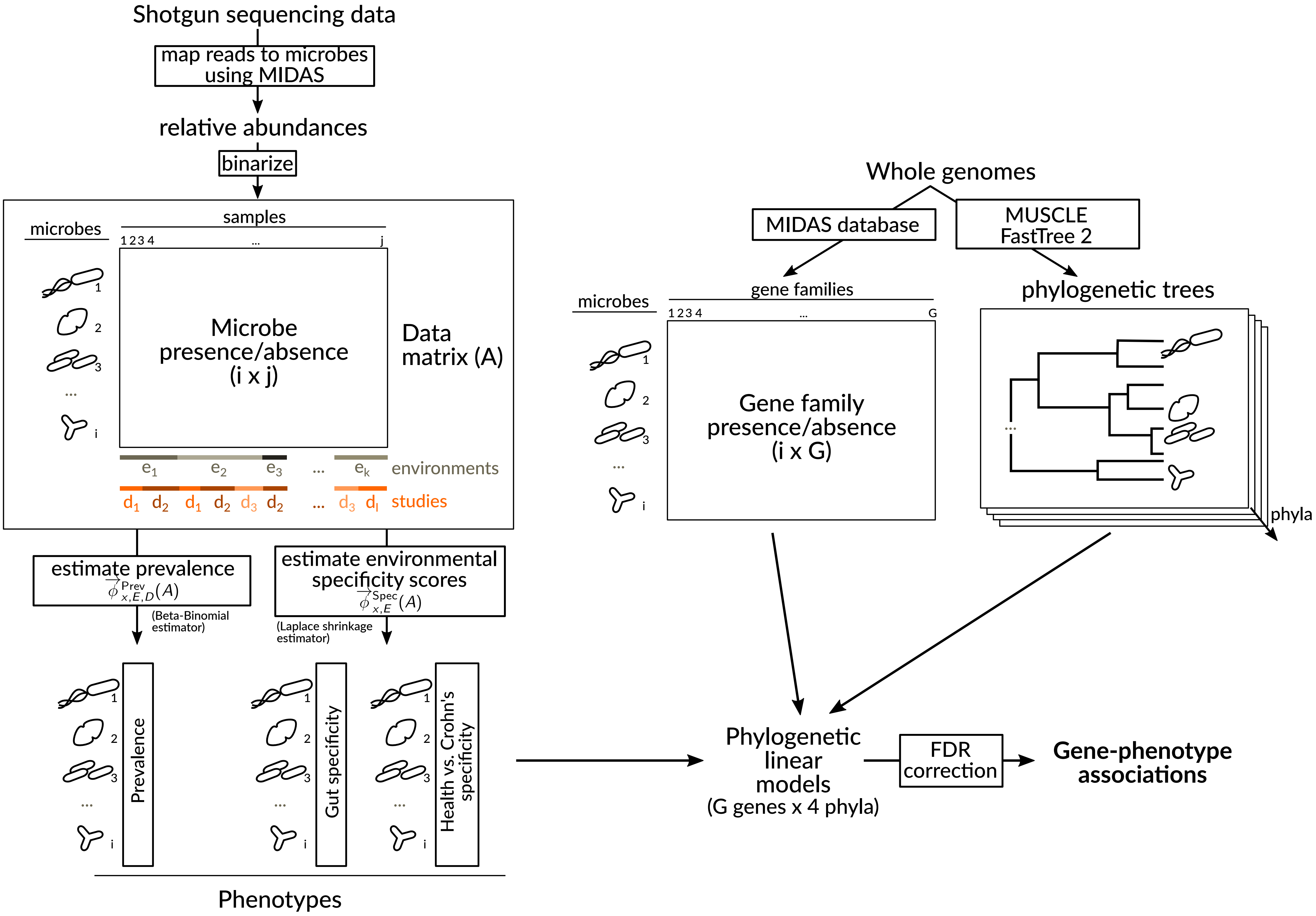
Shotgun sequencing data
map reads to microbes using MIDAS
relative abundances
binarize
samples
microbes
1 2 3 4 ... j
1
2
3
...
i
Microbe presence/absence (i x j)
Data matrix (A)
e1 e2 e3 ... ek environments
d1 d2 d1 d2 d3 d2 ... d3 dl studies
estimate prevalence $\overrightarrow{\phi}^{\text{Prev}}_{x,E,D}(A)$
(Beta-Binomial estimator)
estimate environmental specificity scores $\overrightarrow{\phi}^{\text{Spec}}_{x,E}(A)$
(Laplace shrinkage estimator)
Prevalence
Gut specificity
Health vs. Crohn's specificity
Phenotypes
Whole genomes
MIDAS database
MUSCLE FastTree 2
gene families
microbes
1 2 3 4 ... G
Gene family presence/absence (i x G)
phylogenetic trees
phyla
Phylogenetic linear models (G genes x 4 phyla)
FDR correction
Gene-phenotype associations