Star Formation Sequence in a Hierarchical Universe

ChangHoon Hahn[1,2], Jeremy L. Tinker[2], Andrew R. Wetzel[3,4,5]

changhoon.hahn@lbl.gov

DRAFT --- c4cf42a --- 2018-08-20 --- NOT READY FOR DISTRIBUTION

ABSTRACT

motivation, methodology, impact. In observations star forming galaxies form a tight *log* M_* to *log* SFR relation referred to as the *star formation main sequence* (SFS) out to $z \sim 2$. Beyond the evolution "along" this SFS, however, the star formation histories of star forming galaxies have not been precisely characterized. The SFH of these galaxies govern SMF, SFS, and also observed constraints on the stellar mass to halo mass relation.

By combining high-resolution cosmological N-body simulation with observed evolutionary trends of SF galaxies, we construct a model that tracks the evolution of star forming central galaxies over the redshift $z < 1$. Comparing this model

Observations find a remarkably small scatter in the stellar mass to halo mass relation. Somehow the star formation histories of galaxies must

According to observations, star forming galaxies form a tight *log* M_* to *log* SFR relation referred to as the "star formation main sequence" out to $z \sim 2$.

Subject headings: methods: numerical – galaxies: clusters: general – galaxies: groups: general – galaxies: evolution – galaxies: haloes – galaxies: star formation – cosmology: observations.

1. Introduction

- Motivate why we think SF galaxies evolve along the main sequence
- Discuss the current thought process on galaxy assembly bias
- Explain the limitation of SFH derivable from observations (Claire's fisher matrix paper would be really good; ask her about the details)

[1]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720

[2]Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003

[3]TAPIR, California Institute of Technology, Pasadena, CA USA

[4]Carnegie Observatories, Pasadena, CA USA

[5]Department of Physics, University of California, Davis, CA USA

- in fact we can't constrain sf variability very well even in simulations due to the time resolution (see Hahn et al. (in prep.)).
- Observations also can't provide detail host dark matter halo properties
- So the approach with combining observations with N-body (empirical modeling) is very effective in the context of the halo.
- Maybe talk about how the bigger context of why this is important?
- Why only centrals – because our current best understanding of satellites is that they quench after infall, so it doesn't make sense to look at them
- our model goes from $z < 1$ because beyond that the observations are statistically meaningless.

2. Central Galaxies of SDSS DR7

We construct our galaxy sample following the sample selection of Tinker et al. (2011). We select a volume-limited sample of galaxies with $M_r 5log(h) < 18$ and complete in $M_* > 10^{9.4} M_\odot$ from the NYU Value-Added Galaxy Catalog (VAGC; Blanton et al. 2005) of the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; Abazajian et al. 2009) at $z \approx 0.04$. The stellar masses of these galaxies are estimated using the `kcorrect` code (Blanton & Roweis 2007) assuming a Chabrier (2003) initial mass function. The star formation of the galaxies are estimated spectroscopically using the specific star formation rates (SSFR) from the current release of the MPA-JHU spectral reductions[1] (Brinchmann et al. 2004). Generally speaking, SSFR $> 10^{-11}\mathrm{yr}^{-1}$ are derived from H_α emission, $10^{-11} >$ SSFR $> 10^{-12}\mathrm{yr}^{-1}$ are derived from a combination of emission lines, and SSFR $< 10^{-12}\mathrm{yr}^{-1}$ are based on D_n4000 (see discussion in Wetzel et al. 2013). We note that SSFR $< 10^{-12}\mathrm{yr}^{-1}$ should only be considered upper limits to the true galaxy SSFR (Salim et al. 2007).

From our galaxy sample, we identify the central galaxies using the Tinker et al. (2011) halo-based group-finding algorithm, which is based on the Yang et al. (2005) algorithm and tested in Campbell et al. (2015). The algorithm assigns a probability of being a satellite, $P_{\rm sat}$, to each galaxy in the sample. Galaxies with $P_{\rm sat} \geq 0.5$ are classified as satellites and $P_{\rm sat} < 0.5$ are classified as centrals. In this paper we focus on central galaxies. With any group finding algorithm, galaxies are misassigned due to projection effects and redshift space distortions. The purity of the full central galaxy sample is $\sim 90\%$ with a completeness of $\sim 95\%$ (Tinker et al. 2017). Furthermore, Campbell et al. (2015) find that the algorithm

[1] http://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

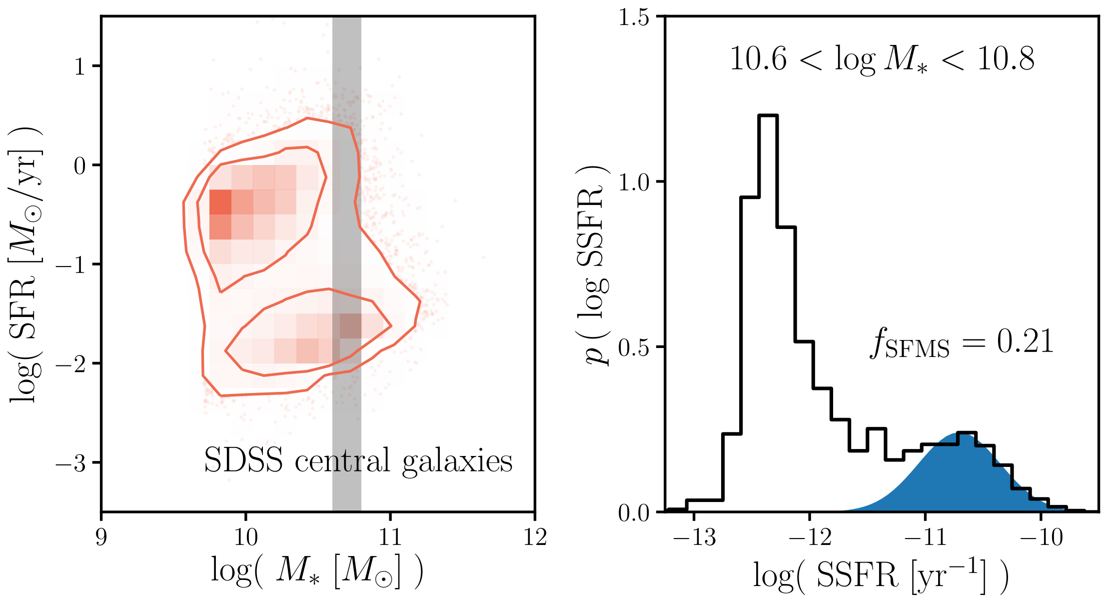

Fig. 1.— Central galaxies of the SDSS DR7 group catalog. *Left:* We plot the SFR-M_* relation of the SDSS central galaxies. The contours illustrate the bimodal distribution of the galaxy properties and mark the star-forming and quiescent populations. The transitioning galaxies lie on the "green" valley between the star-fomring and quiescent modes. *Right:* We plot the distribution of $\log \mathrm{SSFR}$ for SDSS centrals with $10.6 < \log M_* < 10.8$. Shaded in blue, we plot the SFS component of our GMM fit of the SFR-M_* relation described in Section 3.1. Based on this fit, galaxies in the SFS account for approximately $f_{\mathrm{SFS}} = 0.21$ of the central galaxies in the stellar mass bin.

robustly identifies red and blue centrals as a function of stellar mass, which is highly relevant to our analysis.

In the left panel of Figure 1, we plot the SFR-M_* distribution of the SDSS DR7 central galaxies. In the right panel, we plot the distribution of SSFR, $p(\log \mathrm{SSFR})$, for galaxies with $10.6 < \log M_* < 10.8$ (stellar mass range highlighted on the left panel). Both panels of Figure 1 illustrate the bimodality in the galaxy sample. The SFR-M_* distribution also illustrate the correlation between SFR and M_* in star-forming galaxies *i.e.* the star-formation main sequence (SFS).

3. Model: Simulated Central Galaxies

We're interesting in constructing a model that tracks central galaxies and their star formation within the heirarchical growth of their host halos. This requires a cosmological N-body simulation that accounts for the complex dynamical processes that govern the host halos of galaxies. In this paper we use the high resolution N-body simulation from Wetzel et al. (2013) generated using the White (2002) `TreePM` code with flat ΛCDM cosmology ($\Omega_m = 0.274, \Omega_b = 0.0457, h = 0.7, n = 0.95, \mathrm{and} \sigma_8 = 0.8$). From initial conditions at $z = 150$ generated from second-order Lagrangian Perturbation Theory, 2048^3 particles with mass of $1.98 \times 10^8\, M_\odot$ are evolved in a 250Mpc/h box with a Plummer equivalent smoothing of $2.5\,\mathrm{kpc}/h$. For a more detailed description of the simulation, we refer readers to Wetzel et al. (2013, 2014).

From the TreePM N-body simulation, 'host halos' are identified using the Friends-of-Friends (FoF) algorithm of Davis et al. (1985) with linking length of $b = 0.168$ times the mean inter-partcile spacing. Within these host halos, Wetzel et al. (2013) identifies 'subhalos' as overdensities in phase space through a six-dimensional FoF algorithm (FoF6D White et al. 2010). The host halos and subhalos are then tracked across the simulation outputs from $z = 10$ to 0 to build merger trees (Wetzel et al. 2009; Wetzel & White 2010). The most massive subhalos in newly-formed host halos at a given simulation output are defined as the 'central' subhalo. A central subhalo retains its 'central' definition until it falls into a more massive host halo, at which point it becomes a 'satellite' subhalo.

At a given snapshot, we assign stellar masses used only for initializing our model to each subhalo by subhalo abundance matching (SHAM; Conroy et al. 2006; Vale & Ostriker 2006; Yang et al. 2009; Wetzel et al. 2012; Leja et al. 2013; Wetzel et al. 2013, 2014; **?**) to M_peak, the maxmum host halo mass that it ever had as a central subhalo. SHAM, in its simplest form, assumes a one-to-one mapping between subhalo M_peak and galaxy stellar mass M_* that preserves rank order: $n(> M_\mathrm{peak}) > n(> M_*)$. In practice, we apply a 0.2 dex log-normal scatter in M at fixed M_peak based on the observed stellar to halo mass relation (SHMR; **bunch of SMHMR citations**). For $n(> M_*)$, we use observed stellar mass functions (SMFs) at the redshift corresponding to the snapshot. At $z = 0.05$, the lowest redshift snapshot of our model, we use the SMF from Li & White (2009), which is based on the same SDSS NYU-VAGC sample as our group catalog. At higher redshifts, we interpolate between the Li & White (2009) SMF and the SMF from Marchesini et al. (2009) at $z = 1.6$. We choose the Marchesini et al. (2009) SMF, among others, because it produces interpolated SMFs that monotonically increase at $z < 1$. As noted in **?**, at $z \approx 1$, the SMF interpolated between the Li & White (2009) and Marchesini et al. (2009) SMFs is consistent with more recent measurements from Muzzin et al. (2013) and Ilbert et al. (2013). **TBD: Perhaps mention in appendix how we test different SMF assumptions**

Throughout its 45 snapshot outs, `TreePM` simulation tracks the evolution of subhalos back to $z \sim 10$. We restrict ourselves to 15 snapshots from $z = 1.08$ to $z = 0.05$, where we have the most statistically meaningful observations. Furthermore, since we're interested in centrals we only keep subhalos that are classified as centrals throughout the redshift range. This criterion removes "black splash" or "ejected" satellite galaxies (*e.g.* Mamon et al. 2004; Wetzel et al. 2014) misclassified as centrals. Finally, we have a model based on the `TreePM` N-body simulation that tracks the evolution of central subhalos from $z = 1.08$ to $z = 0.05$. Next, we describe how we select and initialize the star forming central galaxies from the central subhalos in our model.

3.1. Selecting Star Forming Centrals

In our model, we're interested in tracking the SFR and stellar mass evolution of SF central galaxies. To construct such a model, we first need to select SF galaxies from the central galaxies in our simulation, described above. Since we want our model to reproduce observations, our selection is based on $f^{\rm cen}_{\rm SFS}(M_*)$, the fraction of central galaxies within the star forming sequence, measured from the SDSS DR7 VAGC (Section 2). Below, we describe how we derive this $f^{\rm cen}_{\rm SFS}(M_*)$ and use it to select SF central galaxies in our model. Afterwards we describe how we initalize the SFRs and M_* of these galaxies in our model.

Often in the literature, an empirical color-color or SFR–M_* cut that separates the two main modes (red/blue or star-forming/quiescent) in the distribution is chosen to classify galaxies (*e.g.* Baldry et al. 2006; Blanton & Moustakas 2009; Drory et al. 2009; Peng et al. 2010; Moustakas et al. 2013; Hahn et al. 2015). The red/quiescent or blue/star-forming fractions derived from this sort of classification, by construction, depend on the choice of cut and neglect galaxy subpopulations such as transitioning galaxies *i.e.* galaxies in the "green valley". Instead, for our $f^{\rm cen}_{\rm SFS}(M_*)$, we use the SFS fitting method from Hahn et al. (in prep.). Hahn et al. (in prep.)uses Gaussian Mixture Models and the Bayesian Information Criteria in order to fit the SFR–M_* relation of a galaxy population and identify its SFS. This data-driven approach relaxes many of the assumptions and hard cuts that go into other methods. Furthermore, as they demonstrate in Hahn et al. (in prep.)by applying to multiple simulations, it can be flexibly applied to a wide range of star formation and M_*. The weight of the SFS GMM component from the fitting provides an estimate of $f^{\rm cen}_{\rm SFS}$. In the right panel of Figure 1, we plot the SFS GMM component (blue shaded region) of the $p(\log \mathrm{SSFR})$ for the SDSS DR7 central galaxies within $10.6 < \log M_* < 10.8$. The SFS constitutes $f^{\rm cen}_{\rm SFS} = 0.21$ of the SDSS central galaxies in this stellar mass bin.

Rather than using the $f^{\rm cen}_{\rm SFS}$ values directly, for selecting SF galaxies, we fit $f^{\rm cen}_{\rm SFS}$ as a linear function of $\log M_*$ similar to Wetzel et al. (2013); Hahn et al. (2017a). We derive the following best-fit:

$$f^{\rm cen}_{\rm SFS,bestfit}(M_*) = -0.627\,(\log M_* - 10.5) + 0.354. \tag{1}$$

We note that this $f^{\rm cen}_{\rm SFS,bestfit}(M_*)$ is in good agreement with the $f^{\rm cen}_{\rm Q}(M_*; z \sim 0)$ fit from **?**. For each central galaxy in our simulation, we assign a probability of it being on the SFS, using Eq. 1 with M_* at $z \sim 0$ assigned through SHAM. Based on these probablities, we randomly identify centrals from our simulation as SF at $z \sim 0$. In our model, we make the assumption that once a SF galaxy quenches its star formation, it remains quiescent. Without any quiescent galaxies rejuvenating their star formation, galaxies on the SFS at $z \sim 0$ are also on the SFS at $z > 0$. Using this assumption the SF centrals we select at $z \sim 0$ are also on the SFS at the intial redshift of our model: $z \sim 1$.

Next, we can initialize the SF centrals at $z \sim 1$ using SHAM M_*s and assign their initial SFRs based on the observed SFR-M_* relation of the SFS. Observations in the literature at these redshifts, however, not only use galaxy properties derived differently from the SDSS VAGC but they also find SFS with significant discrepancies from one another. Speagle et al. (2014) compiles the SFR-M_* relation of the SFS from 25 studies in the literature, each with different methods of deriving galaxy properties. Even *after* their calibration, for a fixed $M_* = 10^{10.5}\,M_\odot$, the SFRs of the SFSs at $z \sim 1$ vary by more than a factor of 2 (see Figure 2 of Speagle et al. 2014). With little consensus on the SFS at $z \sim 1$, and consequently its redshift evolution, we flexibly parameterize the SFS SFR ($\log \mathrm{SFR}_{\rm MS}(M_*, z)$) with free parameters m_{M_*} and m_z that characterize the stellar mass and redshift dependences respectively. We parameterize the mean log SFR of the SFS as,

$$\log \overline{\mathrm{SFR}}_{\rm SFS}(M_*, z) = m_{M_*} * (\log M_* - 10.5) + m_z * (z - 0.05) - 0.11. \tag{2}$$

We assign SFRs to our SF centrals at $z \sim 1$ by sampling a log-normal distribution centered about $\log \overline{\mathrm{SFR}}_{\rm MS}(M_*, z = 1)$ with a constant scatter of 0.3 dex, motivated from observations (Daddi et al. 2007; Noeske et al. 2007; Magdis et al. 2012; Whitaker et al. 2012). Later in our analysis, for the priors of our parameters m_{M_*} and m_z, we conservatively choose a range that encompass the best-fit SFS from Speagle et al. (2014) and measurements from Moustakas et al. (2013) and Lee et al. (2015). With our SF centrals initalized at $z \sim 1$, next, we describe how we evolve their SFR and M_*.

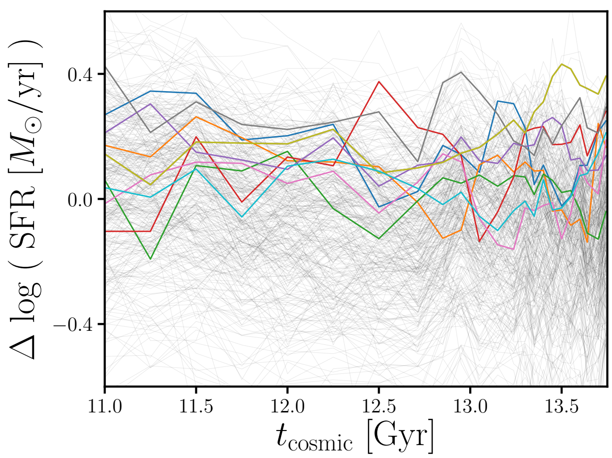

Fig. 2.— Star formation rate with respect to the log $\overline{\rm SFR}_{\rm SFS}$—$\Delta \log$ SFR)—as a function of cosmic time for star-forming galaxies in the Illustris simulation. These galaxies have stellar masses within the range $10^{10.5} - 10^{10.6} M_\odot$ at $z \sim 0$. log $\overline{\rm SFR}_{\rm SFS}$ is fit using the Hahn et al. (in prep.)SFS fitting method, the same method we use for our SDSS centrals in Section 3.1. As the $\Delta \log$ SFR(t)s in color emphasize how the SFRs of Illustris star-forming galaxies fluctuate about the mean SFS. *The* SFR *variability in the SFHs of SF centrals in our model (see Section 3.2) is motivated by this* $\Delta \log$ SFR(t) *behavior in Illustris.*

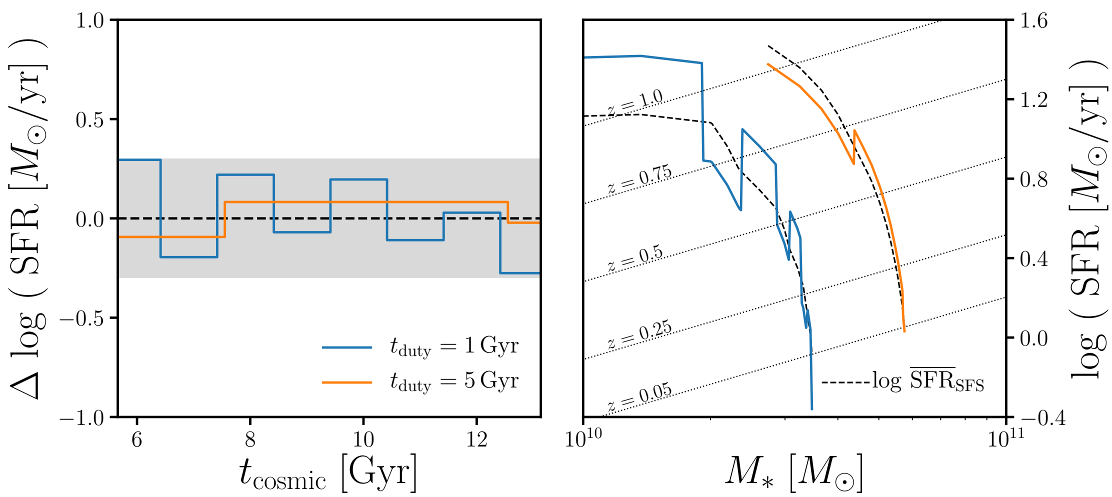

Fig. 3.— *Left*: The $\Delta \log$ SFR evolution of two SF centrals using the fiducial model that incorporates star formation variability through a "star formation duty cycle" on two timescales $t_{\rm duty} = 1$ Gyr (blue) and 5 Gyr (orange). $\Delta \log$ SFR(t)s determine the SFHs (Eq. 3), hence it also determines the M_* growth of the SF central galaxies in our model (Eq. 4). *Right*: The SFR$_i$ and $M_{*,i}$ evolutions of the same two SF central galaxies in our fiducial model with $t_{\rm duty} = 1$ Gyr (blue) and 5 Gyr (orange). For reference we plot $\log \overline{\rm SFR}_{\rm SFS}(M_{*,i}(t), t)$ for each galaxy (black dashed) and $\log \overline{\rm SFR}_{\rm SFS}(M_*)$ at various redshifts between $z = 1$ to 0.05 (dotted lines). As the SFR$_i$ and $M_{*,i}$ evolutions illustrate, *the SF centrals in our model evolve as their SFRs fluctate about the SFS.*

3.2. Evolving along the Star Formation Sequence

The tight correlation between the SFRs and stellar masses of SF galaxies (the so-called SFS) has been observed to span over four orders of magnitude in stellar mass and extends beyond the local universe out to $z > 2$ (*e.g.* Noeske et al. 2007; Daddi et al. 2007; **?**; Salim et al. 2007; Santini et al. 2009; Karim et al. 2011; Whitaker et al. 2012; Moustakas et al. 2013; Lee et al. 2015; see also references in Speagle et al. 2014). Even in hydrodynamic simulations and semi-analytic models, we find well defined SFSs (see Hahn et al. (in prep.)and references therein). The SFRs of galaxies on the SFS, for a given stellar mass follow a log-normal distribution with a roughly constant scatter (~ 0.3 dex in observations). Given its persistence in the local Universe, the SFS provides a anchoring relationship to characterize the SFRs and M_*s of SF galaxies throughout $z < 1$. More specifically, *we can characterize the star formation histories (SFHs) of our SF centrals with respect to the mean* log SFR *of the SFS* (Eq. 2):

$$\log \mathrm{SFR}(M_*, t) = \log \overline{\mathrm{SFR}}_{\mathrm{MS}}(M_*, t) + \Delta \log \mathrm{SFR}(t). \tag{3}$$

Since SFHs determine the M_* growth of galaxies, $\Delta \log$ SFRs in our model dictate the SFHs and M_* evolution of SF centrals. Below, we describe the prescription for our fiducial $\Delta \log \mathrm{SFR}(t)$.

One naive example for $\Delta \log \mathrm{SFR}(t)$ would be to keep $\Delta \log$ SFR fixed from the initial offsets from the $\log \overline{\mathrm{SFR}}_{\mathrm{SFS}}$ in the initial SFRs of our SF central galaxies at $z \sim 1$. SF centrals with higher than average initial SFRs continue evolving above the average SFS, while SF centrals with lower than average initial SFRs continue evolving below the average SFS. In addition to the fact that such a SFH cannot reproduce observations, which we later demonstrate, we do not find such a SFH in SF galaxies of hydrodynamic simulations such as Illustris Vogelsberger et al. (2014); Genel et al. (2014). In Figure 2, we plot $\Delta \log$ SFRs of SF galaxies in Illustris as a function of cosmic time. These galaxies have stellar masses within $10^{10.5} - 10^{10.6} M_\odot$ at $z = 0$. At each simultation output, we derive $\log \overline{\mathrm{SFR}}_{\mathrm{SFS}}$ using the same Hahn et al. (in prep.)fitting method as in Section 3.1 and use it to calculate $\Delta \log$ SFR (Eq. 3). As the $\Delta \log$ SFRs highlighted in color illustrate, the $\Delta \log \mathrm{SFR}(t)$s of SF galaxies do not remain constant, but rather vary about the SFS.

Motivated by the $\Delta \log \mathrm{SFR}(t)$ of Illustris galaxies, we introduce variability in the SFHs of our SF centrals in the form of a "*star formation duty cycle*". We parameterize $\Delta \log$ SFR to flucutate about the mean SFS on some duty cycle timescale t_{duty} with amplitude randomly sampled from a log-normal distribution with 0.3 dex scatter at every t_{duty} timestep. The full SFH of the SF centrals follow Eq. 3. In the left panel of Figure 3, we present $\Delta \log \mathrm{SFR}(t)$ of SF centrals with our fiducial star formation duty cycle prescription. The two $\Delta \log \mathrm{SFR}(t)$s have $t_{\mathrm{duty}} = 1$ Gyr (blue) and 5 Gyr (orange). The shaded region represents the observed 0.3 dex 1σ scatter of the SFS SFR. Although, we do not expect such a simplified model to reflect the exact individual SFHs of SF centrals, for the SF population it captures the stochasticity from gas accretion, star-bursts, and feedback mechanisms. Furthermore, it allows us to measure the timescale of such variabilities. Also this $\Delta \log \mathrm{SFR}$ prescription by construction reproduces the observed log-normal SFR distribution of the SFS at any point in the model.

Next using our $\Delta \log \mathrm{SFR}$ prescription, we now evolve both the SFR and M_* of our SF centrals along the SFS. Based on Eq. 3, the SFRs of our SF centrals are functions of M_*. Meanwhile, M_* is the integral of the SFR over time:

$$M_*(t) = f_{\mathrm{retain}} \int_{t_0}^{t} \mathrm{SFR}(\mathrm{M}_*, \mathrm{t})\, \mathrm{d}t + M_0. \tag{4}$$

t_0 and M_0 are the initial cosmic time and stellar mass at $z \sim 1$, respectively. f_{retain} here is the fraction of stellar mass that is retained after supernovae and stellar winds; we use $f_{\mathrm{retain}} = 0.6$ (Wetzel et al. 2013). By solving the differential equation from combining Eqs. 3 and 4, we evolve the SFR and M_* of our SF centrals. The right panel of Figure 3 presents the SFR_i and $M_{*,i}$ evolutions for the two SF centrals with different t_{duty} timescales in the left panel. For reference, we include $\log \overline{\mathrm{SFR}}_{\mathrm{SFS}}(M_{*,i}(t), t)$ (black dashed) and $\log \overline{\mathrm{SFR}}_{\mathrm{SFS}}(M_*)$ (dotted lines) at various redshifts between $z = 1$ to 0.05. As Figure 3 illustrates, using our $\Delta \log \mathrm{SFR}$ prescription the SF centrals evolve as their SFRs fluctuate along the SFS throughout the timesteps of our model.

We continue to evolve our SF central galaxies until the final $z = 0.05$ snapshot. For the SF centrals in our model, not only do we have SFRs and M_*s that we evolved but we also have their host halo properties from the `TreePM` N-body simulation. Using these properties, we can compare our model to observations and constrain the free parameters using observables such as the quiescent fraction and SMF. Once we have a model that reproduces the standard observables we can use the host halo properties to examine observables such as the SHMR. Next, we present this comparison between our model and observations and present the constraints we derive on the role and timescale of star formation variability in the evolution of SF galaxies.

4. Results

Our model takes `TreePM` central subhalos and tracks their SFR and M_* evolution using a flexible parameterization of the SFS and SFHs that incorporate variability through a star formation duty cycle. At $z = 0.05$, its final timestep, our model provides properties such as the SFR, M_*, and host halo mass, M_h, of central galaxies it classifies as SF. We now use these resulting properties to compare our model to observations and constrain its free parameters—the parameters of the Eq. 2 SFS. Since the focus of our model and this paper is on SF centrals, the main observable we use is the SMF of star forming centrals in SDSS, which we estimate as

$$\Phi^{\rm SDSS}_{\rm SF,cen} = f^{\rm cen}_{\rm SFS} \times f_{\rm cen} \times \Phi^{\rm Li\&White(2009)}. \tag{5}$$

$f^{\rm cen}_{\rm SFS}$ is the fraction of central galaxies on the SFS, which we fit in Eq. 1. $f_{\rm cen}$ is the central galaxy fraction from Wetzel et al. (2013) and $\Phi^{\rm Li\&White(2009)}$ is the SMF of the SDSS from Li & White (2009). If our model reproduces the observed $\Phi^{\rm SDSS}_{\rm SF,cen}$, by construction, it also reproduces the observed quiescent fraction.

For the actual comparison of our model to observation, we use the parameter estimation framework of Approximate Bayesian Computation (ABC). ABC has the advantage over standard approaches to parameter inference in that it does not require evaluating the likelihood. It relies only on a simulation of the observed data and a distance metric to quantify the "closeness" between the observed data and simulation. Many variations of ABC has been used in astronomy and cosmology (*e.g.* **???**Alsing et al. 2018). We use ABC in conjunction with the efficient Population Monte Carlo (PMC) importance sampling as in (**?**Hahn et al. 2017a). For initial range of our ABC particles, *i.e.* the priors of our Eq. 2 SFS parameters A_z **and** m_z, we use uniform distributions spanning the ranges **numbers** and **numbers**, respectively. As we mentioned in Section 3.1, the range of the prior were conservatively chosen to encompass the best-fit SFS from Speagle et al. (2014) and measurements from Moustakas et al. (2013) and Lee et al. (2015) at $z \sim 1$. Finally, for our distance metric, we formulate a distance between the SMF of the SF centrals in our model to the observed $\Phi^{\rm SDSS}_{\rm SF,cen}$ (Eq. 5):

$$\rho_\Phi = \sum_M \left(\frac{\Phi^{\rm sim} - \Phi^{\rm SDSS}_{\rm SF,cen}}{\sigma'_\Phi}\right)^2. \tag{6}$$

$\Phi^{\rm sim}(M)$ is the SMF of the SF centrals in our model and $\sigma'_\Phi(M)$ **is the SMF uncertainty derived using mock catalogs from Li & White (2009)**. For the rest of our ABC-PMC implementation, we strictly follow the implementation of Hahn et al. (2017b) and **?**. We refer reader to those papers for further details.

4.1. The Star Formation Duty Cycle

In Figure 4, we present the SMFs (left), SFSs (middle), and $\sigma_{\log M_*}(\log M_{\rm halo})$ (right) of two models with different SFH prescriptions each run using the medians parameter values of their ABC posterior distributions — $\theta_{\rm median}$. One model has duty cycle of $t_{\rm duty} = 10\,{\rm Gyr}$ (red) while the other has a duty cycle of $t_{\rm duty} = 1\,{\rm Gyr}$ (blue). Both models, as expected from the ABC posteriors, successfully reproduce the SDSS SMF ($\Phi^{\rm SDSS}_{\rm SF,cen}$). By construction, *i.e.* the priors, they also have SFSs consistent with observations. Despite their consistency, however, *the difference in the star formation duty cycle timescales of the models results in dramatically different scatter in* $\log M_*$ *at a given* $\log M_{\rm halo}$ *(*i.e. *scatter in the SHMR)*, particularly below $M_{\rm halo} < 10^{12.5} M_\odot$.

Since the scatter in SHMR of our model depends on the star formation duty cycle timescale, we can compare the scatter from our model to observational constraints on the SHMR in order to constrain the star formation duty cycle and the SFHs of star forming galaxies (Figure 5). More specifically in Figure 5, we present $\sigma_{\log M_*}$, the scatter in $\log M_*$ at fixed halo mass, at $M_h = 10^{12} M_\odot$) and $\sigma_{\log M_h}$, the scatter in $\log M_h$ at fixed stellar mass, at $M_* = 10^{10} M_\odot$ as a function of t_duty. For t_duty ranging from 10 to 0.5 Gyr, $\sigma_{\log M_*}$ ranges from $0.32^{+5.1}_{-4.0}$ **to** $0.26^{+5.1}_{-4.0}$ dex **and** $\sigma_{\log M_h}$ **ranges from number to number**. With a shorter star formation duty timescale, our model produces significantly smaller scatter in the SHMR.

In order to constrain t_duty of our model, we compare the scatter in our model to $\sigma_{\log M_*}$s and $\sigma_{\log M_h}$s from observational constraints in the literature. In the left panel of Figure 5, we include Yang et al. (2009); More et al. (2011); Leauthaud et al. (2012); Reddick et al. (2013); Tinker et al. (2013); Zu & Mandelbaum (2015). These works mainly derive their $\sigma_{\log M*}$ measurements using halo occupation based models. More et al. (2011) use a halo occupation based model and kinematics of satellite galaxies to infer $sigma_{\log M*}$ that is constant across M_h from the SDSS NYU-VAGC for blue central galaxies. In Leauthaud et al. (2012), they analyze weak lesning, galaxy clustering, and galaxy number densities of the COSMOS data to constrain $sigma_{\log M*}$ for all galaxies.

In the right, we include Mandelbaum et al. (2006b); More et al. (2011); **?**); Han et al. (2015).

- No duty cycle model obviously ruled out
- More et al. (2011) use kinematics of satellite galaxies in a dark matter halo to infer the scaling relations between halo mass and central galaxy properties. Their model is based on halo occupation and assumes constant $\sigma_{\log M*}$. They fit this model to SDSS VAGC and get $\sigma_{\log M_*}(M_h \sim 10^{12} M_\odot) = 0.07 - 0.15 - 0.26$ $\sigma_{\log M_h}(M_* \sim 10^{10.5} M_\odot) = 0.055 - 0.195$ for blue centrals.
- Leauthaud et al. (2012) joint analysis of galaxygalaxy weak lensing, galaxy spatial clustering, and galaxy number densities of the COSMOS data. $\sigma_{\log M*}$ is constant. They get the constraint $\sigma_{\log M_*}(M_h \sim 10^{12} M_\odot) = 0.206 \pm 0.031$ at $0.2 < z < 0.48$ for all galaxies.
- Tinker et al. (2013) uses the stellar mass function, galaxy clustering, and galaxy-galaxy lensing within the COSMOS survey to constrain the stellar-to-halo mass relation (SHMR) of star forming and quiescent galaxies over the redshift range $z = [0.2, 1.0]$. They constrain the constant $\sigma_{\log M_*}(M_h \sim 10^{12} M_\odot) = 0.21 \pm 0.06$ at $0.22 < z < 0.48$ for star-forming galaxies.
- Reddick et al. (2013) constrain a model that abundance matches galaxies with the peak circular velocity of their halos with the SDSS NYU-VAGC projected two-point

galaxy clustering and the observed conditional stellar mass functions. The model that best reproduces the data has a scatter of 0.20 ± 0.03 dex for all galaxies.

- Zu & Mandelbaum (2015) `iHOD`: an exteded HOD model, to fit the galaxy clustering and the galaxy-galaxy lensing measured from SDSS main galaxy sample and NYU-VAGC. The best-fit model has $\sigma_{\log\, M_*}(M_h \sim 10^{12} M_\odot) = 0.22 \pm 0.02$ for all galaxies.

- Yang et al. (2009) uses a group catalog constructed from SDSS DR4 to measure conditional SMF. Based on the CSMF for blue centrals they find $\sigma_{\log\, M_*}(M_h \sim 10^{12.16} M_\odot) = 0.122 \pm 0.03$ averaged over two SDSS samples and two group mass difference.

from weak lensing:

- Mandelbaum et al. (2006b) observations constraints from SDSS galaxygalaxy weak lensing, with a total of 351 507 lenses. We use stellar masses derived from spectroscopy and virial halo masses derived from weak gravitational lensing. At $M_* = 1.5 \times 10^1 0 M_\odot$ $M_{\rm halo} = 6.6^{+5.1}_{-4.0} \times 10^{11} h^{-1} M_\odot$.

- ?: weak lensing from imaging data from the $\sim 300\,\deg^2$ overalp between the second Red Sequence Cluster Survey (RCS2) and SDSS DR7 using halo model that accounts for the clustering of the lenses and distinguishes between satellite and centrals. At $\log\, M_* = [10.0 - 10.5]$ $M_{\rm halo} = 0.56^{+1.66}_{-0.55} \times 10^{11} h^{-1} M_\odot$.

- ?: 154 deg2 CFHTLenS lensing and photometric data using a galaxy-galaxy lensing halo model which allows us to constrain the halo mass and satellite fraction. There's some scatter correction which reduces the uncertainties significantly. At $< M_* >= 0.54 \times 10^1 0 M_\odot / h^2$ $M_{\rm halo} = 2^{+0.64}_{-0.62} \times 10^{11} h^{-1} M_\odot$.

- Han et al. (2015): weak lensing analysis of optically selected spectroscopic galaxy groups (G3Cv5) in GAMA survey using background SDSS photometric galaxies.

- We find that by decreasing the timescale of stochasticity on a simple SFH model that traces the overall SFS evolution does in fact decrease the scatter seen in the SMHMR. However, even with timescales less than XXXX, we cannot reproduce observations. Ultimately to reproduce observations, we need to add in assembly bias.

Figure 5

4.2. The need for a galaxy assembly bias

- discuss how t_{duty} is not enough to be consistent with σ_{M_*}.

- first clarify what you mean by galaxy assembly bias

- discuss implementation of galaxy assembly bias
- Figure (pedagogical) of dlogSFR versus dMh dt for different correlation amounts
- Figure of different tdelay and dtabias
- Figure of sigma M star as a function of duty cycle and realistic dt abias and t delay

5. Discussion

5.1. Rethinking the Main Sequence?

- Test the SMHMR for Louis's SFHs

6. Summary

A. $z \sim 1$ observations

Much of the results presented in this paper are based on comparison between our model and observations at $z \sim 0$.. Our model is initalized at $z \sim 1$. Therefore, in this section we test some of the choices we make in our intializations.

- Test impact of $z \sim 1$ SMF
- Test impact of $z \sim 1$ $\sigma_{\log M_*}$

Acknowledgements

It's a pleasure to thank Louis Abramson, Shy Genel, **more acknowledgements** for valueable discussions.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, The Astrophysical Journal Supplement Series, 182, 543

Alsing, J., Wandelt, B., & Feeney, S. 2018, arXiv:1801.01497 [astro-ph], arXiv:1801.01497 [astro-ph]

Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, Monthly Notices of the Royal Astronomical Society, 373, 469

Blanton, M. R., & Moustakas, J. 2009, Annual Review of Astronomy and Astrophysics, 47, 159

Blanton, M. R., & Roweis, S. 2007, The Astronomical Journal, 133, 734

Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, The Astronomical Journal, 129, 2562

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, Monthly Notices of the Royal Astronomical Society, 351, 1151

Campbell, D., van den Bosch, F. C., Hearin, A., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 444

Chabrier, G. 2003, Publications of the Astronomical Society of the Pacific, 115, 763

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, The Astrophysical Journal, 647, 201

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, The Astrophysical Journal, 670, 156

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, The Astrophysical Journal, 292, 371

Drory, N., Bundy, K., Leauthaud, A., et al. 2009, The Astrophysical Journal, 707, 1595

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 175

Hahn, C., Tinker, J. L., & Wetzel, A. R. 2017a, The Astrophysical Journal, 841, 6

Hahn, C., Vakili, M., Walsh, K., et al. 2017b, Monthly Notices of the Royal Astronomical Society, 469, 2791

Hahn, C., Blanton, M. R., Moustakas, J., et al. 2015, The Astrophysical Journal, 806, 162

Han, J., Eke, V. R., Frenk, C. S., et al. 2015, Monthly Notices of the Royal Astronomical Society, 446, 1356

Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, Astronomy and Astrophysics, 556, A55

Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, The Astrophysical Journal, 730, 61

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, The Astrophysical Journal, 744, 159

Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, The Astrophysical Journal, 801, 80

Leja, J., van Dokkum, P., & Franx, M. 2013, The Astrophysical Journal, 766

Li, C., & White, S. D. M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 2177

Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, The Astrophysical Journal, 760, 6

Mamon, G. A., Sanchis, T., Salvador-Solé, E., & Solanes, J. M. 2004, Astronomy and Astrophysics, 414, 445

Mandelbaum, R., Seljak, U., Cool, R. J., et al. 2006a, Monthly Notices of the Royal Astronomical Society, 372, 758

Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006b, Monthly Notices of the Royal Astronomical Society, 368, 715

Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, The Astrophysical Journal, 701, 1765

More, S., van den Bosch, F. C., Cacciato, M., et al. 2011, Monthly Notices of the Royal Astronomical Society, 410, 210

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, The Astrophysical Journal, 767, 50

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, The Astrophysical Journal, 777, 18

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, The Astrophysical Journal Letters, 660, L43

Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, The Astrophysical Journal, 721, 193

Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, The Astrophysical Journal, 771, 30

Salim, S., Rich, R. M., Charlot, S., et al. 2007, The Astrophysical Journal Supplement Series, 173, 267

Santini, P., Fontana, A., Grazian, A., et al. 2009, Astronomy and Astrophysics, 504, 751

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, The Astrophysical Journal Supplement Series, 214, 15

Tinker, J., Wetzel, A., & Conroy, C. 2011, ArXiv e-prints, 1107, arXiv:1107.5046

Tinker, J. L., Hahn, C., Mao, Y.-Y., & Wetzel, A. R. 2017, arXiv:1705.08458 [astro-ph], arXiv:1705.08458 [astro-ph]

Tinker, J. L., Leauthaud, A., Bundy, K., et al. 2013, The Astrophysical Journal, 778, 93

Vale, A., & Ostriker, J. P. 2006, Monthly Notices of the Royal Astronomical Society, 371, 1173

Velander, M., van Uitert, E., Hoekstra, H., et al. 2014, Monthly Notices of the Royal Astronomical Society, 437, 2111

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Monthly Notices of the Royal Astronomical Society, 444, 1518

Wetzel, A. R., Cohn, J. D., & White, M. 2009, Monthly Notices of the Royal Astronomical Society, 395, 1376

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, Monthly Notices of the Royal Astronomical Society, 424, 232

Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, Monthly Notices of the Royal Astronomical Society, 432, 336

—. 2014, Monthly Notices of the Royal Astronomical Society, 439, 2687

Wetzel, A. R., & White, M. 2010, Monthly Notices of the Royal Astronomical Society, 403, 1072

Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, The Astrophysical Journal Letters, 754, L29

White, M. 2002, The Astrophysical Journal Supplement Series, 143, 241

White, M., Cohn, J. D., & Smit, R. 2010, Monthly Notices of the Royal Astronomical Society, 408, 1818

Yang, X., Mo, H. J., & van den Bosch, F. C. 2009, The Astrophysical Journal, 695, 900

Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005, Monthly Notices of the Royal Astronomical Society, 356, 1293

Zu, Y., & Mandelbaum, R. 2015, Monthly Notices of the Royal Astronomical Society, 454, 1161

This preprint was prepared with the AAS LaTeX macros v5.2.

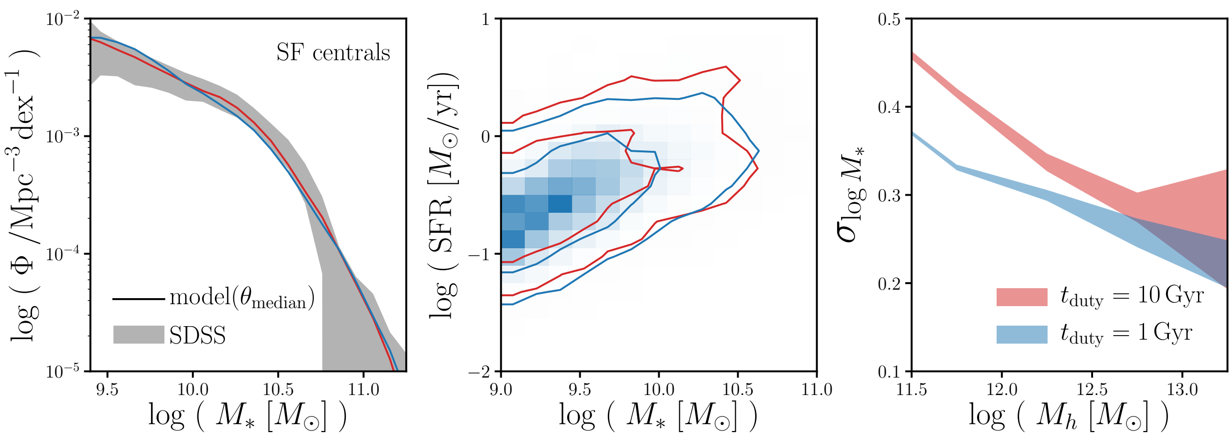

Fig. 4.— Our models with different star formation duty cycle timescales (blue: $t_{\rm duty}$=1 Gyr; red: $t_{\rm duty}$=10 Gyr) run with median values of their ABC posterior distribution ($\theta_{\rm median}$) have SMFs (left) and SFSs (middle) consistent with observations. *They however produce significantly different scatter in* $\log M_*$ *at fixed* $\log M_{\rm halo}$ *— scatter in the SHMR (right)*. By comparing the scatter in SHMR of our models to observational constraints on the SHMR, we can constrain the timescale of the star formation duty cycle and thereby the SFHs of star forming galaxies.

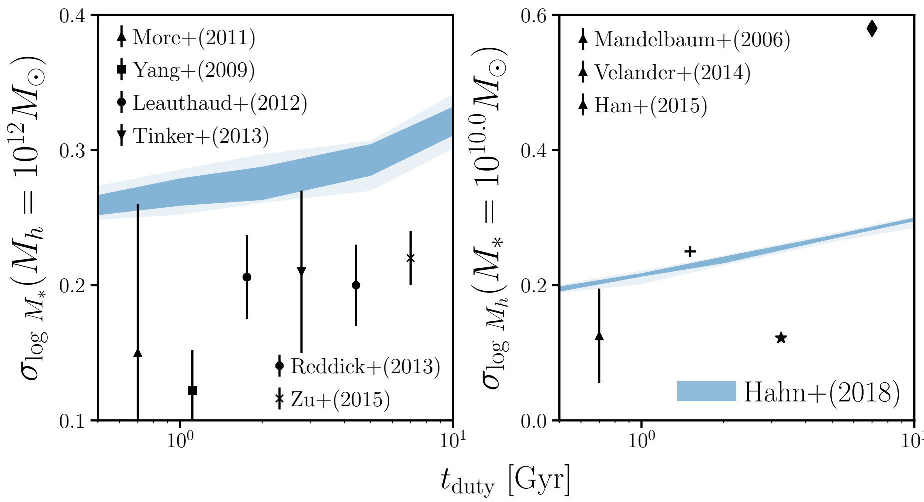

Fig. 5.— With shorter star formation duty cycle timescales, $t_{\rm duty}$, our model produces smaller scatter in the SHMR—$\sigma_{\log M_*}$ at M_h=$10^{12}M_\odot$ (left) and $\sigma_{\log M_h}$ at M_*=$10^{10}M_\odot$. For $t_{\rm duty}$ = 10 Gyr, $\sigma_{\log M_*}$ = 0.32 dex and $\sigma_{\log M_h}$ = 0.30 dex. Meanwhile for $t_{\rm duty}$ = 0.5 Gyr, $\sigma_{\log M_*}$ = 0.26 dex and $\sigma_{\log M_h}$ = 0.20. We include $\sigma_{\log M_*}$ constraints from Yang et al. (2009); More et al. (2011); Leauthaud et al. (2012); Reddick et al. (2013); Tinker et al. (2013); Zu & Mandelbaum (2015) and $\sigma_{\log M_h}$ constraints from Mandelbaum et al. (2006a); More et al. (2011); Velander et al. (2014); Han et al. (2015). The Mandelbaum et al. (2006a); Han et al. (2015) $\sigma_{\log M_h}$ constraints serve as upper limits. A short star formation duty cycle timescale is necessary to produce a tight SHMR roughly consistent with constraints from observations.

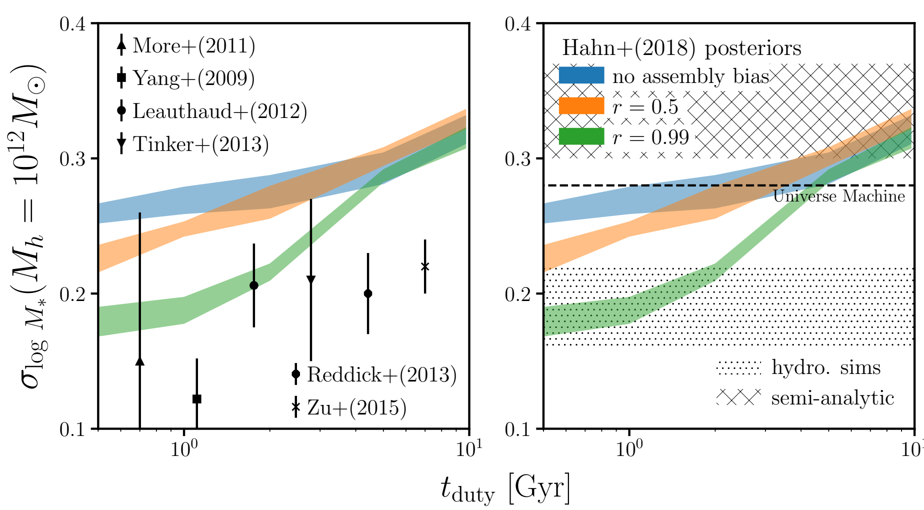

Fig. 6.—