Inferring repertoire dynamics and using them to identify responsive clones

Maximilian Puelma Touzel and Aleksandra Walczak
Laboratoire de Physique Théorique, ENS-PSL Research University, Paris, France

Thierry Mora
Laboratoire de Physique Statistique, ENS-PSL Research University, Paris, France,

Abridged: Here, we present a method to infer repertoire dynamics from repertoire-sequenced receptor RNA. We analyze the structure of the model used, justifying the ingredients and apply it to answer questions regarding the repertoire dynamics of yellow fever.

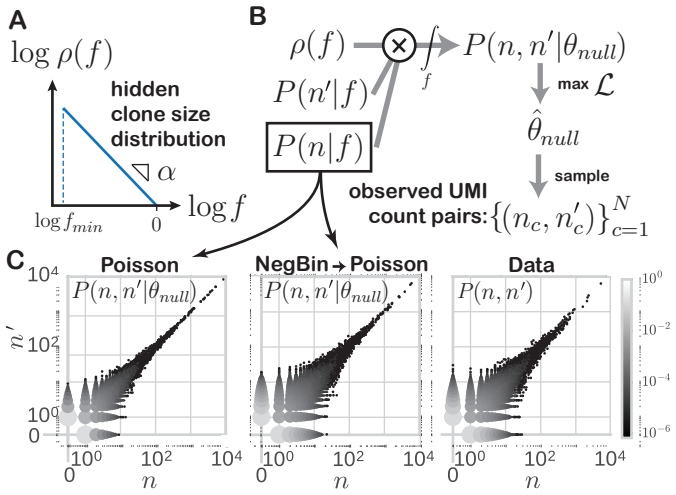

Figure 1. *Learning a null model.* **A** Clone frequency distribution, $\rho(f)$, is set as a power law, parameterized by the power, α and the minimum frequency, f_{min}. **B** Null model parameters are learned by marginalizing over clone frequency, f, and maximizing this marginal likelihood with respect to the parameters. Sampled repertoires can then be generated with the ML estimate. **C** Models and data comparison using molecule pair count statistics (example donor: S2; day 0-day 0 comparison). The learned model for Poisson distributed $P(n|f)$ (left) and for $P(n|f)$ set as a negative binomial distribution in cell counts controlling the scale parameter of a Poisson distribution of molecule counts (center; fig.2). Right: empirical pair count histogram. Gray scale bar denotes the empirical frequency of a (n, n') pair.

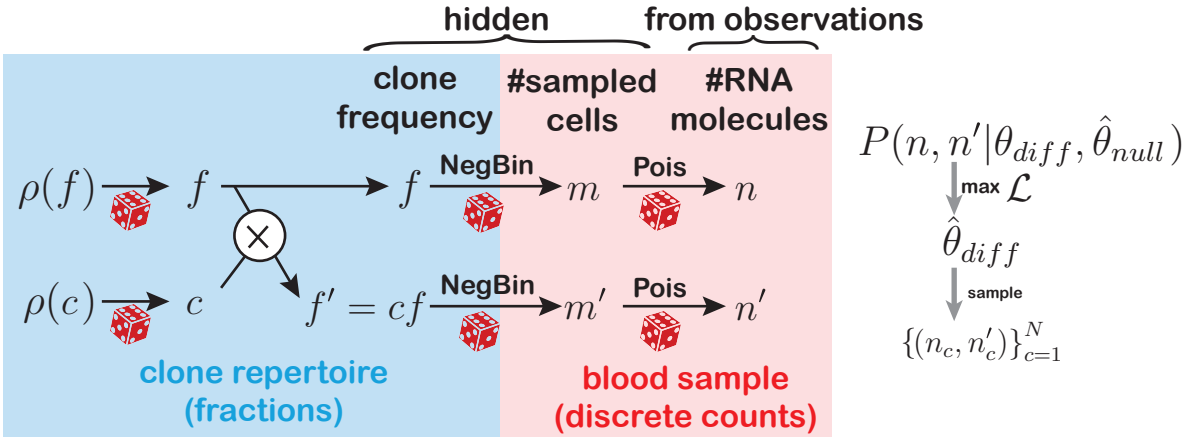

Figure 2. *Differential expression model structure and learning procedure.* $P(n|f)$ is set as a negative binomial distribution in cell counts controlling the scale parameter of a Poisson distribution of molecule counts. In the differentially expressed condition (prime-decorated quantities), the parameters, θ_{diff}, of the prior distribution, $\rho(c)$, of fold-change, c, are learned by maximizing the marginal likelihood (see section *Fold change prior*) with respect to θ_{diff}, keeping the remaining parameters, θ_{null}, fixed to their ML estimates, $\hat{\theta}_{null}$, previously obtained using same-day replicate data (see fig. 1).

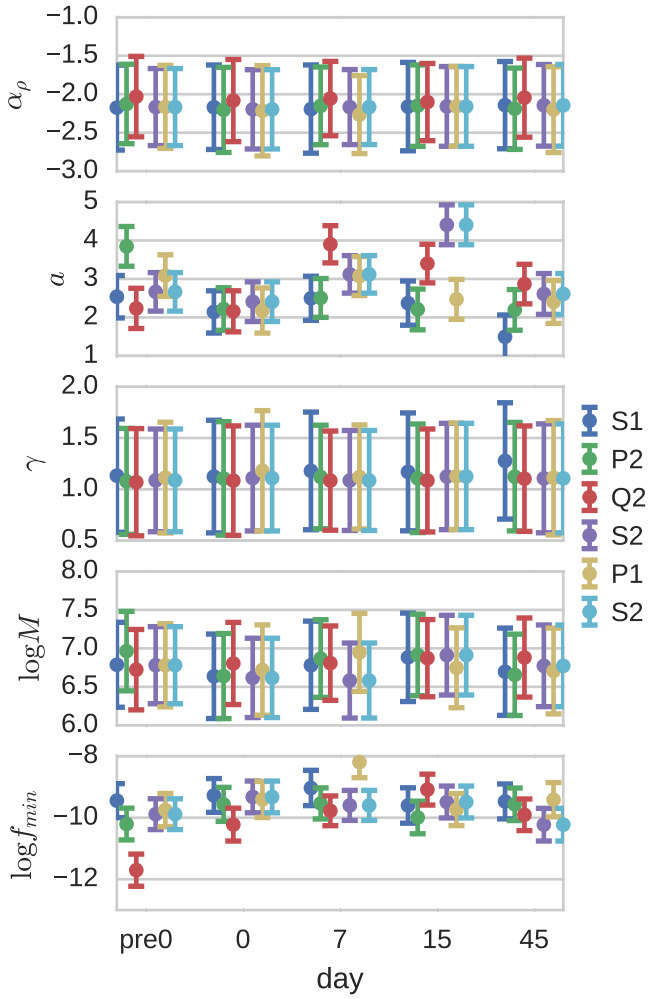

Figure 3. *Learned null model parameters.* Shown is the same data from fig. **??**, plotted separately for each donor and time point. Error bars are the inverse standard deviation of a Gaussian approximation around the maximum of the likelihood, i.e. the Cramer-Rao lower bound of the variance of the estimator.

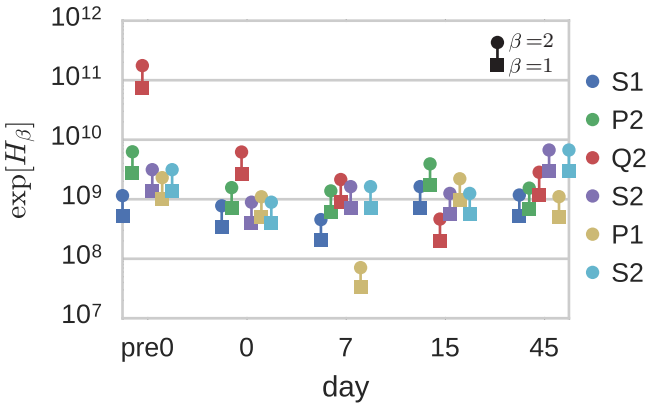

Figure 4. *Diversity estimates.* Shown are diversity estimates obtained from the Renyi entropies, H_β, of the inferred clone frequency distributions for $\beta = 1$ (Shannon entropy) and $\beta = 2$ (Simpson index), across donors and days.