

FAULDS: A Bayesian MCMC Framework for Likelihood-Based Decomposition of DNA Mixtures

Taehyeon Kim^{1,2}, Sang Un Park³, Hwan Young Lee^{3,4,*} and Sangwoo Kim^{1,2,*}

¹*Yonsei University College of Medicine, Graduate School of Medical Science, Brain Korea 21 FOUR Project for Medical Science, Seoul, South Korea*

²*Yonsei University College of Medicine, Department of Biomedical Systems Informatics, Seoul, South Korea*

³*Department of Forensic Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea*

⁴*Institute of Forensic and Anthropological Science, Seoul National University College of Medicine, Seoul, Korea*

*Corresponding author: hylee192@snu.ac.kr, SWKIM@yuhs.ac

Interpreting DNA mixtures remains a major computational challenge because mixture proportions, contributor-specific genotype combinations, and the number of contributors (NoC) are all inherently uncertain. We introduce FAULDS, a SNP-based Bayesian framework that integrates Gibbs sampling with reversible jump Markov chain Monte Carlo (RJMCMC) to jointly estimate contributor proportions, genotypes, and NoC. The model represents observed SNP data as clusters of genotype combinations, which serve as the basis for reconstructing contributor structure. To ensure robustness in noisy sequencing environments, the model incorporates explicit correction for sequencing errors and allele-specific bias in its emission probabilities. Mixture proportions are sampled from Dirichlet priors, and genotype updates are accepted through Metropolis–Hastings ratios. RJMCMC further enables the model to move across hypotheses with different contributor counts via analytically derived Jacobian transformations. In *in vitro* and *in silico* experiments, FAULDS reliably identified minor contributors down to 1% of a mixture and accurately recovered both NoC and mixture ratios, even when applied to a downsampled SNP set of only about 1,000 loci. To our knowledge, this work provides the first RJMCMC-based approach for direct NoC inference, offering a probabilistic decomposition strategy broadly applicable to forensic DNA analysis as well as other complex DNA mixture problems.