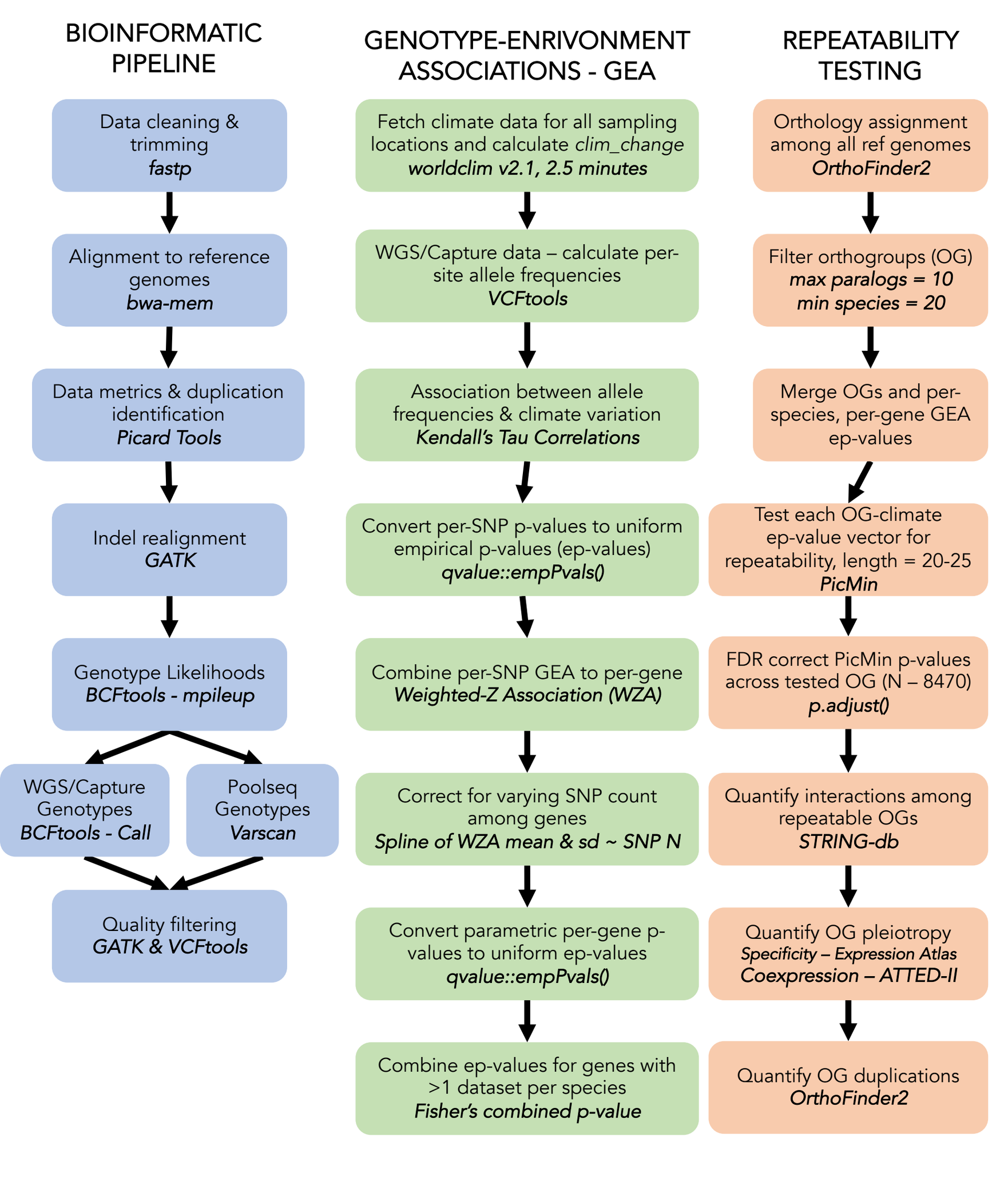
BIOINFORMATIC PIPELINE
Data cleaning & trimming
fastp
Alignment to reference genomes
bwa-mem
Data metrics & duplication identification
Picard Tools
Indel realignment
GATK
Genotype Likelihoods
BCFtools - mpileup
WGS/Capture Genotypes
BCFtools - Call
Poolseq Genotypes
Varscan
Quality filtering
GATK & VCFtools
GENOTYPE-ENRIVONMENT ASSOCIATIONS - GEA
Fetch climate data for all sampling locations and calculate clim_change
worldclim v2.1, 2.5 minutes
WGS/Capture data – calculate per-site allele frequencies
VCFtools
Association between allele frequencies & climate variation
Kendall's Tau Correlations
Convert per-SNP p-values to uniform empirical p-values (ep-values)
qvalue::empPvals()
Combine per-SNP GEA to per-gene
Weighted-Z Association (WZA)
Correct for varying SNP count among genes
Spline of WZA mean & sd ~ SNP N
Convert parametric per-gene p-values to uniform ep-values
qvalue::empPvals()
Combine ep-values for genes with >1 dataset per species
Fisher's combined p-value
REPEATABILITY TESTING
Orthology assignment among all ref genomes
OrthoFinder2
Filter orthogroups (OG)
max paralogs = 10
min species = 20
Merge OGs and per-species, per-gene GEA ep-values
Test each OG-climate ep-value vector for repeatability, length = 20-25
PicMin
FDR correct PicMin p-values across tested OG (N – 8470)
p.adjust()
Quantify interactions among repeatable OGs
STRING-db
Quantify OG pleiotropy
Specificity – Expression Atlas
Coexpression – ATTED-II
Quantify OG duplications
OrthoFinder2