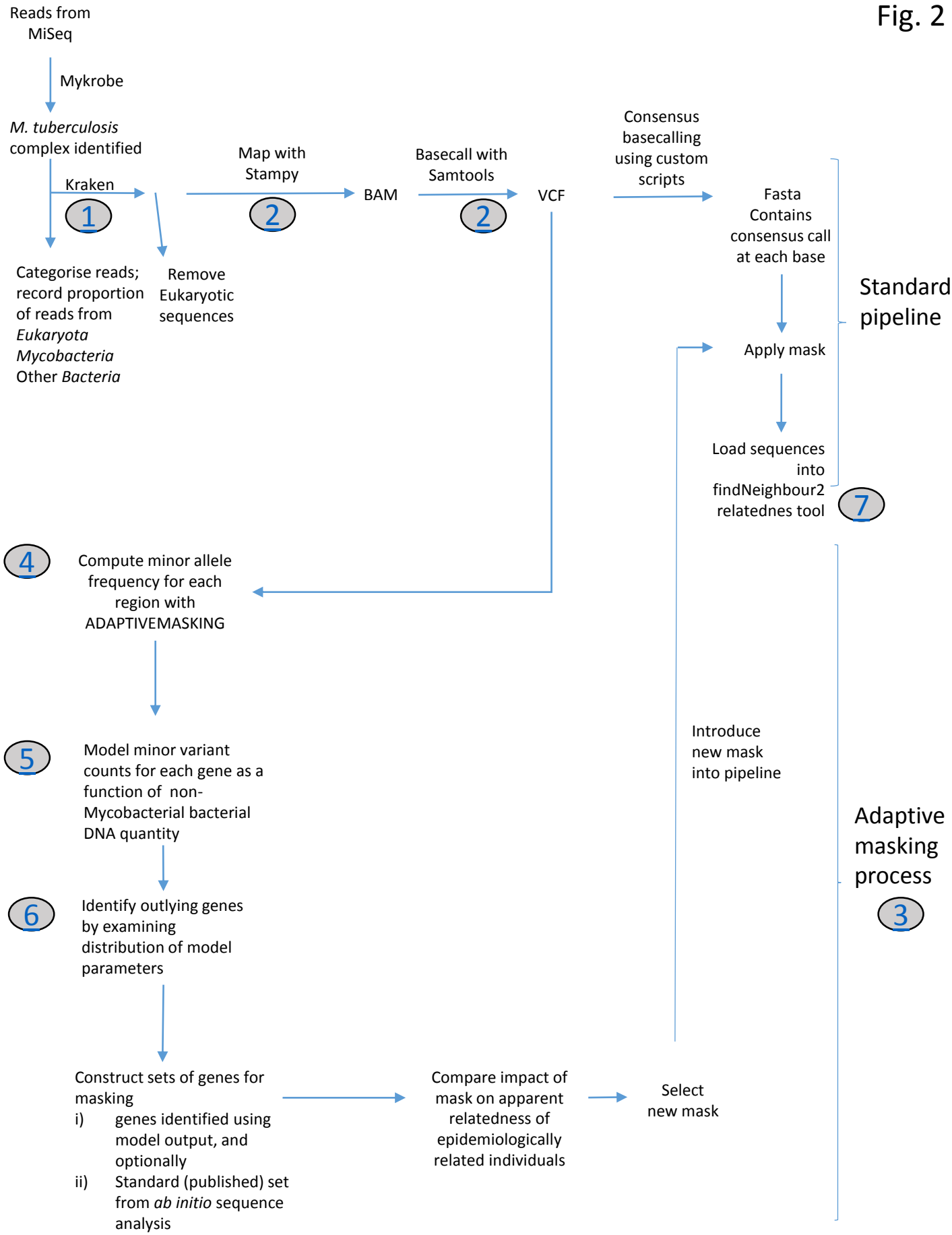
Fig. 2


Reads from MiSeq
Mykrobe
M. tuberculosis complex identified
Kraken
1
Categorise reads; record proportion of reads from Eukaryota Mycobacteria Other Bacteria
Remove Eukaryotic sequences
Map with Stampy
2
BAM
Basecall with Samtools
2
VCF
Consensus basecalling using custom scripts
Fasta Contains consensus call at each base
Apply mask
Load sequences into findNeighbour2 relatednes tool
7
Standard pipeline
4
Compute minor allele frequency for each region with ADAPTIVEMASKING
5
Model minor variant counts for each gene as a function of non-Mycobacterial bacterial DNA quantity
6
Identify outlying genes by examining distribution of model parameters
Construct sets of genes for masking
i) genes identified using model output, and optionally
ii) Standard (published) set from ab initio sequence analysis
Compare impact of mask on apparent relatedness of epidemiologically related individuals
Select new mask
Introduce new mask into pipeline
Adaptive masking process
3