Supplementary Material

Globular cluster simulations

We evolve a large number of globular cluster models using `cBHBd`, which includes realistic BH initial mass functions [62] and a model for the cluster formation rate and evolution through redshift [89]. We consider 25 different values of metallicity between $0.01Z_\odot$ and $Z_\odot$. These models follow the evolution of clusters and the formation of BH binary mergers through 3-body interactions. The BHs are paired using the paring functions derived in [22] using Heggie's theoretical interaction rate formulae [58]. In all models considered here we adopt the delayed supernova mechanism [62] and an initial cluster half-mass density $\rho_{\rm h} = 10^5 M_\odot {\rm pc}^3$. The densities of Milky Way globular clusters reach up to approximately an order of magnitude lower than we have assumed [90]. However, it is important to note that the initial densities of these clusters were higher than their current densities. This aligns with our models, which evolve to significantly lower densities by the end of the simulations. Similarly, some nuclear clusters exhibit present-day densities near and above the assumed value [91]. When generating our model predictions, we take in account the uncertainties in the initial cluster mass function and in the number density of clusters. This results in theoretical error bars in the merger rate estimates as in [8].

We report the results from four models that differ by the choice of the initial BH spin distribution. In one model the initial BH spins are all set to zero. Moreover, we do not include any prescription for pair-instability processes. Nevertheless, we find that no 1G BH is formed with a mass above $m_{\rm cut} \simeq 70M_\odot$, implying a mass truncation set by the initial upper limit on the stellar mass function (here at $130M_\odot$) and by stellar wind mass-loss prior to the formation of the BHs. In a second model, we assume that the initial BH spin distribution follows a beta distribution with $\alpha = 2$ and $\beta = 18$, which peaks at $a \simeq 0.06$ and has a median at $\simeq 0.1$. In a third model, the initial BH spin distribution follows a beta distribution with $\alpha = 2$ and $\beta = 5$, which peaks at $a \simeq 0.2$ and has a median at $\simeq 0.26$. In a fourth model, we use a beta distribution with $\alpha = 2$ and $\beta = 5$, which peaks at 0.5 and has a median at 0.5. In these latter three models, $m_{\rm cut} \simeq 51M_\odot$, as determined by the adopted prescriptions for pair instability [taken from 92]. All models result in a sharp truncation of the initial BH mass function at $m_{\rm cut}$. We note that certain high-mass star properties not included in our models, such as rotation or chemical composition, could result in the formation of BHs exceeding this limit. Therefore, the existence of a strict mass cut-off should be regarded as an assumption.

In Fig. S1 we show the differential merger rate of BH binaries as a function of $\chi_{\rm eff}$, $\chi_{\rm p}$ and m_1. We separately plot the populations of 1G+1G mergers, 1G+2G mergers, 2G+2G mergers, and the remaining mergers that involve higher generation BHs, the majority of which are 3G+1G and 3G+2G mergers. The figure shows that the simulated populations cannot explain the peak at $10M_\odot$ inferred from the GW data under the PowerLaw + Peak model from GWTC-3 [6], but can account for all mergers with primary mass $m_1 \gtrsim 30M_\odot$. The lower edge of the mass gap can be identified as the upper edge of the m_1 distribution where the 1G+1G merger rate drops to zero, and naturally coincides with the initial upper limit of the initial BH mass function that is set by our stellar evolution recipes. In all models the merger rate above this limit is dominated by 1G+2G mergers. As expected based on our theoretical considerations, the left panel of Fig. S1 shows that their $\chi_{\rm eff}$ distribution is nearly uniform (i.e., flat) within $|\chi_{\rm eff}| < 0.5$; the distribution of $\chi_{\rm p}$ is peaked at $\simeq 0.7$; and as the 1G BH spin increases, the overall contribution of 1G+2G binaries to the merger rate decreases. Naturally, 1G+2G as well as 2G+2G mergers have a null contribution to the merger rate above $m_1 > 2m_{\rm cut}$. In this range of masses, the merging binaries are 3G+1G, 3G+2G and 4G+1G.

We note that Ref. [31] used a phenomenological population model and the data from the second LIGO–Virgo Gravitational-Wave Transient Catalog to infer the median relative merger rates of 1G+2G and 2G+2G to 1G+1G mergers to be $\sim 10^{-2}$ and $\sim 10^{-5}$. These numbers are broadly consistent with the number ratio between the two populations found in our models.

Fig. S2 shows $\chi_{\rm eff}$ vs $\chi_{\rm p}$ for a fraction of the binaries that merge in each model. The mean correlation between these two parameters is well described by our simple model. The model with $(\alpha, \beta) = (2, 2)$ demonstrates significant overlap in the spin parameter distributions for 1G+1G and 1G+2G mergers, while still showing a substantial merger rate contribution from 1G+2G mergers. Thus, a confidently measurable value of $\chi_{\rm p}$ alone cannot be used to identify a single detected binary as hierarchically formed if BH natal spins are high. On the other hand, the population properties can be clearly used to put a hierarchical formation scenario to the test, as we do in this work.

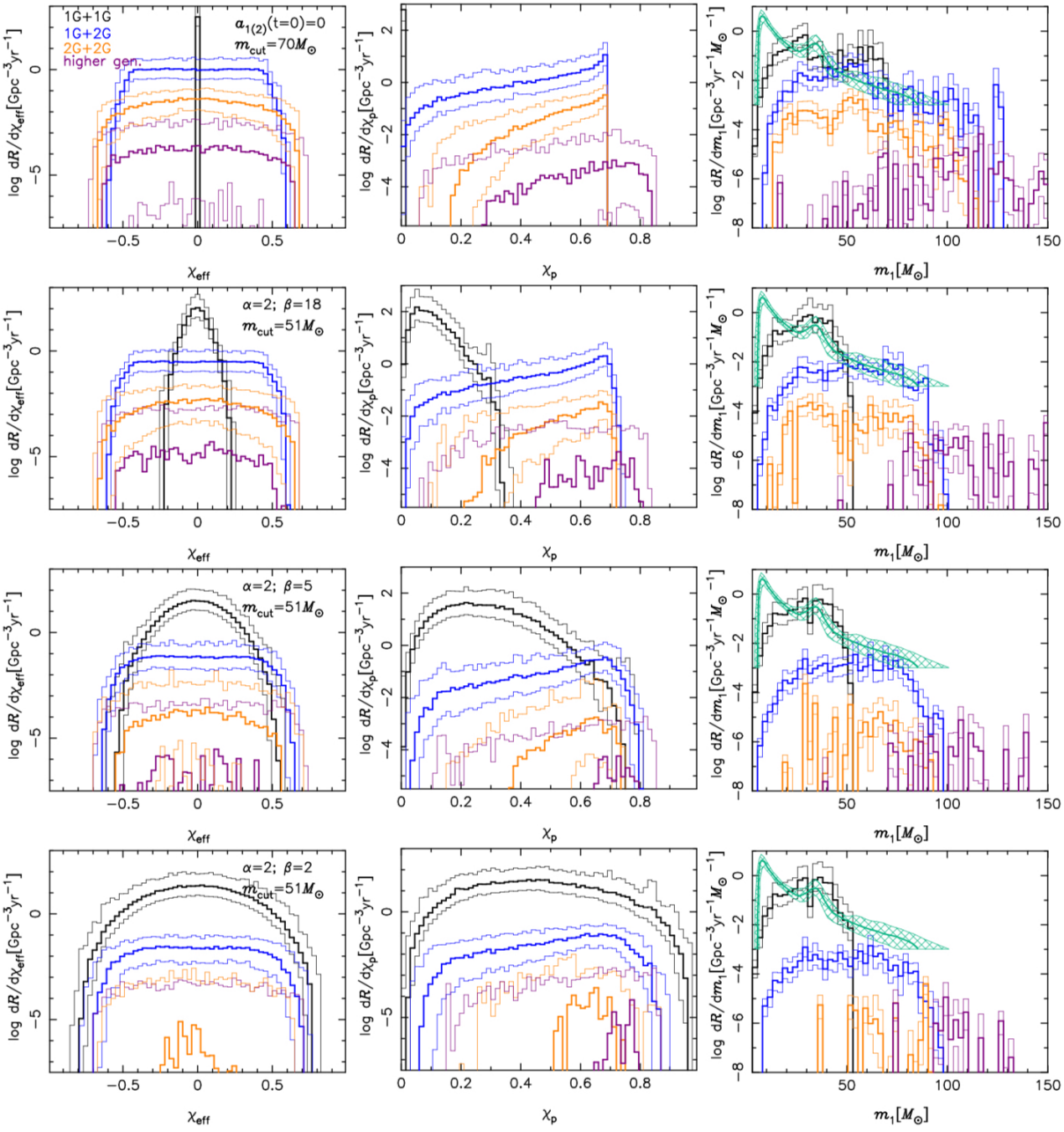

FIG. S1. Differential merger rate of BH binaries formed in globular cluster simulations as a function of binary effective spin parameters $\chi_{\rm eff}$ (left panels), $\chi_{\rm p}$ (central panels) , and primary BH mass m_1 (right panels). The upper panels show the results for a model where the first generation BHs are formed with no spin, while in the other models the BH birth spins follow a beta distribution with the α and β parameters as shown. The plot gives the median value of the merger rate in each bin, together with the 95 and 5 percentiles as thinner solid lines. These errors reflect uncertainties in the initial cluster mass function and number density of clusters. In the right panel, the green curve and hatched regions show the mean merger rate as inferred from the GW data and the corresponding confidence intervals as reported by [6] under the PowerLaw + Peak model.

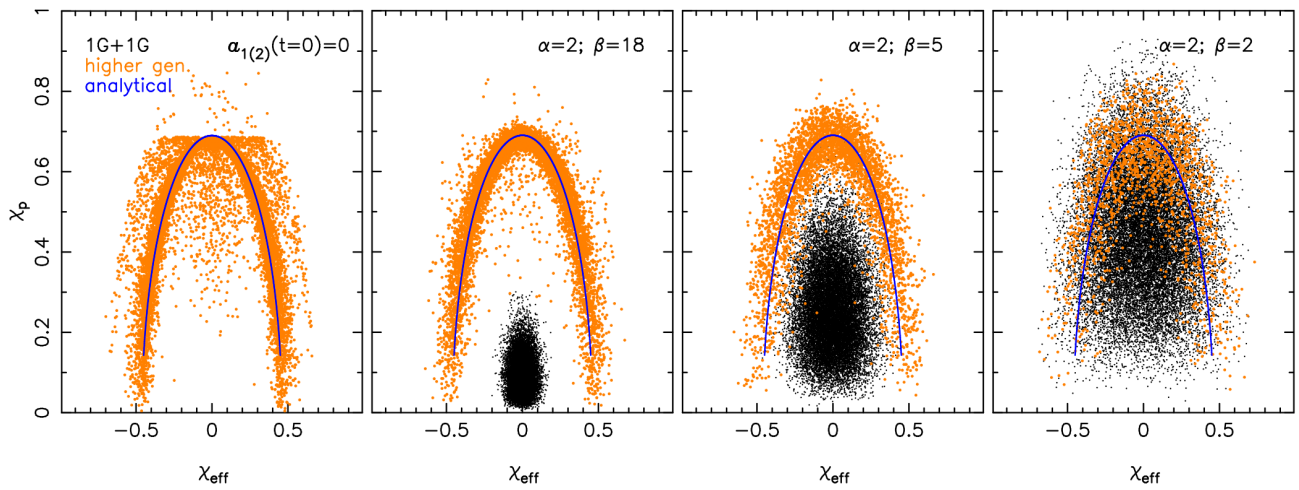

FIG. S2. Values of $\chi_{\rm eff}$ and $\chi_{\rm p}$ for a fraction of randomly selected binary BH mergers formed in our cluster models. The blue line shows the mean correlation predicted by our simple analytical model.

Parameter	Prior	$\mathcal{N}+\mathcal{U}$	$\mathcal{N}+\mathcal{U}_w$	$\mathcal{N}+\mathcal{N}\mathcal{U}_w$	$\mathcal{N}_m+\mathcal{U}_w$	$\chi_{\rm p}$
$\tilde{m}$	$\mathcal{U}(20,100)$	✓	✓	✓	✓	✓
w	$\mathcal{U}(0,1)$	–	✓	✓	✓	✓
μ	$\mathcal{U}(-1,1)$	✓	✓	✓	✓	✓
ζ	$\mathcal{U}(0,1)$	–	–	✓	–	–
$\mu_{\rm u}$	$\mathcal{U}(-1,1)$	–	–	✓	–	–
$\log_{10}\sigma$	$\mathcal{U}(-1.5,0)$	✓	✓	✓	✓	✓
$\log_{10}\sigma_{\rm u}$	$\mathcal{U}(-1.5,0)$	–	–	✓	–	–
$\delta\mu$	$\mathcal{U}(-2.5,1)$	–	–	–	✓	–
$\delta\log\sigma$	$\mathcal{U}(-2,1.5)$	–	–	–	✓	–
$\mu_{\rm p,u}$ and $\mu_{\rm p,l}$	$\mathcal{U}(0.02,1)$	–	–	–	–	✓
$\log_{10}\sigma_{\rm p,u}$ and $\log_{10}\sigma_{\rm p,l}$	$\mathcal{U}(-1.2,0)$	–	–	–	–	✓

TABLE S1. Priors adopted on the hyperparameters describing the effective spin distribution of binary BHs. For each parameter, we additionally indicate the model(s) in which the parameter appears.

Parameter	Prior	Defined in
β_q	$\mathcal{N}(0,3)$	equation (S7)
κ	$\mathcal{N}(0,6)$	equation (S8)
μ_m	$\mathcal{U}(50\,M_\odot, 100\,M_\odot)$	equation (S4)
σ_m	$\mathcal{U}(2\,M_\odot, 15\,M_\odot)$	equation (S4)
f_p	$\mathcal{U}(0,1)$	equation (S4)
λ	$\mathcal{N}(-2,3)$	equation (S4)
$m_{\rm max}$	$\mathcal{U}(60\,M_\odot, 100\,M_\odot)$	equation (S4)
$m_{\rm min}$	$\mathcal{U}(5\,M_\odot, 15\,M_\odot)$	equation (S4)
$\log_{10} dm_{\rm max}/\,M_\odot$	$\mathcal{U}(0.5,1.5)$	equation (S4)
$\log_{10} dm_{\rm min}/\,M_\odot$	$\mathcal{U}(-1,1)$	equation (S4)

TABLE S2. Priors adopted for the hyperparameters with which we describe the primary mass, mass ratio, and redshift distributions of the binary BH population.

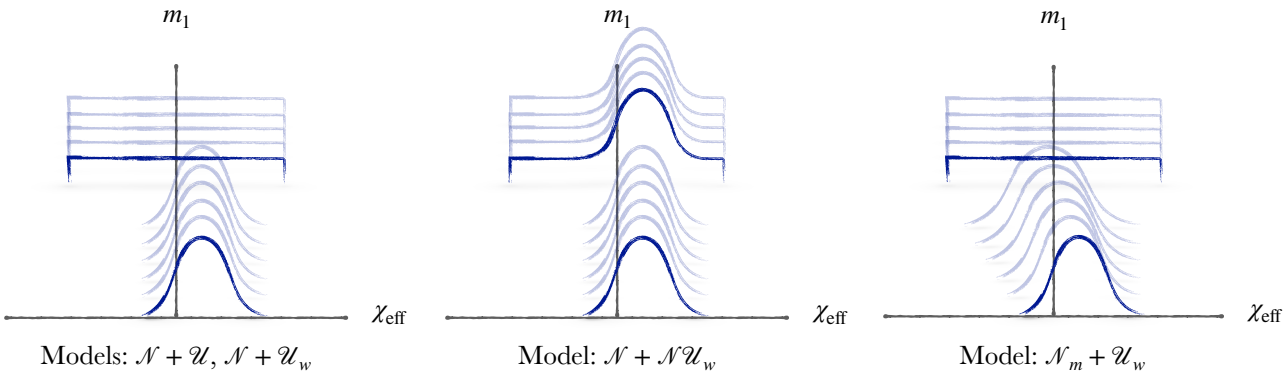

FIG. S3. Cartoon illustrating the main models examined in the main text. Blue lines show a representation of the parametric models used in our inference analysis.

Population models, data and hierarchical inference

We perform hierarchical Bayesian inferences to fit the data of the observed events with a given population model. Here, we summarise the main ingredients of the models, the data we use, and our inference analysis method.

As described in the main text, we investigate several different models for the $\chi_{\rm eff}$ distribution of binary BHs as a function of their m_1. These are sketched in Fig. S3. Illustrated in the left-hand subplot, models $\pi_{\mathcal{N}+\mathcal{U}}$ and $\pi_{\mathcal{N}+\mathcal{U}_w}$ (see equation (5)) assume that the effective spin distributions transitions from a Gaussian to a flat uniform distribution above some threshold mass $\tilde{m}$. Shown in the middle subplot, model $\pi_{\mathcal{N}+\mathcal{N}\mathcal{U}_w}$ (see equation (6)) instead assumes that the Gaussian spin distribution persists across both low- and high-mass systems, but includes a uniform component above $\tilde{m}$. Finally, in the right-hand panel, model $\pi_{\mathcal{N}_m+\mathcal{U}_w}$ (see equation (7)) adopts a low-mass Gaussian whose mean and log-standard deviation vary linearly as a function of mass. The priors on the hyperparameters defining each model are listed in Table S1.

We perform hierarchical inference on the binary BH population using the Hamiltonian Monte Carlo sampling algorithm implemented `numpyro`, a probabilistic programming library based on `jax`. Hamiltonian Monte Carlo methods require a likelihood that is a differentiable function of the population hyperparameters. However, the piecewise equations that define our spin models (e.g. equation (5)) are not differentiable; the discontinuities at $m_1 = \tilde{m}$ cause the likelihood to itself discontinuously as a function of $\tilde{m}$. To remedy this, in practice we implement these piecewise models as sharp but continuous transitions in the effective spin distributions above and below $\tilde{m}$. Our baseline model $\pi_{\mathcal{N}+\mathcal{U}}$ defined in equation (5), for example, is approximated as

$$\pi_{\mathcal{N}+\mathcal{U}}(\chi_{\rm eff}|m_1) = [1-\eta(m_1)]\,\mathcal{N}(\chi_{\rm eff};\mu,\sigma) + \eta(m_1)\,\mathcal{U}(\chi_{\rm eff};w=0.47), \tag{S1}$$

where

$$\eta(m_1) = \left[1+\exp\left(-\frac{(m_1-\tilde{m})}{3\,M_\odot}\right)\right]^{-1} \tag{S2}$$

is a logistic function that is approximately equal to zero below $\tilde{m}$ and unity above. Other spin models are implemented analogously. Similarly, a true truncated uniform distribution at high masses would cause the likelihood to change discontinuously with w. When implementing $\mathcal{U}(\chi_{\rm eff};w)$ we therefore do not use an infinitely sharp truncation at the boundaries but instead exponentially suppress the distribution beyond $|\chi_{\rm eff}| = w$:

$$\mathcal{U}(\chi_{\rm eff};w) \propto \begin{cases} 1 & (|\chi_{\rm eff}| \leq w) \\ \exp\left[-\frac{(|\chi_{\rm eff}|-w)^2}{2(0.1)^2}\right] & (|\chi_{\rm eff}| > w). \end{cases} \tag{S3}$$

Our conclusions do not depend on the precise scales over which the above smoothing takes place.

Alongside the effective spin distribution, we hierarchically measure the distribution of binary BH primary masses m_1, mass ratios q, and redshifts z. We model the m_1 distribution as a mixture between a power law and a Gaussian,

with exponential tapering functions at high and low masses:

$$p(m_1) \propto T_{\rm l}(m_1) T_{\rm h}(m_1) \Big[(1-f_p) P(m_1) + f_p \mathcal{N}(m_1) \Big]. \tag{S4}$$

Here, $P(m_1) \propto m_1^{\lambda}$ and $\mathcal{N}(m_1) \propto \exp\left[-\frac{(m_1-\mu_m)^2}{2\sigma_m^2}\right]$ are power-law and Gaussian distributions, each normalized over the interval $2\,M_\odot \leq m_1 \leq 100\,M_\odot$. The tapering functions are defined as

$$T_{\rm l}(m_1) = \begin{cases} \exp\left[-\frac{(m_1-m_{\rm min})^2}{2dm_{\rm min}^2}\right] & (m_1 < m_{\rm min}) \\ 1 & (m_1 \geq m_{\rm min}) \end{cases} \tag{S5}$$

and

$$T_{\rm h}(m_1) = \begin{cases} 1 & (m_1 \leq m_{\rm max}) \\ \exp\left[-\frac{(m_1-m_{\rm max})^2}{2dm_{\rm max}^2}\right] & (m_1 > m_{\rm max}) \end{cases} \tag{S6}$$

We assume that the secondary mass m_2 distribution follows [35]

$$p(m_2|m_1) \propto m_2^{\beta_q} \quad (2\,M_\odot \leq m_2 \leq m_1)\ . \tag{S7}$$

Finally, we assume a distribution of z that is proportional to the differential comoving volume dV_c/dz, with a possible evolution in the merger rate towards higher z [93, 94]

$$p(z) \propto \frac{1}{1+z} \frac{dV_c}{dz} (1+z)^{\kappa}\ . \tag{S8}$$

The priors placed on the hyperparameters governing the m_1, q, and z distributions are listed in Table S2.

We perform our inference using the subset of binary BH events from GWTC-3 with false alarm rates below $1\,{\rm yr}^{-1}$. We exclude GW170817, GW190425, GW190426, GW190814, GW190917, GW200105, GW200115 [7, 95], as they have at least one component with a mass $< 3\,M_\odot$ that are most likely neutron stars [6]. This leaves a total of $N_{\rm det} = 69$ binary black holes in our sample. We use parameter estimation samples made publicly available through the Gravitational-Wave Open Science Center. For events first published in GWTC-1 [5], we use the "`Overall_posterior`" parameter estimation samples. For events first published in GWTC-2 [95] and GWTC-2.1 [96], we adopt the "`PrecessingSpinIMRHM`" samples, and for events in GWTC-3 [7], we use the "`C01:Mixed`" samples [97]. These selections correspond to a union of samples obtained with different waveform families. All samples account for spin precession effects, while the `PrecessingSpinIMRHM` and `C01:Mixed` samples from GWTC-2, GWTC-2.1, and GWTC-3 additionally include the effects of higher order modes (parameter estimation incorporating higher order modes was not available in GWTC-1). We assess the detection efficiency using the set of successfully recovered binary BH injections, provided by the LIGO-Virgo-KAGRA collaborations, spanning their first three observing runs [63].

We perform our analysis using standard hierarchical inference. Let $p(\theta_i|d_i)$ be posteriors on the individual parameters θ_i (e.g. component masses, redshift, etc.) of each gravitational-wave event, conditioned on its observed data d_i. The corresponding posterior on the population parameters Λ is [e.g., 35, 93, 98]

$$p(\Lambda\,|\,\{d_i\}) \propto p(\Lambda)\,\xi^{-N_{\rm det}}(\Lambda) \prod_{i=1}^{N_{\rm det}} \int d\theta_i\, p(\theta_i|d_i) \frac{p(\theta_i|\Lambda)}{p_{\rm pe}(\theta_i)}\ , \tag{S9}$$

where $p_{\rm pe}(\theta_i)$ is the prior adopted for purposes of parameter estimation and $p(\Lambda)$ is our prior on the population-level parameters. We use the priors listed in Tables S1 and S2. We replace integration over $p(\theta_i|d_i)$ with an ensemble average taken over the posterior samples associated with each event:

$$p(\Lambda\,|\,\{d_i\}) \propto p(\Lambda)\,\xi^{-N_{\rm det}}(\Lambda) \prod_{i=1}^{N_{\rm det}} \left\langle \frac{p(\theta_i|\Lambda)}{p_{\rm pe}(\theta_i)} \right\rangle. \tag{S10}$$

Given a number $N_{\rm inj}$ of injected signals drawn from some reference distribution $p_{\rm inj}(\theta_{\rm i})$ The detection efficiency $\xi(\Lambda)$ quantifies the total fraction of events that we expect to pass our detection criteria:

$$\xi(\Lambda) = \frac{1}{N_{\rm inj}} \sum_{i=1}^{N_{\rm found}} \frac{p(\theta_i|\Lambda)}{p_{\rm inj}(\theta_{\rm i})}, \tag{S11}$$

summing over the $N_{\rm found}$ injections that pass our detection criteria, and where $N_{\rm inj}$ is the total number of injections (including those that are not recovered). We estimate $\xi(\Lambda)$ using the injection campaign reported in [63, 96], selecting successfully found injections (with recovered false alarm rates below 1 yr^{-1} in at least one pipeline) and reweighting to the proposed population Λ as in [35]. We sample over equation (S10) using the NUMPYRO's [86, 87] implementation of the "No U-Turn" Hamiltonian Monte Carlo algorithm [99].

The full posterior distributions on the key parameters that enter in our population analysis are given in Figs. S4, S5, S6, and S7. In particular, we note that although parameter ζ in $\pi_{\mathcal{N}+\mathcal{N}\mathcal{U}_w}$ favors a value of zero (i.e. a full transition to a uniform $\chi_{\rm eff}$ distribution above $\tilde{m}$), it has large uncertainties, being constrained to lie in the range of 0.02–0.82 in 90% credible intervals. This might initially appear to reduce the robustness of our conclusions, implying that the data are unable to distinguish with confidence a uniform from a Gaussian distribution above $\tilde{m}$. This is not surprising, however, and suggests that our inference is not particularly sensitive to how the high mass population is parametrized. In fact, we consider a model where the high mass population is represented by a Gaussian (by setting $\zeta = 1$), and find a consistent distribution of $\chi_{\rm eff}$ and $\tilde{m} = 45^{+5}_{-4} M_\odot$. Finally, we note that in the $\pi_{\mathcal{N}+\mathcal{N}\mathcal{U}_w}$ model the parameters of the Gaussian are essentially unconstrained, which means that this distribution is probably not required.

In the main text, we quote Bayes factors between our models and one in which the binary BH effective spin distribution is modeled as a single Gaussian, $\pi_{\mathcal{N}}$, with no distinct high-mass population. Bayes factors are obtained by sampling mixture models of the form $\pi_{\rm mix}(\chi_{\rm eff}|m_1) = \xi\,\pi_{\mathcal{A}}(\chi_{\rm eff}|m_1) + (1-\xi)\,\pi_{\mathcal{N}}(\chi_{\rm eff}|m_1)$, where $\pi_{\mathcal{A}}(\chi_{\rm eff}|m_1)$ is one of our models under investigation. The quantity ξ is not continuous, but is a *categorical* variable that can be either 0 or 1, corresponding to populations described entirely by $\pi_{\mathcal{N}}$ or $\pi_{\mathcal{A}}$, respectively. We sample this likelihood using the `DiscreteHMCGibbs` sampler implemented in `numpyro`; this performs Hamiltonian Monte Carlo over the continuous parameters defining models $\pi_{\mathcal{N}}$ and $\pi_{\mathcal{A}}$ while sampling the categorical variable ξ using Metropolis-Hastings. The Bayes factor between models $\pi_{\mathcal{A}}$ and $\pi_{\mathcal{N}}$ is then given by the ratio of posterior probabilities $p(\xi = 1|\{d_i\})/p(\xi = 0|\{d_i\})$.

We also perform several leave-one-out analyses to determine whether our results are driven primarily by a small number of unusual events. In particular, the events GW170729 and GW190517 could conceivably be driving the preference for a broad spin distribution at high masses; both events have confidently large spins ($\chi_{\rm eff} = 0.36^{+0.21}_{-0.25}$ and $0.52^{+0.20}_{-0.19}$, respectively, under default priors) and have primary masses (50^{+16}_{-10} and $36^{+12}_{-8}\,M_\odot$) near our inferred values of w and $\tilde{m}$. We repeat our analysis under the $\pi_{\mathcal{N}+\mathcal{U}_w}$ model excluding one or both of GW170729 and GW190517. In all cases, results remain consistent with those shown in Fig. 2, increasing our confidence that we are identifying a feature inherent in the broader black hole population.

Finally, in order to identify possible issues due to finite sampling effects when estimating Eq. S10, we track the number of "effective samples" $N_{\rm eff}$ informing the Monte Carlo estimates of the likelihood for every event. Given a set of N_i posterior samples $\{\lambda_{i,j}\}_{j=1}^{N_i}$ for each event i, the $N_{{\rm eff},i}$ under a proposed population Λ is

$$N_{{\rm eff},i}(\Lambda) \equiv \frac{\left[\sum_{j=1}^{N_i} w_{i,j}(\Lambda)\right]^2}{\sum_{j=1}^{N_i} \left[w_{i,j}(\Lambda)\right]^2}\,, \tag{S12}$$

where $w_{i,j}(\Lambda) = p(\lambda_{i,j}|\Lambda)/p_{\rm pe}(\lambda_{i,j})$. Small $N_{{\rm eff},i}(\Lambda) \lesssim 10$ indicates that the given event is sparsely sampled, and hence the likelihood may be dominated by sampling variance [100]. In this case, we should not necessarily trust the results of Monte-Carlo-based hierarchical inference. We do not impose any cut on $N_{{\rm eff},i}(\Lambda)$. Instead, we compute and track $\mathcal{N} \equiv \min \log_{10}\left[N_{{\rm eff},i}(\Lambda)\right]$ for each model we consider. In all cases we find that effective sample counts are large, with $\min\left[N_{\rm eff}\right] > 10$ – for $\pi_{\mathcal{N}+\mathcal{U}}$, $\pi_{\mathcal{N}+\mathcal{U}_w}$, $\pi_{\mathcal{N}+\mathcal{N}\mathcal{U}_w}$, and $\pi_{\mathcal{N}_m+\mathcal{U}_w}$ we find $\mathcal{N} = 1.9^{+0.5}_{-0.6}$, $1.8^{+0.5}_{-0.5}$, $1.8^{+0.5}_{-0.4}$, and $1.4^{+0.6}_{-0.4}$ respectively; computing $\min\left[N_{{\rm eff},i}(\Lambda)\right]$ for the high mass population only (i.e., setting $\eta(m_1) = 1$ in equation S12), we find $\mathcal{N} = 2.4^{+0.3}_{-0.4}$, $2.3^{+0.3}_{-0.3}$, $2.4^{+0.3}_{-0.4}$ and $2.0^{+0.3}_{-0.3}$. This gives us confidence that the Monte Carlo averages over posterior samples yields accurate results. The posteriors of $\mathcal{N}$ and $N_{\rm inj}$ for these models are shown in Figs. S4, S5, S6, and S7.

Particular care must be given when interpreting the results from the $\pi_{\chi_{\rm p}}$ model. In this model we find that setting a zero lower bound on the prior of $\mu_{\rm p,u(l)}$ results in an unacceptably low $\mathcal{N} \approx 0.1$ and in a very narrow Gaussian for the $\chi_{\rm p}$ distribution of the low-mass population. To unsure a sufficiently large sampling size and to avoid singularities, we follow Ref. [101, 102] and exclude zero from the prior range of $\mu_{\rm p,u(l)}$ as reported in Table S1, but we caution that the inference of $\chi_{\rm p}$ may be subject to increased Monte Carlo averaging error and the assumed lower bounds for the parameter priors. We note that all other results (in particular the $\chi_{\rm eff}$ and $\tilde{m}$ results on which our paper is focused) are unchanged, and similar sampling issues did not present for other parameters. The posteriors for a selections of parameters for the $\pi_{\chi_{\rm p}}$ model are shown in Fig. S8.

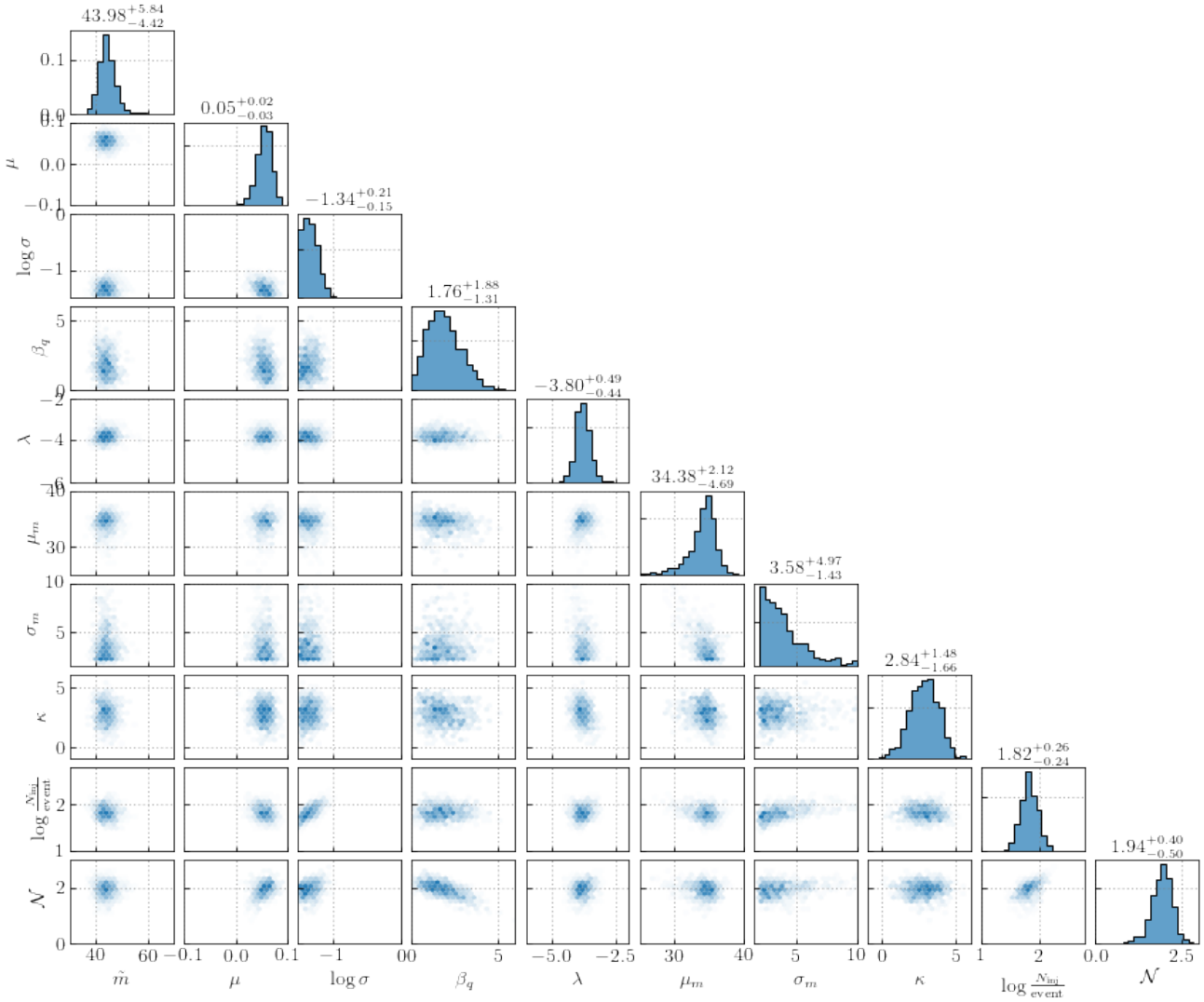

FIG. S4. Posteriors on the parameters that govern the hierarchical model where the $\chi_{\rm eff}$ distribution is given by equation (5): a Gaussian below $\tilde{m}$ and a Uniform distribution with half width $w = 0.47$ above $\tilde{m}$. Here log indicates $\log_{10}$ and $N_{\rm inj}/event$ gives the total number of injection divided by the number of detections.

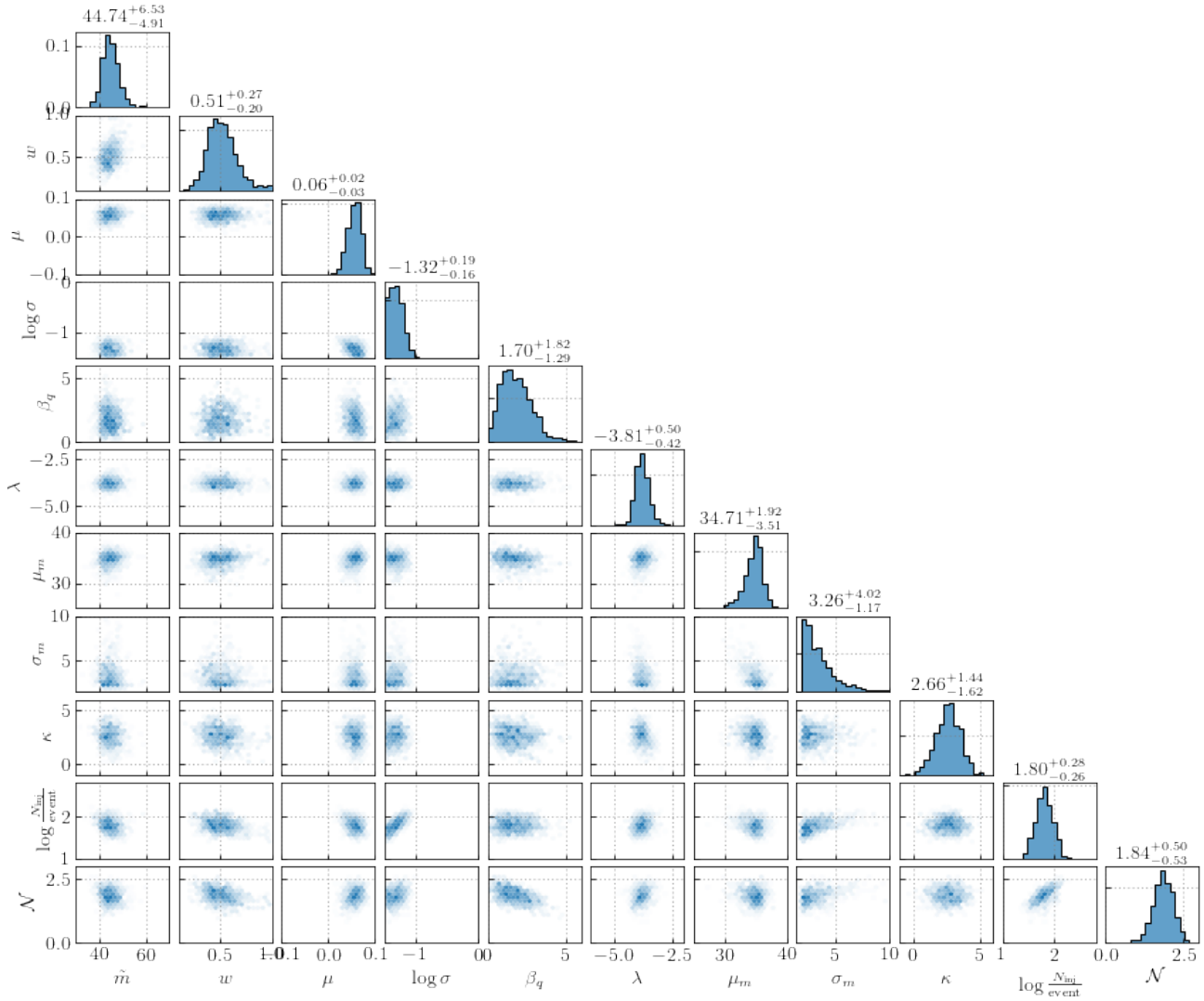

FIG. S5. Posteriors on the parameters that govern the hierarchical model where the $\chi_{\rm eff}$ distribution is represented by a fixed Gaussian below $\tilde{m}$ and a Uniform distribution with variable half width, w, above $\tilde{m}$. Here log indicates $\log_{10}$ and $N_{\rm inj}/event$ gives the total number of injection divided by the number of detections.

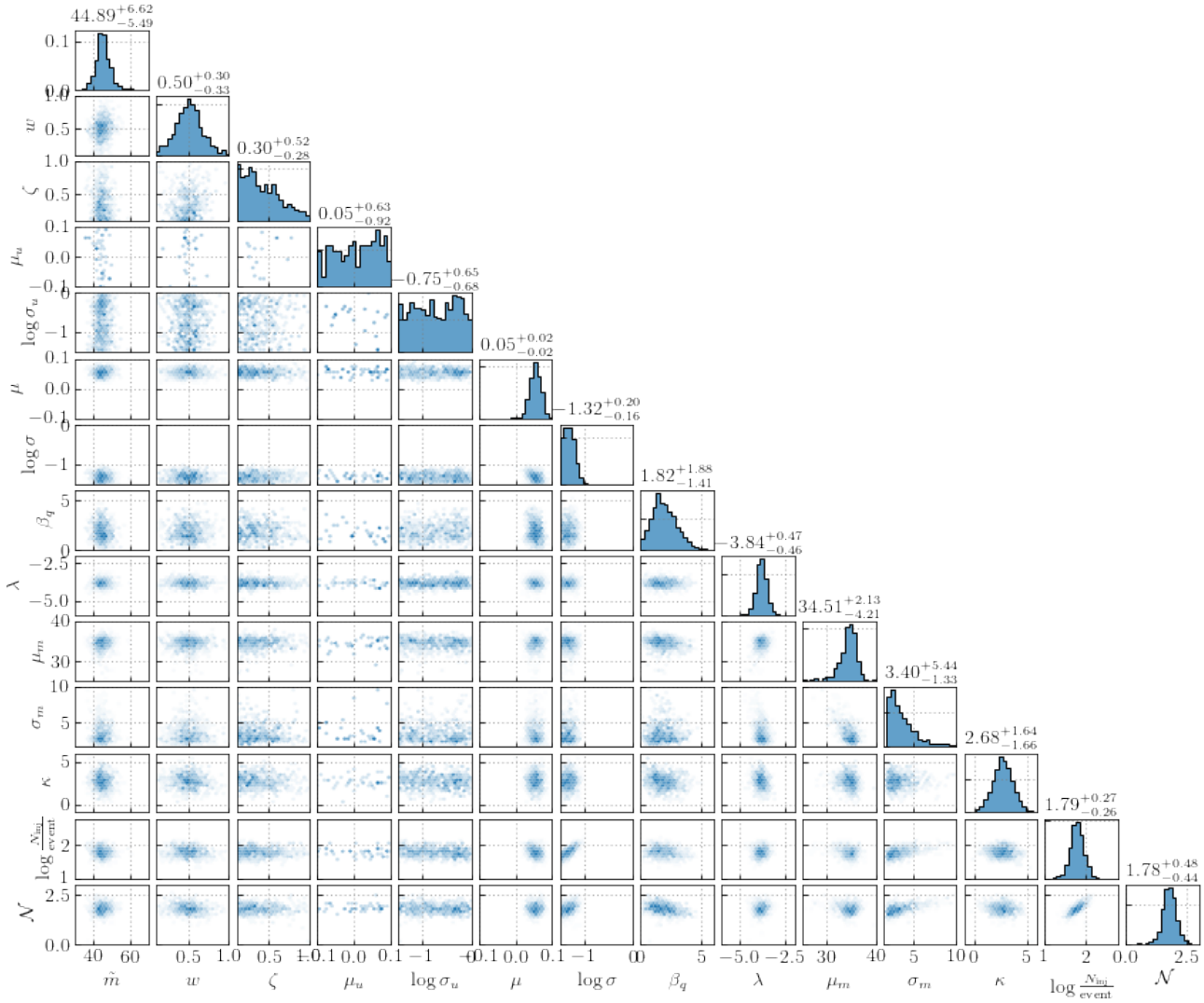

FIG. S6. Posteriors on the parameters that govern the hierarchical model where the $\chi_{\rm eff}$ distribution is represented by equation (6): a fixed Gaussian below $\tilde{m}$ and a Uniform distribution with variable half width, w, plus a Gaussian above $\tilde{m}$. Here log indicates $\log_{10}$ and $N_{\rm inj}/event$ gives the total number of injection divided by the number of detections.

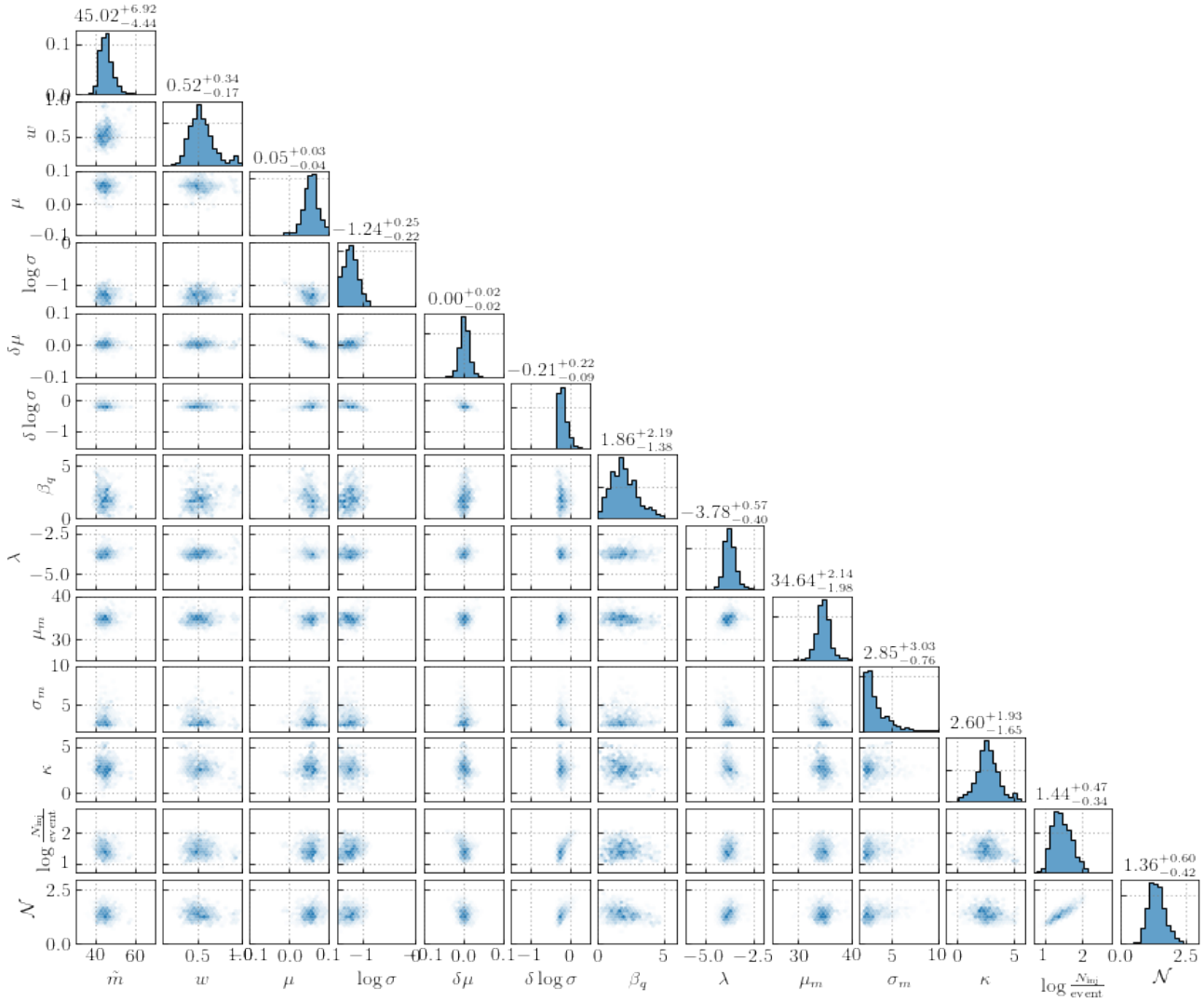

FIG. S7. Posteriors on the parameters that govern the hierarchical model where the $\chi_{\rm eff}$ distribution is represented by equation (7): a Gaussian with mass dependent mean and variance below $\tilde{m}$ and a fixed Uniform distribution with half width w above $\tilde{m}$. Here log indicates $\log_{10}$ and $N_{\rm inj}/event$ gives the total number of injection divided by the number of detections.

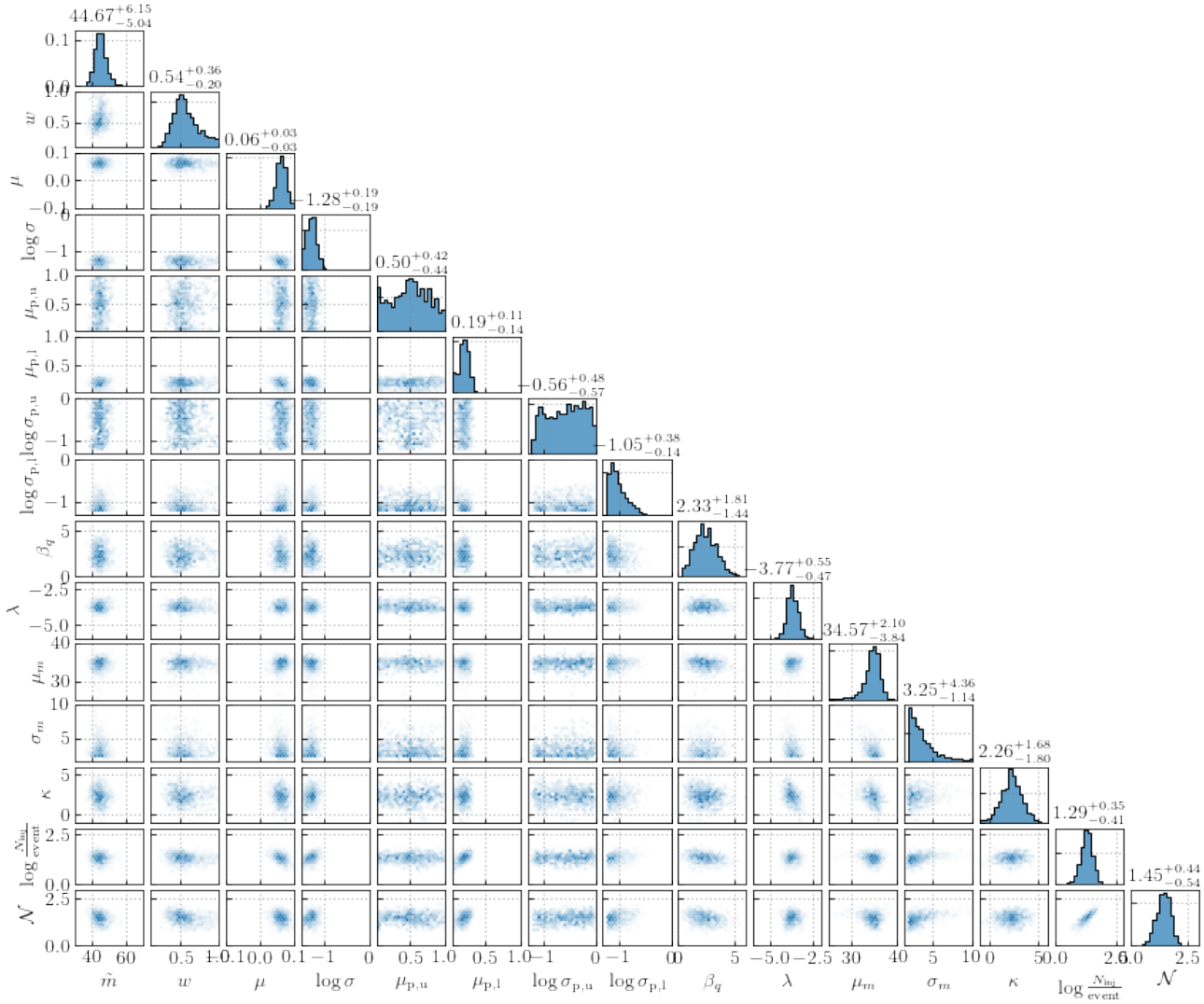

FIG. S8. Posteriors on the parameters that govern the hierarchical model where the χ_{eff} and χ_{p} distributions are represented using equation 8. Here log indicates $\log_{10}$ and $N_{\mathrm{inj}}/event$ gives the total number of injection divided by the number of detections.