Required

- **chr** What chromosome to run. Should match BAM header
- **posfile** Where to find file with positions to run. File is tab separated with no header, one row per SNP, with col 1 = chromosome, col 2 = physical position (sorted from smallest to largest), col 3 = reference base, col 4 = alternate base. Bases are capitalized. STITCH only handles bi-allelic SNPs
- **K** Integer, how many founder / mosaic haplotypes to use
- **nGen** Estimated number of generations since founding. In uncertain, estimate using 4 * Ne / K, where Ne is the effective population size
- **outputdir** What output directory to use / where output files go
- **bamlist** (one of bamlist or cramlist required) Path to file with BAM file locations. File is one row per entry, path to BAM files. BAM index files should exist in same directory as for each BAM, suffixed either .bam.bai or .bai
- **cramlist** (one of bamlist or cramlist required) Same as bamlist, but path is to CRAM locations. If used, requires reference variable to be set with path to fasta file

Optional

- **method** How to run imputation, either diploid or pseudoHaploid, the former being more accurate but having quadratic time complexity in K and the later being less accurate but having linear time complexity in K.
- **switchModelIteration** When selected, the iteration to switch from pseudoHaploid to diploid. Note that one EM iteration is defined as first using the parameters to estimate hidden phase, and secondly to use hidden phase to update parameters. So a choice of 39 with iterations = 40 would means 38 complete pseudo-haploid iterations, a 39th iteration of both estimating hidden phase and updating parameters, and a 40th iteration of updating hidden phase, and from this estimating dosages (parameter updates on the 40th iteration have no influence on dosages). Therefore, we say that a choice of 39 gives 38 pseudo-haploid iterations and 2 diploid iterations
- **genfile** Path to gen file with high coverage results. Empty for no genfile. File has a header row with a name for each sample, matching what is found in the bam file. Each subject is then a tab seperated column, with 0 = hom ref, 1 = het, 2 = hom alt and NA indicating missing genotype, with rows corresponding to rows of the posfile. Note therefore this file has one more row than posfile which has no header
- **regionStart** When running imputation, where to start from
- **regionEnd** When running imputation, where to stop
- **buffer** Buffer of region to perform imputation over. Imputation is run from bases including regionStart - buffer to regionEnd + buffer, including the bases, with 1-based positions. After imputation, the VCF is shrunk to only include positions from regionStart to

regionEnd, inclusive

- **reference_haplotype_file** When initializing using a reference panel, path to haplotype files in IMPUTE2 format (see example R script for an example download). Note that the haplotypes are just used for initialization and not for updating, so after 1 iteration only reads from the avaiable samples will be used for updating the hidden parameters and for subsequent calculations
- **reference_legend_file** When initializing using a reference panel, path to legend file in IMPUTE2 format
- **reference_sample_file** When initializing using a reference panel, path to sample file in IMPUTE2 format
- **reference_populations** When initializing using a reference panel, vector of populations from sample file POP column to use (see example R script for an example)
- **inputBundleBlockSize** How many sample input files to bundle together to reduce number of temporary files. Default NA or not used. Recommended to set to 100 or greater when using large sample sizes (> ~5000)
- **vcf_output_name** Override the default VCF output name with this given file name. Please note that this does not change the names of inputs or outputs (e.g. RData, plots), so if outputdir is unchanged and if multiple STITCH runs are processing on the same region then they may over-write each others inputs and outputs

Output

- VCF named stitch....vcf.gz, or if no regionStart and regionEnd is given stitch..vcf.gz
- **vcf_output_name** overrides default name to give VCF output