Field isolates	Targeted amplicon sequencing	Population-level diversity	Within-host repertoire diversity	Repertoire diversity between hosts

each ⬤ = unique *var* DBLα type

each 🟪🟨🟦🟥🟩 = *var* DBLα repertoire
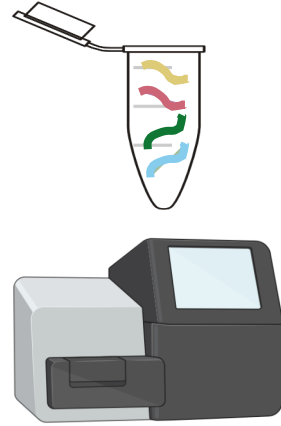
UniMelb-Day-Lab / **DBLaCleaner**
UniMelb-Day-Lab / **clusterDBLalpha**
UniMelb-Day-Lab / **classifyDBLalpha**

Upper bound 95% HDPI
Posterior mean
Lower bound 95% HDPI

P_{TS} BP_{TS}

DNA extraction from parasite isolates collected from malaria-infected individuals

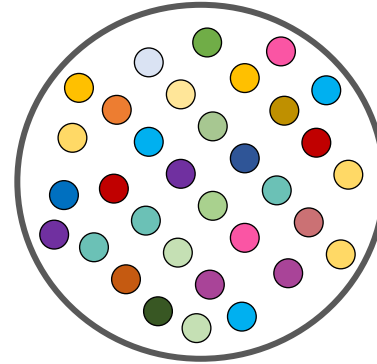PCR amplification of DBLα domain of *var* genes using degenerate primers

Each isolate is individually barcoded and pooled

High-throughput Illumina sequencing
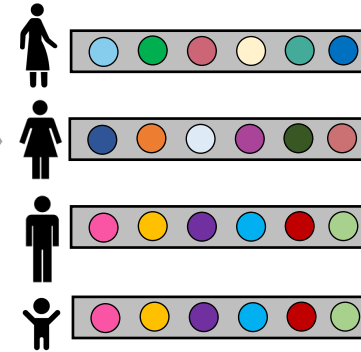
Bioinformatic pipelines for sequence processing

The pool of unique *var* DBLα types in the population is defined by clustering all DBLα sequences at 96% sequence identity

Type-specific population frequencies can be calculated

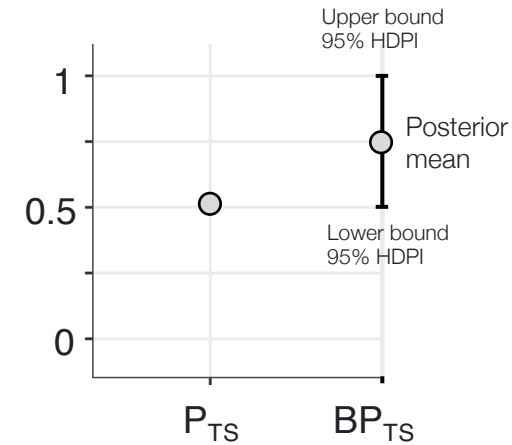Types can be further classified into upsA and non-upsA groups based on DBLα sub-domains

Repertoires are constructed for each isolate based on the total number of unique *var* DBLα types identified

Repertoires missing data (i.e. types) can occur due to the use of degenerate primers for the *var* DBLα PCR

Pairwise type sharing (P_{TS}) statistics measure genetic similarity and relatedness between two repertoires

Bayesian pairwise type sharing (BP_{TS}) accounts for missing data and provides a posterior mean estimate and high-density posterior intervals (HDPI)

Specific thresholds can be applied to define *var*codes (e.g. ≥0.90) and recombinant *var*codes (e.g. ≥0.50)
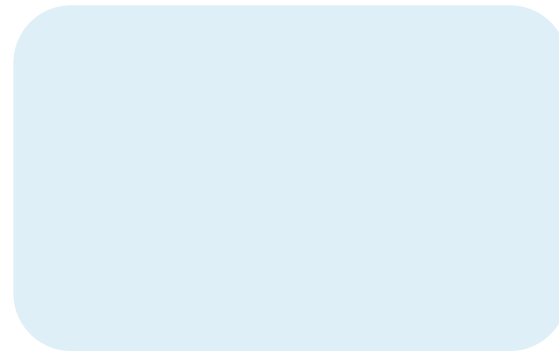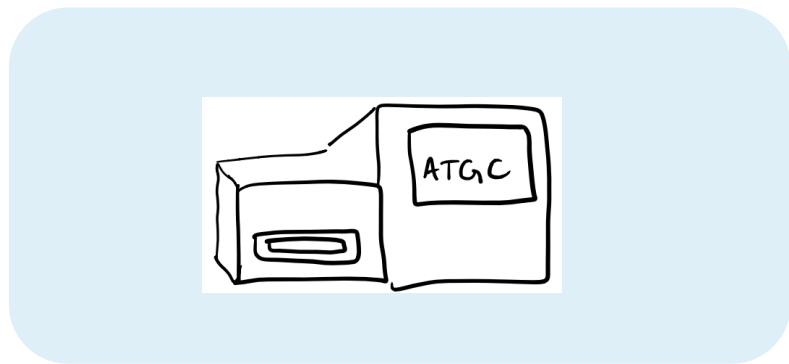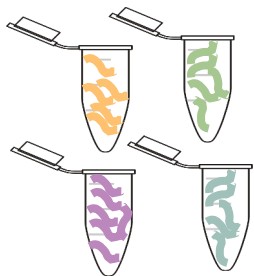
#88CCEE	#CC6677	#DDCC77	#117733
#332288	#AA4499	#44AA99	#999933
#882255	#661100	#6699CC	#888888

ATGC