The Astrophysical Journal, 877:141 (22pp), 2019 June 1 https://doi.org/10.3847/1538-4357/ab1d53

Massive Dead Galaxies at $z \sim 2$ with *HST* Grism Spectroscopy. I. Star Formation Histories and Metallicity Enrichment

T. Morishita[1], L. E. Abramson[2], T. Treu[3], G. B. Brammer[1,4], T. Jones[5], P. Kelly[6], M. Stiavelli[1], M. Trenti[7,8], B. Vulcani[9], and X. Wang[3]

[1] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA; tmorishita@stsci.edu
[2] The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[3] Department of Physics and Astronomy, UCLA, 430 Portola Plaza, Los Angeles, CA 90095-1547, USA
[4] Cosmic Dawn Centre, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
[5] University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA
[6] School of Physics and Astronomy, University of Minnesota,116 Church Street SE, Minneapolis, MN 55455, USA
[7] School of Physics, Tin Alley, University of Melbourne, VIC 3010, Australia
[8] ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions, Australia
[9] INAF—Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122, Padova, Italy
Received 2018 October 4; revised 2019 April 11; accepted 2019 April 26; published 2019 June 5

Abstract

Observations have revealed massive ($\log M_*/M_\odot \gtrsim 11$) galaxies that were already dead when the universe was only ∼2 Gyr. Given the short time before these galaxies were quenched, their past histories and quenching mechanism(s) are of particular interest. In this paper, we study star formation histories (SFHs) of 24 massive galaxies at $1.6 < z < 2.5$. A deep slitless spectroscopy and imaging data set collected from multiple *Hubble Space Telescope* surveys allows robust determination of their spectral energy distributions and SFHs with no functional assumption on their forms. We find that most of our massive galaxies had formed >50% of their extant masses by ∼1.5 Gyr before the time of observed redshifts, with a trend where more massive galaxies form earlier. Their stellar-phase metallicities are already compatible with those of local early-type galaxies, with a median value of $\log Z_*/Z_\odot = 0.25$ and scatter of ∼0.15 dex. In combination with the reconstructed SFHs, we reveal their rapid metallicity evolution from $z \sim 5.5$ to ∼2.2 at a rate of ∼0.2 dex Gyr^{-1} in $\log Z_*/Z_\odot$. Interestingly, the inferred stellar-phase metallicities are, when compared at half-mass time, ∼0.25 dex higher than observed gas-phase metallicities of star-forming galaxies. While systematic uncertainties remain, this may imply that these quenched galaxies have continued low-level star formation, rather than abruptly terminating their star formation activity, and kept enhancing their metallicity until recently.

Key words: galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: star formation

1. Introduction

In the local universe, early-type galaxies dominate the massive end of the galaxy mass function, $\log M_*/M_\odot \gtrsim 11.5$ (Cole et al. 2001; Bell et al. 2003). Those galaxies consist of old and chemically enriched stellar populations, indicating that most of their star formation activities ended ≳10 Gyr (Kauffmann et al. 2003; Thomas et al. 2003, 2010; Gallazzi et al. 2005; Treu et al. 2005b). In fact, observations have revealed that some galaxies are already massive and passively evolving at $z \gtrsim 2$ (Cimatti et al. 2004; Daddi et al. 2005; van Dokkum et al. 2008; Kriek et al. 2009; Belli et al. 2014; Straatman et al. 2014; Marsan et al. 2015; Glazebrook et al. 2017). Given the short time since the big bang and their stellar mass, their earlier star formation must be extremely intense, followed by a rapid cessation of their star formation activity, which we here refer to as quenching.[10]

However, these episodes still remain observationally indirect. What were their star formation histories (SFHs) like? How and why did they stop forming stars, especially at the peak time of the cosmic star formation? Are they already enriched in metallicity as local counterparts, or do any post-quenching processes play key roles over the following 10 Gyr? These are the central questions we aim to answer in this series of papers. In this first paper, we focus on their SFHs.

A number of studies have investigated SFHs of massive galaxies using different approaches. For example, observations of high-redshift galaxies provide an analogy to their past properties, especially when they were actively forming stars. While sufficient valuable information can be obtained from high-z populations (e.g., star formation rate (SFR), number density, metallicity; e.g., Hamann & Ferland 1999; Tacconi et al. 2008; Toft et al. 2014), it is limited by its rather indirect aspect, where connecting different objects at different epochs may introduce systematic uncertainties (e.g., Wellons et al. 2015; Torrey et al. 2017).

Another approach is based on the archeological information of local galaxies, or fossil record (Thomas et al. 2003, 2010; Heavens et al. 2004; McDermid et al. 2015). Detailed information about their stellar population (e.g., age and chemical abundances) provides their past histories, characterizing their formation redshift to ≳2. This information from the local galaxies is, however, limited up to several Gyr with current observing facilities (e.g., Worthey 1994; see also Conroy 2013), which is short for exploring the SFHs of galaxies that are already dead at $z \sim 2$.

[10] The term may refer to different phenomena in different contexts. For example, one may also refer to keeping star formation at very low levels after an initial decline, or the (rapid) decline of star formation itself (see Man & Belli 2018, for a recent review). We use the term to describe any decline of galaxy star formation activity regardless of speed.

To explore evolution histories in the earlier epoch, we need a method that combines some of the virtues of both approaches, that is, the archeological study of high-z galaxies. For example, Kelson et al. (2014) performed spectral energy distribution (SED) modeling of galaxies at $z \sim 1$ by using low-resolution spectra and broadband photometry and reconstructed their SFHs back to $z \sim 1.5$ (also Dressler et al. 2016, 2018). Chauke et al. (2018) recently attempted a similar approach to galaxies at $z \lesssim 1$ but with higher-resolution spectra taken with a ground-based spectrograph and successfully revealed their formation histories back to $z \sim 2.5$. In the current study, we target galaxies at higher redshift, aiming at earlier evolution histories up to their formation redshift from the fossil record obtained with the low spectral resolution yet high-sensitivity *Hubble Space Telescope* (*HST*) spectrophotometric data set.[11]

Stellar metallicity is another key parameter that provides further details of physical mechanisms. In particular, since both the cosmic metallicity and individual gas-phase metallicity are still pristine at these redshifts (e.g., Erb et al. 2006; Maiolino et al. 2008; Lehner et al. 2016), the enrichment process within such massive systems has to be substantial to explain the observed solar/supersolar metallicity of lower-redshift galaxies (Onodera et al. 2012; Choi et al. 2014; Gallazzi et al. 2014; Lonoce et al. 2015), whereas the process is highly dependent on SFHs (e.g., Peng et al. 2015).

With such demands, we here improve our previous methodology of SED modeling, which is free from functional forms of SFHs, by increasing the flexibility in metallicity. We collect 24 massive quenched galaxies at $z \sim 2$ that have deep WFC3/G102 and G141 grism spectra coverage in their rest-frame 4000 Å. The combination of grism spectra and wide broadband photometry (0.2–8.0 μm by *HST* and *Spitzer*) provides a unique opportunity to constrain not only age but also metallicity from the entire SED shape.

We proceed as follows. In Section 2, we describe the data used in this study and their reduction process. In Section 3, we introduce our method for the SED modeling. In Section 4, we show the results. We discuss our results and interpretation in Section 5 and close in Section 6. Further details, including intensive simulation tests of SED modeling and comparison with functional SFHs, are presented in the Appendices. Throughout the text, magnitudes are quoted in the AB system (Oke & Gunn 1983; Fukugita et al. 1996); $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $H_0 = 72$ km s^{-1}Mpc^{-1} for the cosmological parameters; and $Z_\odot = 0.0142$ (Asplund et al. 2009) for the solar metallicity.

2. Data

To achieve our goal of constraining galaxy SFHs, it is essential to cover the wavelength range surrounding 4000 Å, where the spectral features are most age-sensitive, with sufficiently deep spectra and rest-frame NUV-optical-NIR wavelength with broadband photometry. Therefore, we limit ourselves to only fields where deep *HST* grism data are available, which are MACS 1149.6+2223 (hereafter M1149) and the GOODS-North/South (GDN/GDS). To collect the initial photometric sample galaxies, we use publicly available photometric catalogs. We use *HST*/WFC3 G102 and G141 grism data (that covers $\lambda_{\rm obs} \sim$ 8000–17000 Å) taken in various surveys in these fields.

[11] Belli et al. (2019) recently presented SFHs of $1.5 < z < 2.5$ galaxies reconstructed with ground-based spectroscopic data. None of their sample galaxies overlaps with ours in this study.

2.1. Initial Photometric Sample

M1149 is a sightline of a massive cluster of galaxies at $z = 0.544$. The data were taken in CLASH (Postman et al. 2012), Hubble Frontier Fields (Lotz et al. 2017), GLASS (Schmidt et al. 2014; Treu et al. 2015), and the SN Refsdal follow-up campaigns (Kelly et al. 2015, 2016). The combination of its gravitational magnification power and those very deep observations provides a unique opportunity among other clusters of GLASS. We use the photometric catalog used in Morishita et al. (2017, 2018), which consists of *HST* photometry taken in all *HST* surveys above, as well as *Spitzer* IRAC (3.6 + 4.5 μm; PIs: T. Soifer and P. Capak) and ground-based K_S-band imaging (Brammer et al. 2016) and publicly available spectroscopic redshifts by GLASS (Schmidt et al. 2014).

For GDN and GDS, we use the publicly available catalog by 3DHST (Skelton et al. 2014). The photometric catalog consists of fluxes, spectroscopic/photometric redshifts, rest-frame colors, and stellar mass based on data taken in CANDELS and 3DHST (Grogin et al. 2011; Koekemoer et al. 2011; van Dokkum et al. 2013a; Skelton et al. 2014; Momcheva et al. 2016), as well as photometric fluxes obtained in ground-based surveys. We use *HST* photometry for optical/near-IR range, ground-based K_S-band photometry, and *Spitzer* IRAC photometry in this study. Other ground-based fluxes listed in the 3DHST catalog, most of which are in the optical range, are not used, as the deep *HST* photometry is sufficient to constrain optical SEDs.

Broadband fluxes of *HST* are measured in a fixed aperture ($r = 0\farcs7$) and then scaled to the total flux by multiplying $C = f_{\rm AUTO}/f_{\rm aper}$, where $f_{\rm AUTO}$ is the AUTO flux of SExtractor (Bertin & Arnouts 1996), as in Skelton et al. (2014) and Morishita et al. (2017).

From these photometric catalogs, we choose those that satisfy $m_{140} < 24$, $1.6 < z < 3.3$, and $\log M_*/M_\odot > 10.8$. We also apply $(U - V)_{\rm rest} > 1$ mag to select quiescent galaxies. By setting a slightly bluer color for $U - V$ than in the literature ($\sim$1.4 mag), our sample also contains quenching galaxies. With the criteria, we found 17, 51, and 73 galaxies in M1149, GDN, and GDS, respectively, as an initial photometric sample (Figure 1).

2.2. HST *Grism Spectrum*

In M1149, the grism data were taken through GLASS (Schmidt et al. 2014; Treu et al. 2015). GLASS is a spectroscopic survey with *HST*/WFC3 G102 and G141 grisms (10 and 4 orbits, respectively). In addition, we supplement with the follow-up HST-GO/DDT campaign (Proposal ID 14041; PI: P. Kelly) of the multiply imaged supernova SN Refsdal (Kelly et al. 2015, 2016), which adds another 30 orbits of G141 data to the original GLASS observation.

In GDN/GDS, we retrieve the public data through MAST. In addition to the 3DHST data (van Dokkum et al. 2013b; Momcheva et al. 2016) that cover entire CANDELS's GDN/GDS fields, we add those taken in FIGS (13779; PI: S. Malhotra Pirzkal et al. 2017), CLEAR (14227; PI: C. Papovich, also Estrada-Carpenter et al. 2018), and other follow-ups (12099 and 12461, PI: A. Riess; 12190, A. Koekemoer; 13420, PI: G. Barro; 13871, PI: P. Oesch).

We extract 1D spectra from all fields in a consistent way by using the latest version of Grizli (Brammer 2018). During the

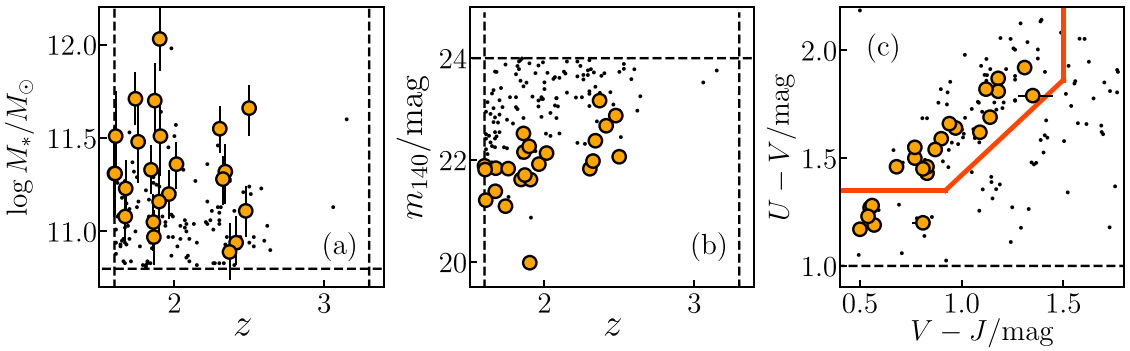

Figure 1. Physical properties of the initial photometric sample galaxies (dots) and final 24 galaxies (orange circles). The criteria used for the initial photometric selection are shown with dashed lines. (a) Stellar-mass distribution as a function of redshift. (b) The F140W magnitude distribution as a function of redshift. (c) The *UVJ* color diagram. Our sample consists of both quenched (top left region bounded by solid lines) and quenching (those $U-V>1$ but outside the boundary) galaxies.

extraction, the code automatically models neighboring objects, which are flagged in pre-provided SExtractor segmentation maps, and produces clean spectra for a target galaxy. The clean, optimal extracted spectra from each position angle (PA) are then stacked in a refined wavelength grid of 45 Å pixel^{-1}. The pixel scale is slightly finer than the Nyquist sampling of the G141 grism, since we have many samplings over different orbits, each of which slightly shifts in the dispersion direction. Each spectrum is convolved with the image of the source to match the morphological difference in different PAs.

For the aperture correction of broadband photometry, we match the pseudo-broadband flux extracted from grism spectra by convolving with the corresponding filters (F140W/F105W) to the observed broadband flux (Section 3.2).

In addition to the random uncertainty in flux, we also estimate the uncertainty associated with the stacking of different PAs by following Onodera et al. (2015), and we integrate this to the random noise in quadrature for conservative estimates. The uncertainty accounts of ∼20% of the random uncertainty. Signal-to-noise ratios (S/Ns) of the final 1D spectrum range up to ∼50. Median values of each spectral element are S/N ∼ 18 at 4200–5000 Å and ∼4 at 3400–3800 Å (Table 1).

2.3. Additional Photometric Data

In addition to photometric fluxes collected in 3DHST, we add WFC3/*UVIS* photometry. The rest-frame UV coverage is important to constrain SEDs with/without the UV upturn, which depends on metallicity (e.g., Yi et al. 1997; Treu et al. 2005a). We use WFC3/*UVIS* images from the HDUV legacy survey (Oesch et al. 2018) that cover parts of the GDN/GDS fields with the F275W and F336W filters. The *UVIS* images in the GDS consist of the previous data taken in the UDF by the UVUDF team (Teplitz et al. 2013). We run SExtractor on the public imaging data to conduct photometry and use the flux measured in a fixed aperture of 0.″72 diameter.

3. Spectrophotometric SED Fitting

3.1. Basic Templates

Our SED fitting method (Grism SED Fitter, `gsf`[12]) is based on the canonical template fitting, where the best-fit parameters are determined by minimizing the residuals of observed and model SEDs. One major difference from most other works is the way we construct the model templates. The SED templates are often constructed with a functional form for SFHs, such as the exponential declining model, $\psi \propto A\exp[-(t-T_0)/\tau]$, where A, T_0, and τ are free parameters. However, it is known that such a simplification may not represent real galaxy SFHs by observations (e.g., Pacifici et al. 2016; Iyer & Gawiser 2017) and simulations (e.g., Diemer et al. 2017). As such, here we avoid any functional forms and adopt an alternative method to generate model templates. The core of the method is to find the best combination of amplitudes, $\{a_i\}$, for a set of composite stellar population templates of different ages, $\{t_i\}$, that matches the data, as previously performed by Morishita et al. (2018). This type of SED modeling has been used in previous studies (Heavens et al. 2004; Cid Fernandes et al. 2005; Panter et al. 2007; Tojeiro et al. 2007; Kelson et al. 2014; Dressler et al. 2018), some of which demonstrated its strength and validity with intensive simulation tests.

To generate the template with different parameters, we use the flexible stellar population synthesis code (FSPS; Conroy et al. 2009; Conroy & Gunn 2010; Foreman-Mackey et al. 2014) to generate ith templates with ages of t_i, based on MIST isochrones (Choi et al. 2016) and the MILES stellar library. As found by Morishita et al. (2018), different isochrones may return different results, in addition to a systematic difference in assumed metallicities.

We set the number of age "pixels" to 10, with [0.01, 0.03, 0.1, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, 3.0] Gyr (Figure 2), doubling the number from those adopted in Morishita et al. (2018). While we set the equal width of the template in lognormal space (∼0.5) following previous studies (e.g., Cid Fernandes et al. 2005), we added extra bins at intermediate age, where most of our galaxies are located, to increase the flexibility of SFHs.

The template is generated by assuming a short constant SFR within each bin width (∼30 Myr), rather than a simple stellar population (SSP). The reason we do not adopt the SSP model is that, while it is simple, it is unrealistic for real galaxies. Changing the width of constant star formation in each bin would result in a minor but systematic shift in reconstructed SFHs. The uncertainty in bin width is considered in calculation of parameters (e.g., age) by randomly fluctuating values within the width.

We also set the metallicity of each age pixel as a free parameter in a range of $\log Z_*/Z_\odot \in [-0.8:0.6]$, as opposed to one global value in Morishita et al. (2018). While the determination of metallicity at each age pixel (i.e., metallicity histories of individual galaxies) is more challenging (see Appendix A), this gives extra

[12] The code is available at https://github.com/mtakahiro/gsf.

Table 1
Summary of Physical Parameters

Obj. ID	R.A. (deg)	Decl. (deg)	$z_{\rm grism}$	$\log M_*$ ($M_\odot$)	$\log Z_*$ ($Z_\odot$)[c]	$\log T_*$ (Gyr)	A_V (mag)	$U-V$ (mag)	$V-J$ (mag)	S/N[a] Blue	S/N[a] Red	$t_{\rm G102}$[b] (s)	$t_{\rm G141}$[b] (s)
					MACS J1149.6+2223[d]								
00141	1.77403e+02	2.24185e+01	$1.96^{+0.01}_{-0.01}$	$11.21^{+0.13}_{-0.14}$	$0.43^{+0.18}_{-0.20}$	$-0.25^{+0.11}_{-0.12}$	$1.72^{+0.20}_{-0.15}$	$1.79^{+0.01}_{-0.01}$	$1.35^{+0.10}_{-0.07}$	4.2	24.5	9529	75987
00227	1.77407e+02	2.24162e+01	$2.41^{+0.01}_{-0.01}$	$10.94^{+0.15}_{-0.13}$	$-0.28^{+0.19}_{-0.17}$	$0.23^{+0.14}_{-0.14}$	$0.74^{+0.07}_{-0.08}$	$1.20^{+0.01}_{-0.01}$	$0.81^{+0.00}_{-0.05}$	6.8	17.4	19758	80399
					GOODS-North								
06215	1.89029e+02	6.21726e+01	$2.30^{+0.02}_{-0.02}$	$11.55^{+0.13}_{-0.13}$	$0.36^{+0.18}_{-0.17}$	$0.01^{+0.12}_{-0.09}$	$0.65^{+0.08}_{-0.09}$	$1.43^{+0.01}_{-0.01}$	$0.83^{+0.01}_{-0.01}$	2.8	8.5	5011	6117
07276	1.89306e+02	6.21791e+01	$2.50^{+0.01}_{-0.01}$	$11.65^{+0.13}_{-0.14}$	$0.14^{+0.22}_{-0.20}$	$0.20^{+0.11}_{-0.10}$	$0.32^{+0.08}_{-0.08}$	$1.50^{+0.01}_{-0.01}$	$0.77^{+0.01}_{-0.01}$	6.1	14.1	5011	5011
11470	1.89066e+02	6.21987e+01	$1.67^{+0.01}_{-0.01}$	$11.07^{+0.12}_{-0.12}$	$0.33^{+0.19}_{-0.17}$	$0.08^{+0.09}_{-0.11}$	$0.13^{+0.10}_{-0.09}$	$1.17^{+0.01}_{-0.01}$	$0.50^{+0.01}_{-0.01}$	3.0	20.1	10023	8723
17599	1.89121e+02	6.22289e+01	$2.15^{+0.01}_{-0.01}$	$11.16^{+0.15}_{-0.15}$	$0.23^{+0.25}_{-0.24}$	$0.17^{+0.12}_{-0.10}$	$0.41^{+0.10}_{-0.10}$	$1.77^{+0.12}_{-0.17}$	$0.86^{+0.03}_{-0.03}$	3.5	20.6	33482	39394
17735	1.89061e+02	6.22290e+01	$1.84^{+0.01}_{-0.01}$	$11.33^{+0.13}_{-0.13}$	$0.36^{+0.17}_{-0.17}$	$0.14^{+0.14}_{-0.16}$	$0.50^{+0.11}_{-0.10}$	$1.46^{+0.01}_{-0.01}$	$0.83^{+0.00}_{-0.00}$	3.8	23.5	5011	8123
19341	1.89087e+02	6.22367e+01	$1.86^{+0.01}_{-0.01}$	$10.98^{+0.14}_{-0.16}$	$0.02^{+0.26}_{-0.26}$	$0.29^{+0.12}_{-0.09}$	$0.04^{+0.07}_{-0.04}$	$1.19^{+0.01}_{-0.01}$	$0.57^{+0.01}_{-0.01}$	4.1	31.1	5011	38494
19850	1.89090e+02	6.22392e+01	$1.86^{+0.01}_{-0.01}$	$11.06^{+0.14}_{-0.16}$	$0.25^{+0.23}_{-0.20}$	$0.15^{+0.13}_{-0.14}$	$0.53^{+0.13}_{-0.15}$	$1.59^{+0.01}_{-0.01}$	$0.90^{+0.01}_{-0.01}$	2.1	22.2	5011	38494
22774	1.89128e+02	6.22537e+01	$2.01^{+0.01}_{-0.01}$	$11.37^{+0.12}_{-0.12}$	$0.14^{+0.19}_{-0.19}$	$0.16^{+0.12}_{-0.10}$	$0.78^{+0.10}_{-0.13}$	$1.64^{+0.01}_{-0.01}$	$0.97^{+0.01}_{-0.01}$	2.2	18.4	4811	31476
23006	1.89351e+02	6.22547e+01	$2.48^{+0.01}_{-0.01}$	$11.11^{+0.12}_{-0.14}$	$0.11^{+0.21}_{-0.20}$	$0.18^{+0.11}_{-0.10}$	$0.05^{+0.07}_{-0.05}$	$1.27^{+0.01}_{-0.01}$	$0.55^{+0.01}_{-0.01}$	2.8	8.3	5011	4911
23249	1.89064e+02	6.22560e+01	$2.37^{+0.01}_{-0.01}$	$10.89^{+0.15}_{-0.14}$	$0.10^{+0.25}_{-0.28}$	$0.21^{+0.18}_{-0.27}$	$0.34^{+0.18}_{-0.22}$	$1.28^{+0.01}_{-0.01}$	$0.56^{+0.02}_{-0.03}$	2.9	7.7	5011	15841
24033	1.89115e+02	6.22594e+01	$1.67^{+0.01}_{-0.01}$	$11.21^{+0.14}_{-0.15}$	$0.47^{+0.18}_{-0.19}$	$0.19^{+0.12}_{-0.10}$	$0.25^{+0.09}_{-0.08}$	$1.54^{+0.01}_{-0.01}$	$0.87^{+0.00}_{-0.00}$	3.2	29.2	4811	39494
33780	1.89202e+02	6.23172e+01	$1.87^{+0.01}_{-0.01}$	$11.71^{+0.17}_{-0.19}$	$0.48^{+0.22}_{-0.20}$	$0.30^{+0.08}_{-0.09}$	$0.51^{+0.12}_{-0.12}$	$1.92^{+0.01}_{-0.01}$	$1.31^{+0.01}_{-0.01}$	2.9	14.8	33282	14635
					GOODS-South								
09704	5.32857e+01	-2.78641e+01	$1.74^{+0.01}_{-0.01}$	$11.71^{+0.14}_{-0.14}$	$0.40^{+0.19}_{-0.18}$	$0.28^{+0.12}_{-0.12}$	$0.71^{+0.14}_{-0.21}$	$1.69^{+0.02}_{-0.02}$	$1.14^{+0.00}_{-0.00}$	12.5	17.4	98073	4711
23073	5.31231e+01	-2.78034e+01	$2.34^{+0.01}_{-0.01}$	$11.32^{+0.15}_{-0.18}$	$0.34^{+0.22}_{-0.21}$	$0.10^{+0.10}_{-0.10}$	$0.15^{+0.09}_{-0.09}$	$1.46^{+0.01}_{-0.01}$	$0.68^{+0.01}_{-0.01}$	5.4	11.9	0	9423
24569	5.31588e+01	-2.77972e+01	$1.90^{+0.01}_{-0.01}$	$11.17^{+0.16}_{-0.18}$	$-0.14^{+0.25}_{-0.23}$	$0.35^{+0.09}_{-0.09}$	$0.20^{+0.07}_{-0.08}$	$1.55^{+0.00}_{-0.01}$	$0.77^{+0.01}_{-0.01}$	1.0	22.6	103246	23358
31397	5.31410e+01	-2.77667e+01	$1.91^{+0.01}_{-0.01}$	$11.51^{+0.21}_{-0.21}$	$0.45^{+0.26}_{-0.25}$	$0.15^{+0.08}_{-0.10}$	$0.38^{+0.10}_{-0.08}$	$1.66^{+0.00}_{-0.00}$	$0.94^{+0.00}_{-0.00}$	5.9	32.3	0	21552
39364	5.30628e+01	-2.77265e+01	$1.61^{+0.01}_{-0.01}$	$11.53^{+0.23}_{-0.21}$	$0.17^{+0.23}_{-0.25}$	$0.44^{+0.12}_{-0.18}$	$0.22^{+0.08}_{-0.07}$	$1.62^{+0.01}_{-0.01}$	$1.09^{+0.00}_{-0.00}$	8.9	23.6	27270	8923
41021	5.31874e+01	-2.77192e+01	$2.32^{+0.01}_{-0.01}$	$11.28^{+0.12}_{-0.13}$	$0.08^{+0.16}_{-0.17}$	$0.16^{+0.10}_{-0.12}$	$0.27^{+0.09}_{-0.09}$	$1.23^{+0.01}_{-0.01}$	$0.54^{+0.01}_{-0.01}$	4.9	15.7	0	4711
41148	5.31279e+01	-2.77189e+01	$1.76^{+0.01}_{-0.01}$	$11.47^{+0.20}_{-0.21}$	$0.35^{+0.26}_{-0.26}$	$0.39^{+0.10}_{-0.10}$	$0.08^{+0.07}_{-0.07}$	$1.87^{+0.02}_{-0.02}$	$1.18^{+0.01}_{-0.01}$	3.4	10.6	23058	4611
41520	5.31527e+01	-2.77163e+01	$1.60^{+0.01}_{-0.01}$	$11.31^{+0.25}_{-0.22}$	$0.20^{+0.25}_{-0.27}$	$0.46^{+0.13}_{-0.18}$	$0.14^{+0.07}_{-0.07}$	$1.81^{+0.01}_{-0.01}$	$1.18^{+0.00}_{-0.00}$	4.7	10.2	27470	4711
43005	5.31085e+01	-2.77101e+01	$1.60^{+0.01}_{-0.01}$	$11.32^{+0.23}_{-0.22}$	$0.38^{+0.29}_{-0.25}$	$0.28^{+0.10}_{-0.09}$	$0.39^{+0.08}_{-0.08}$	$1.82^{+0.00}_{-0.01}$	$1.12^{+0.00}_{-0.00}$	3.9	13.1	23058	9323
43114	5.30624e+01	-2.77069e+01	$1.90^{+0.01}_{-0.01}$	$12.06^{+0.19}_{-0.20}$	$0.07^{+0.21}_{-0.19}$	$0.37^{+0.10}_{-0.10}$	$0.27^{+0.07}_{-0.07}$	$1.45^{+0.00}_{-0.00}$	$0.81^{+0.00}_{-0.00}$	25.2	49.4	27270	7720

Notes.
[a] Average S/Ns of grism spectral element measured at blue (3400–3800 Å) and red (4200–5000 Å) wavelength ranges.
[b] Total exposure time in G102 and G141 observations.
[c] $Z_\odot = 0.0142$ (Asplund et al. 2009).
[d] Stellar masses are corrected for magnifications by the foreground cluster.

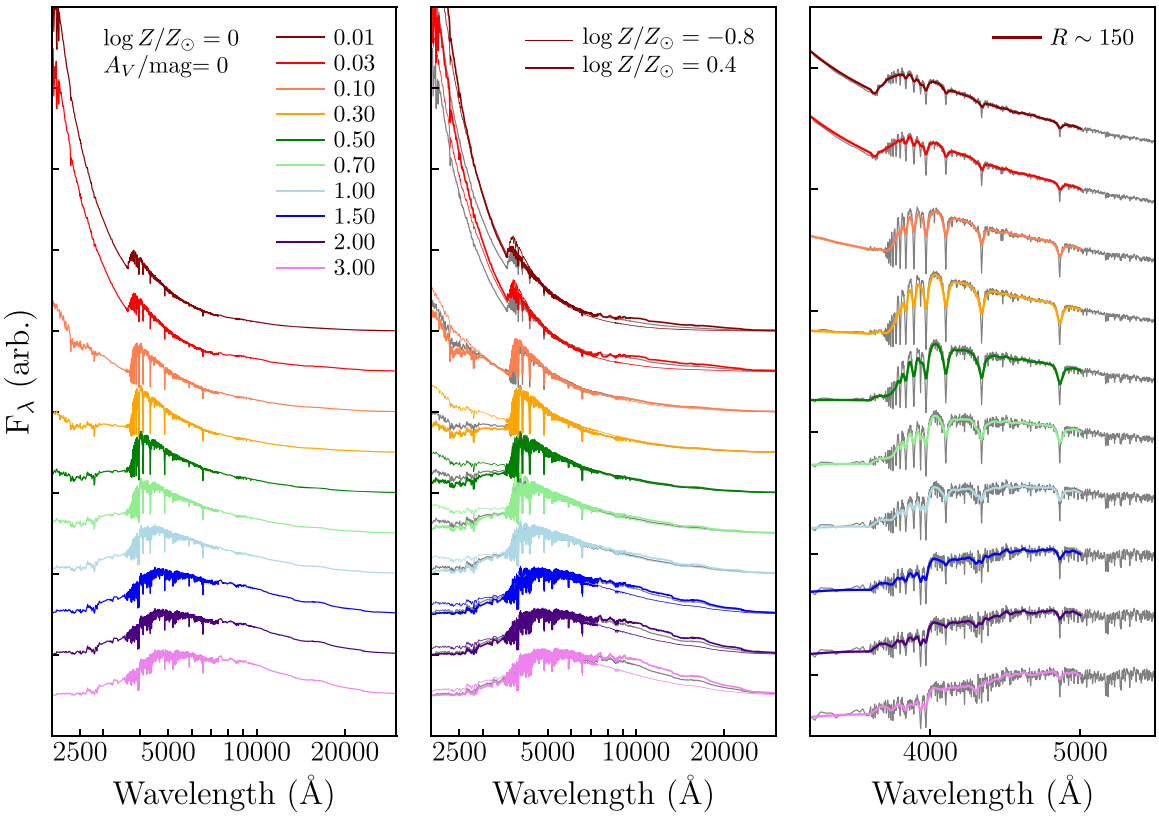

Figure 2. Left: original spectral templates used in fitting. Only those with $\log Z_*/Z_\odot = 0$ without dust attenuation ($A_V = 0$) are shown here, with arbitral shifts in flux. Middle: same as left panel (gray solid lines) but also showing templates with different metallicities for comparison (thin lines for $\log Z_*/Z_\odot = -0.8$; thick lines for $\log Z_*/Z_\odot = 0.4$). Templates are normalized at 5000 Å. Right: same as left panel (gray solid lines) but also showing spectra that are degraded to match the observation that accounts for the G102/G141 resolutions and source morphology ($R \sim 150$; colored lines).

flexibility in fitting templates and allows a reasonable estimate of uncertainty in SFHs (Section 5.3).

It is noted that metallicity-sensitive lines (such as Fe and Mg) are not measured at our spectral resolution. Our method instead relies on the entire spectral shape, with grism spectra and wide broadband photometry, that spans from the NUV, to the optical (that are sensitive to age), to the NIR (to metallicity) wavelength range (see Figure 2).

Templates generated with MIST are uniformly set to the solar-scaled abundance (Asplund et al. 2009; i.e., $[\alpha/\mathrm{Fe}] = 0$). It is noted that galaxies at high z may have an α-enhanced chemical composition (e.g., Onodera et al. 2015; Kriek et al. 2016), as found in local early-type galaxies (e.g., Thomas et al. 2005; Walcher et al. 2015). In fact, enhancement of the α-element has a similar effect as that of iron in UV and NIR continuum slopes (e.g., Vazdekis et al. 2015), while our low-resolution spectra cannot capture a detailed difference in each absorption line (i.e., Lick indices), and both abundances are degenerated in our total metallicity measurement, $\log Z_*/Z_\odot$.[13] As such, our total metallicity should remain similar to those with, e.g., α-enhanced templates (see also Walcher et al. 2009).

We assume a Salpeter (1955) initial mass function and Calzetti et al. (2000) dust law, where the dust attenuation, A_V, is a global parameter that is applied to all age templates equally. Redshift is set as a free parameter at this step but within the 3σ range estimated in the previous step. In sum, the fits have $10 \times 2 + 1 + 1 = 22$ free parameters.

The degree of freedom of our fitting is worth noting. The number of spectral data points for each of our galaxies is >200 (with ∼16 for broadband photometric data points), where the spectral element is set to 45 Å in this study. Considering the correlation due to morphology (which is ∼104 Å for the mean size of our galaxies, $r \sim 0.''3$) and the large number of parameters, our spectra still have ∼100 independent data points.

Our updated method here has a few advantages over Morishita et al. (2018). First, it is more flexible than an a priori assumption of the SFH and robust to systematic bias in derived parameters (e.g., Wuyts et al. 2012). Second, it is flexible to a complex shape of SFHs, such as those with multiple bursts and sudden declines (e.g., Boquien et al. 2014). An SFH with multiple peaks cannot be reproduced by the exponential declining model (see Section 4). Third, it is flexible to the metallicity evolution. Methods with functional forms often have a fixed metallicity over the entire history. In our method, each template at a different age has a flexibility in metallicity as a free parameter that also provides metallicity enrichment histories, though the uncertainty in each age is typically large for most of our data sets in this study (Appendix A).

3.2. SED Parameter Exploration

The combination of templates is controlled by changing each amplitude, a_i, as free parameters during the fit. A challenging part is the large number of parameters (over a dozen, compared to $\lesssim$5 parameters with functional SFHs), which could be

[13] Total metallicity is often inferred with $[\mathrm{Z/H}] = [\mathrm{Fe/H}] + A[\alpha/\mathrm{Fe}]$, where $A \sim 0.9$ depending on abundance ratios (e.g., Trager et al. 2000; Jimenez et al. 2007).

trapped in local minima. To sufficiently, yet efficiently, explore the parameter space, we adopt the Markov chain Monte Carlo (MCMC) method.

The fitting process of `gsf` is twofold: (1) initial redshift determination based on visual inspection of absorption lines and (2) MCMC realization to estimate the probability distribution for all parameters.

First, `gsf` determines the redshift by fitting the model templates to the observed grism spectrum. At this point, it only generates model templates in the wavelength range of grism spectra to minimize the computational cost. The templates are convolved to the resolution of the spectra with a Moffat function derived from the observed source morphology for spectra. It searches the best-fit redshift by minimizing χ^2, as well as visual inspection, to avoid catastrophic errors. During the visual inspection, we rely on the major absorption lines in the observed range (i.e., Hδ, Hγ, Hβ), and thus those without clear absorption features (i.e., low S/N) are discarded here. At this step, `gsf` also determines the scale of the G102/G141 spectra so that each matches to the broadband photometry in F105W and F140W at a given template. The added scale for our sample is small ($\lesssim$10%) thanks to the accurate sky-background estimation in Grizli.

Then, `gsf` generates a template library at the redshift determined in the previous step and fits SEDs at the entire wavelength range. Here `gsf` fits the observed spectra and broadband photometry simultaneously by using *emcee* (Foreman-Mackey et al. 2013), as in Morishita et al. (2018). Redshift is also explored at this step by shifting and refining the template wavelength grid at a proposed redshift of each MCMC step. Emission lines, if detected, are masked during the fit. Those lines are modeled with a Gaussian function after subtracting the best-fit SED template to estimate the line flux and equivalent width (EW; Section 4.2).

We set the number of walkers to 100 and the number of realizations, $N_{\rm mc}$, to 10^5. We adopt a uniform prior for each parameter over the parameter ranges, $a_i \in [0: 1000]$, $\log Z_*/Z_\odot \in [-0.8: 0.6]$, and $A_V/{\rm mag} \in [0: 4]$. The effect of the amplitude prior (i.e., SFH; see Figure 9) should be minimal due to the wide constrained range, as seen in the simulation in Appendix A. While some previous studies set a prior in metallicity histories from the local mass–metallicity relation (e.g., Pacifici et al. 2016; Leja et al. 2018; i.e., increasing metallicity as a function of time), we find that our result reproduces this behavior without such priors. However, the age–metallicity degeneracy could also mimic this trend. The test using an artificial SED data set in Appendix A revealed that, while the trend is fairly reproduced, nonnegligible scatter occurs in derived metallicity at each age pixel, and the metallicity histories of individual galaxies remain less promising (see also below).

During the fit, we let *emcee* run with the parallel tempering sampling, with $n_{\rm temp} = 5$. With this, *emcee* samples the parameter spaces but with $n_{\rm temp}$ samplers in parallel. Each sampler has a different value for the temperature parameter in a Metropolis–Hastings step; i.e., a higher temperature makes a larger step in a parameter space. With this, the sampler suffers less from the local minima. We note, however, that a sufficient number of MCMC realizations (which is inferred from a simulation test for our case) is necessary to estimate reliable uncertainties. Otherwise, the derived uncertainties would become inappropriately small, with possible biases in the best-fit values (see Cid Fernandes 2018).

The first half realization of the sampled chain is discarded to avoid biased results from initial input values. We take the 50th and 16th/84th percentiles of marginalized distributions as the best fit and uncertainty range.

To check the reliability of the SED parameters and reconstructed histories, we conducted a simulation test with a mock data set (Appendix A). From the test, we found that the scatter in the reproduced amplitude and metallicity at each age bin is $\sim$0.5 and $\sim$0.25 dex, respectively. To account for this, we add the estimated scatter to the reconstructed star formation and metallicity histories, which are also propagated to other SED parameters (Section 4.3).

4. Results

4.1. 24 Galaxies as the Final Sample

In Figure 1, we summarize the basic parameters of the final sample galaxies. The final sample consists of 2, 12, and 10 galaxies from M1149, GDN, and GDS, after the rest of the initial samples are visually discarded because of poor redshift fitting quality (Section 3.2). Due to the partial coverage of the grism observations, as well as random contamination from neighboring galaxies, our samples have nonuniform exposure times in the grism observation and S/N (Table 1). The two galaxies in M1149 are those previously reported by Morishita et al. (2018).

In panels (a) and (b) of Figure 1, we plot the distribution of our sample galaxies in redshift–stellar mass/F140W–magnitude spaces, respectively. Due to the increasing sensitivity of the G141 grism with wavelength, galaxies at higher redshift are fainter in F140W than those at lower redshift. In (c), we show the *UVJ* color–color diagram for diagnosing galaxy quiescence (e.g., Williams et al. 2009). While most of the sample locates within the passive category, there are six galaxies that fall below the passive/star-forming boundary in $U - V$. These galaxies are in transition between star-forming ($U - V \lesssim 1$) and passive ($\gtrsim$1.4), akin to the green valley (or "quenching") galaxies in the local universe (Kauffmann et al. 2003; Schawinski et al. 2014).

The characteristic age, which is about the time since a system reached its half-mass, of each galaxy represents the mass-weighted value from the reconstructed SFH,

$$T_* = \sum_i t_i a_i \Psi_i / \sum_i a_i \Psi_i, \quad (1)$$

where t_i is the median age, a_i is the best-fit amplitude, and Ψ_i is the mass-to-light ratio of the ith template. An estimated error for each amplitude from the simulation test (Appendix A) is added in quadrature. The typical error in mass-weighted age is $\sim$0.2 dex.

The derived parameters are summarized in Table 1. It is noted that the estimated errors in stellar mass are larger ($\sim$0.2 dex) than those listed in the original catalogs ($\lesssim$0.1 dex; Skelton et al. 2014). This is due to the fact that our fitting accounts for an additional uncertainty originated in the flexibility of SFHs. While one can implement this by, e.g., repeating a SED fitting analysis with different SFHs (Wuyts et al. 2012; Morishita et al. 2015), the stellar-mass measurement in the original catalog (as well as many others) is based on one functional form for SFHs; thus, the quoted errors are solely from photometric error and redshift (see also Appendix B).

4.2. Diagnostics from the SED Shape

In the left panels of Figure 3, we show the SEDs of the sample galaxies with the best-fit templates. Our galaxies are well characterized with $k + a$ (Dressler & Gunn 1983) and quiescent spectra, as is expected from the sample selection and redshift range. Deep spectra successfully capture the spectral features of these types of galaxies, such as absorption lines and the 4000 Å/Balmer break. The wide broadband coverage well captures the spectral features, such as a blue UV slope from a young population ($\sim$1 Gyr) and near-IR excess from an old population ($\gtrsim$2 Gyr), that is consistent with the derived mass-weighted age (T_*; see Table 1).

Six galaxies have moderately detected ($\sim 1.5\sigma$) weak emission lines, such as [O II], Hδ, Hγ, and Hβ, which is a signature of ongoing star formation. We first fit each emission line with a Gaussian after subtracting the best-fit spectrum. The EW is then measured with the total flux from the Gaussian fit and the best-fit template as a continuum. For the [O II] line, we use Equation (3) in Kennicutt (1998) to estimate the SFRs. For the Hβ line, we use a recombination coefficient in Case B (Osterbrock 1989) and then Equation (2) in Kennicutt (1998). These weak emission lines indicate a specific SFR (SFR/M_*) of $\lesssim 10^{-10}$ yr^{-1}. While the detection is tentative, the low level of star formation activity is also observed in previous findings (e.g., Belli et al. 2017), possibly providing more detailed pictures of quenching mechanisms (Section 5). Two of the emission-detected galaxies (IDs 19341 and 19850) have strong [O III] lines (4959 + 5007 Å) with a relatively weak Hβ line, suggesting the existence of active galactic nuclei (AGNs). While Hα and N II are beyond our wavelength coverage, the line ratio of the Hβ and [O III] lines ($\log$Hβ/O III = 0.15 ± 0.03 and 0.13 ± 0.05) implies that these galaxies are AGNs in the mass-excitation diagram (Juneau et al. 2014).

On the other hand, we find six galaxies that consist of very old populations with mass-weighted ages $\gtrsim$2 Gyr and dominate the high-mass end of our sample. While such massive galaxies are rare ($n \sim 3 \times 10^{-5}$ Mpc^{-3}; Muzzin et al. 2013; Tomczak et al. 2014), it is also true that some ancient galaxies that formed a long time ago have more chances to experience ex situ processes, e.g., merger and gas accretion, especially at this high redshift. Such ancient galaxies would be smuggled into the younger population and become indistinguishable when seen with, e.g., a light-weighted age. Our reconstructed SFHs have the ability to investigate this.

4.3. SFHs

In the right three columns of Figure 3, we show the reconstructed SFHs, mass accumulation histories, and metallicity enrichment histories for our sample galaxies. We reconstruct SFRs in each time bin by dividing the amount of stellar mass formed (including the lost mass by the time of observation) by the bin length. Thus, the SFRs at each bin represent its average values over time ($\sim$30 Myr for the youngest template to $\sim$1 Gyr for the oldest one). It is also noted that the derived SFR cannot distinguish between in situ (stars formed in the system) or ex situ (those obtained via mergers).

Some galaxies are worth highlighting. For example, ID 43114, the most massive galaxy in our sample ($\log M_*/M_\odot \sim 12$), had already formed about 50% of the final mass at $z \gtrsim 5$ ($\sim$2 Gyr ago). The galaxy was at low star formation activity for $\sim$1.5 Gyr and then started active star formation ($\sim 1000\, M_\odot$ yr^{-1}) at $z \sim 2.5$, $\sim$300 Myr ago. The significant star formation is comparably as high as that of submillimeter galaxies at this redshift (Younger et al. 2007; Tacconi et al. 2008). In fact, its morphology shows a tidal feature with two close objects at the outer part, suggesting a recent (major) merger. Its star formation activity seen in the reconstructed SFH is consistent with a typical merger timescale at this redshift (Lotz et al. 2011; Snyder et al. 2017). Interestingly, the dust attenuation of this galaxy is relatively low ($A_V \sim 0.3$ mag) compared to typical submillimeter galaxies and starburst galaxies at this redshift ($A_V > 3$ mag; Riechers et al. 2013; Toft et al. 2014), suggesting that post-processes might have cleared a large amount of dust.

Other galaxies with clearly disturbed morphology (e.g., IDs 09701 and 00141) also show recent intense star formation at $\lesssim$1 Gyr before $t_{\rm obs}$ that may provide an independent constraint on the merger timescale and induced star formation activity from follow-up kinematical studies.

Interestingly, many of our galaxies show extended star formation activity to $\sim$0.3 Gyr before their observed redshifts. This differs from the previous understanding of massive early-type galaxies, whose star formation activity was believed to decline rapidly or become truncated after forming the bulk of the stars in a very short time. While our sample size here is too small to generalize this (and is also possibly biased toward compact morphology), it is curious to see how previously adopted functional form SFHs behave toward such an extended feature, as well as other features like dual-peak SFHs.

We repeat the same analysis but with functional forms of SFHs to compare with our SFHs. In Figure 3, we show the best-fit SFHs obtained with a functional form for the SFH τ model with $\propto A \exp[-(t - t_0)/\tau]$. In addition to τ and t_0, we allocate one parameter for metallicity and one for dust attenuation. In most cases, it is clear that SFHs derived from `gsf` cannot be reproduced by the τ model. For example, the τ model cannot capture the dual peaks observed in some of our galaxies (e.g., ID 43114). The τ model also fails to capture extended star formation at both the young and old age sides. This is due to the fact that the functional SFH is light-weighted, where the best-fit parameters are more sensitive to differential amounts of light. Such a qualitative discrepancy in fact results in a quantitative discrepancy in the best-fit parameters, with χ^2/ν systematically larger than `gsf` (see also Carnall et al. 2018, who argued limitation by a functionally defined SFH). Appendix B summarizes the comparison of the τ model, as well as results with the delayed τ model.

The metallicity histories shown in Figure 3 represent mass-weighted accumulated metallicity,

$$Z_j = \sum_{i}^{i \leqslant j} Z_i a_i \Psi_i / \sum_{i}^{i \leqslant j} a_i \Psi_i, \quad (2)$$

where i covers older age pixels than j. Individual metallicity values at each age pixel often suffer from large uncertainty ($\sim$0.25 dex; Appendix A), while this is partially attributed to a small amount of light in those age pixels (i.e., small a_i). We therefore avoid discussing metallicity enrichment histories of individual galaxies and instead focus on (more robust) total metallicity in this study (Section 5.3). It is noted, however, that having parameters for metallicity at each age bin allows flexibility in fitting and a more reasonable estimate (i.e., larger error bars) in SFHs and SED parameters.

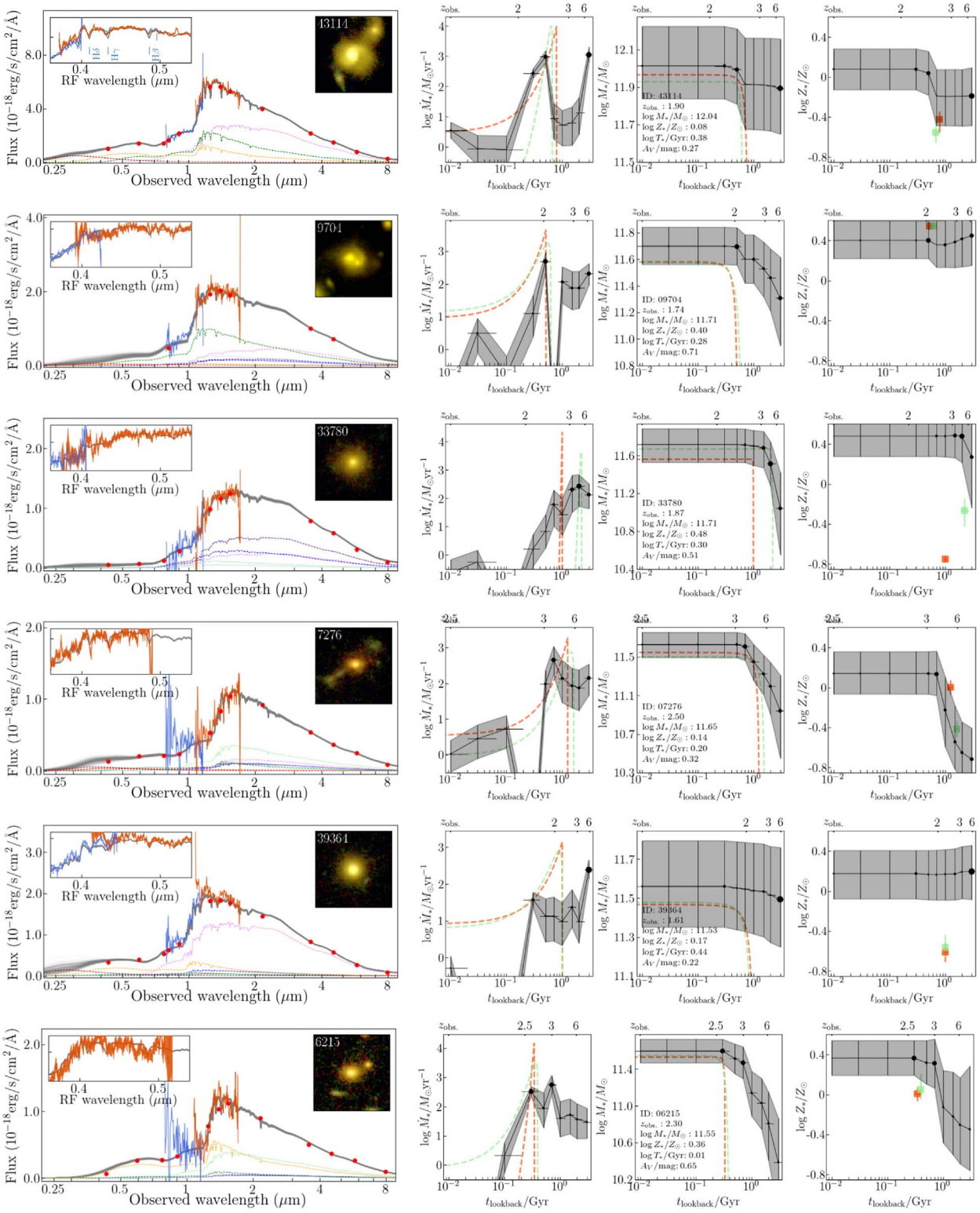

Figure 3. Results of SED fitting for the final 24 galaxies sorted in order of stellar mass. First column: observed spectra (blue and red lines with error bars for G102 and G141) and photometric data points (red circles with black error bars). The best-fit templates of each age bin (dashed lines, colored as in Figure 2) and the sum (gray dashed lines) are shown. Different total templates randomly reproduced from the MCMC are also shown. Pseudo-color stamps (F160/125/814W for R/G/B; 4.8 × 4.8 arcsec^2) are shown. Emission lines, when detected, are shown with fitted Gaussian curves in the inset (blue solid lines). Second column: reconstructed SFHs, with the 50th (black circles) and 16th/84th percentile ranges (gray hatched region) shown. The size of the symbols at each age bin (black circles) represents the amplitude. The best-fit SFHs with functional forms (τ and delayed-τ models) are also shown for comparison (red and green dashed lines). Third column: same as second column but for stellar-mass accumulation histories. Fourth column: same as second column but for mass-weighted stellar metallicity. The best-fit metallicities derived by the two functional models are shown at the best-fit age (red/green squares with error bars).

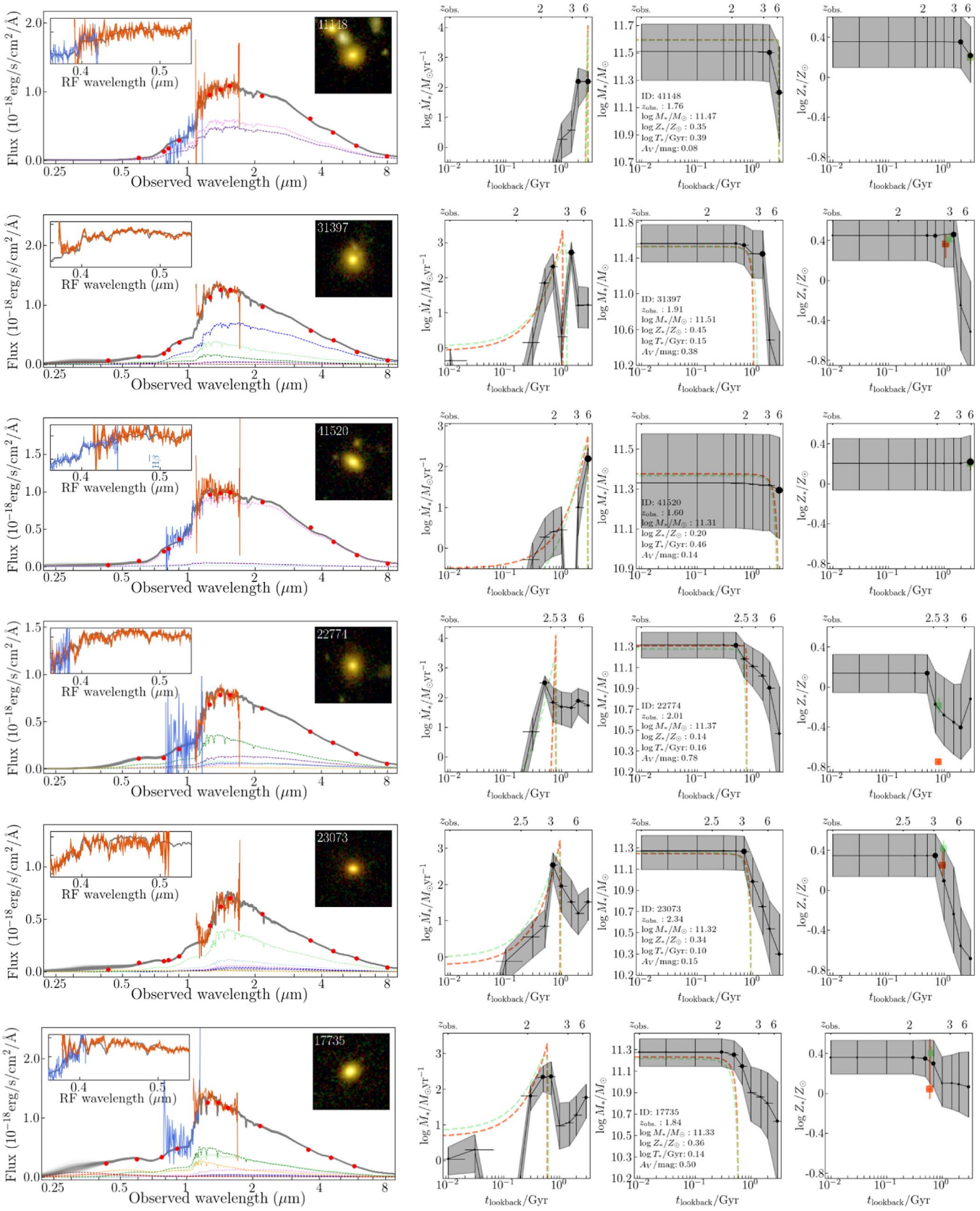

Figure 3. (Continued.)

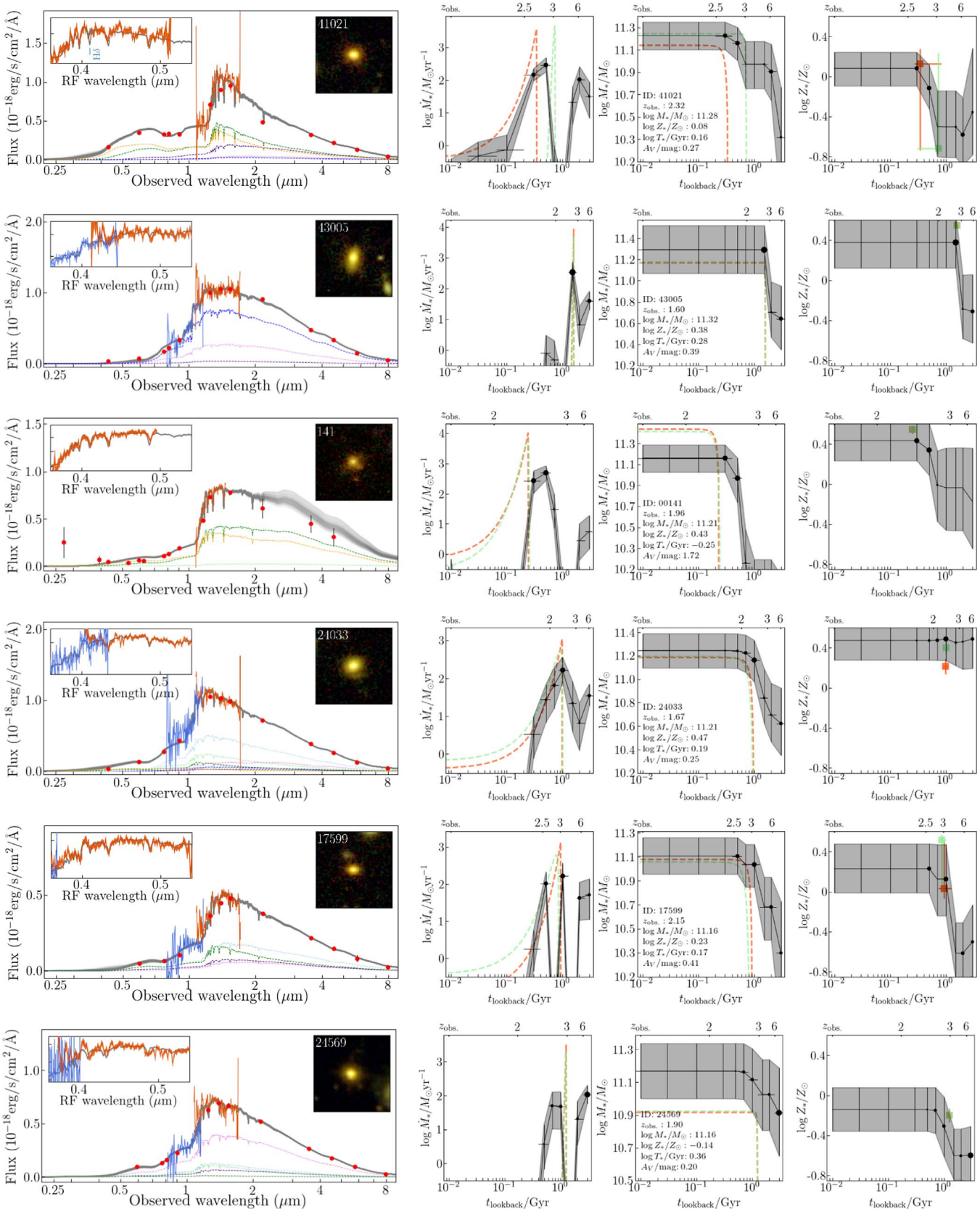

Figure 3. (Continued.)

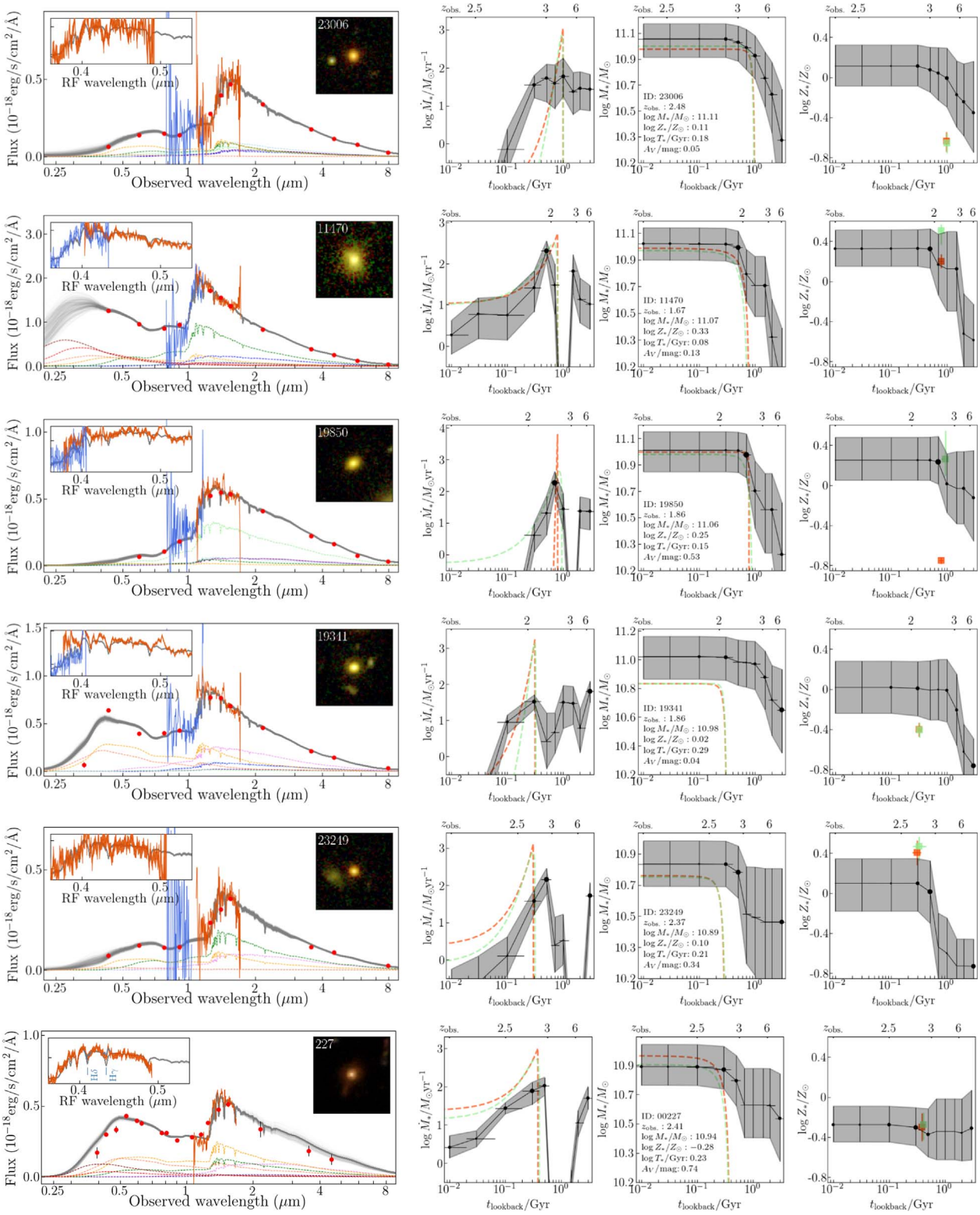

Figure 3. (Continued.)

5. Discussion

5.1. Timescale of Star Formation

In Figure 4, we summarize the mass accumulation histories as a function of lookback time from the observed redshift. Most of our sample galaxies formed >50% of their extant mass by ∼1.5 Gyr prior to the observation ($z \sim 2.5$–5, depending on the observed redshift), which is quantitatively consistent with recent studies at similar and higher redshift (Domínguez Sánchez et al. 2016; Estrada-Carpenter et al. 2018; Schreiber et al. 2018; Belli et al. 2019).

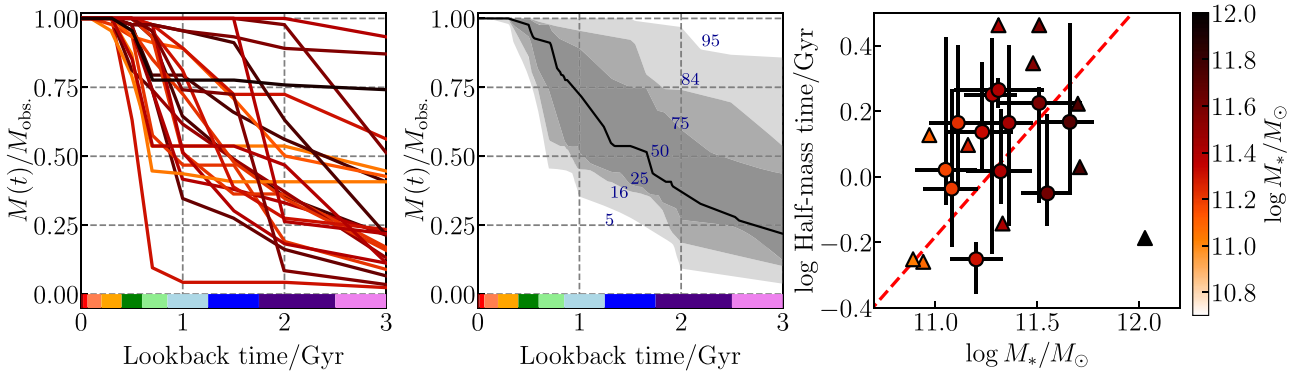

Figure 4. The SFHs as a function of lookback time from the observed redshift. Left: cumulative stellar-mass evolution of individual galaxies. Color corresponds to the final stellar mass. Ages that correspond to the SED templates are indicated with color bars at the bottom (colored as in Figure 2). Middle: summary of individual cumulative stellar-mass evolution, where contour boundaries and the line correspond to the 5th/16th/25th/75th/84th/95th percentiles and median. Right: half-mass time (lookback time from $t_{\rm obs}$; t_{50} in the main text) as a function of observed stellar mass. Those with lower limits are shown with triangles. Symbol color corresponds to the final stellar mass. A positive correlation, $\log t_{50}/{\rm Gyr} \propto 0.5 \log M_*/M_\odot$, is seen (red dashed line; i.e., downsizing).

We estimate the half-mass time, t_{50} (lookback time from the observed redshift), from individual SFHs in the right panel of Figure 4. It is estimated in each step of the MCMC, and thus its uncertainty represents those in SFHs and individual age bin widths. For some galaxies, we only estimate the lower limit, as >50% of stellar mass is in the oldest template. Higher-resolution spectra by, e.g., the *James Webb Space Telescope* are required to reveal ancient histories at a higher time resolution.

Still, we see a trend where more massive galaxies form earlier, known as downsizing (Cowie et al. 1996; Heavens et al. 2004; Treu et al. 2005b), with a linear fit of $\log t_{50}\,{\rm Gyr}^{-1} \propto 0.5 \log M_*/M_\odot$. The measured standard deviation (∼0.16 Gyr) is comparable to the redshift range of our galaxies ($\Delta \log T \sim$ 0.14 Gyr). The fact that the downsizing trend exists in the early time of the universe provides hints about the galaxy evolution at even earlier epochs, when they were star-forming galaxies, and how observed luminous galaxies form (Zitrin et al. 2015; Oesch et al. 2016) in relation to, e.g., their environments (Harikane et al. 2019).

5.2. Stellar Mass–Metallicity Relation at z ∼ 2

The stellar mass–metallicity relation is a key diagnostic of galaxies' chemical and mass maturation histories. The relation encodes the coevolution of stellar mass and chemical enrichment among galaxies and provides an independent clue to the past evolution of SFHs. The relation is known to hold from the local universe (Gallazzi et al. 2005; Panter et al. 2008; González Delgado et al. 2014) in a wide range of mass (Kirby et al. 2013), up to $z \sim 1$ (Choi et al. 2014; Gallazzi et al. 2014; Leethochawalit et al. 2018). Beyond the redshift, however, it is still studied with a small sample of galaxies and is not clear (Onodera et al. 2012, 2015; Kriek et al. 2016; Morishita et al. 2018). While the relation in gas-phase metallicity at $z \gtrsim 1$ may suggest that the relation may remain universal to a higher redshift (Tremonti et al. 2004; Kewley & Ellison 2008; Maiolino et al. 2008; Mannucci et al. 2010; Yabe et al. 2014; Zahid et al. 2014; Onodera et al. 2016; Wang et al. 2017), the observed scatter is still large due to a selection bias, different tracer of metallicity, and different physical state of gas-phase metallicity (e.g., Andrews & Martini 2013). Here we overview the relation at $z > 1.6$ for the first time.

In Figure 5, we show the distribution of our galaxies in the stellar mass–metallicity plane. The metallicity here is a mass-weighted value, as for the age (Equation (1)). While no significant mass dependency of the metallicity is observed, this is because our galaxies occupy the high-mass end and do not span a wide mass range. In fact, the flattening behavior at a high-mass range is consistent with the relation at lower redshift (Gallazzi et al. 2005, 2014). The observed metallicity is significantly high (median of $\log Z_*/Z_\odot \sim 0.25$) and tight (scatter of ∼0.15–0.25 dex around the median), which is comparable to the observed scatter in the local relation.

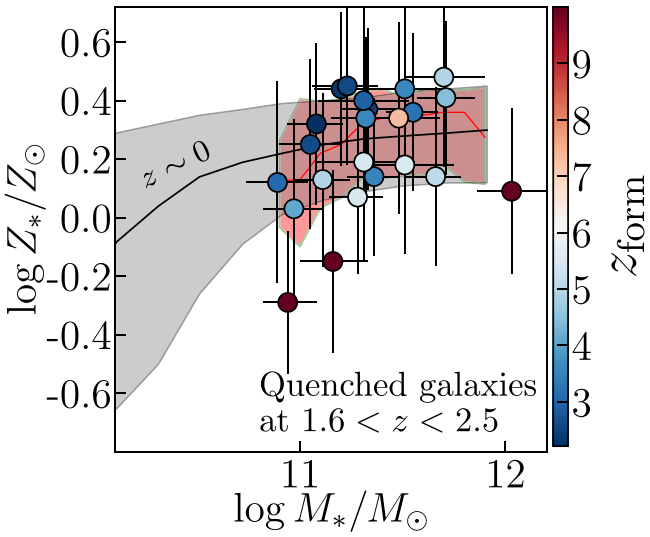

Figure 5. Stellar mass–metallicity relation of 24 galaxies in this study (circles). The symbols are color-coded by formation redshift, $z_{\rm form}$. A running median (∼0.25) is shown with the 16th–84th percentile range (red line with shaded region; $\sigma_{\log Z} \sim 0.15$). Most of our galaxies at $z \sim 2$ already have values consistent with the local metallicity value (the gray shaded region represents the 16th–84th percentile range from Gallazzi et al. 2005, with a calibration shift for +0.15 dex; Section 5.2).

While challenging, it is still of particular interest to compare our metallicity measurement with the local relation. To do this, we need to calibrate the absolute value of metallicity. Although Gallazzi et al.'s (2005) stellar-phase metallicity measurement is based on the total metallicity, as in this study (as opposed to an element abundance–based measurement), there is a systematic difference due to the adopted isochrones (MIST versus Padova), each of which has a different definition of solar metallicity ($Z_\odot = 0.0142$ versus 0.0190). We correct this by

applying a +0.15 dex offset to the Gallazzi et al. (2005) measurement.

Another systematics is the α-abundance of the template. While the templates used in these studies are set to the solar composition (Choi et al. 2016), it is not clear, due to the low resolution of our spectra, if this metallicity is enhanced by α-elements, despite its completely different origin from iron (e.g., Thomas et al. 2005). In fact, Leethochawalit et al. (2018) recently found an ∼0.16 dex decrease in the iron abundance of massive galaxies at $z \sim 0.4$ compared to the local value. Given the time of the universe (when iron was relatively deficient) and short timescale of star formation for our galaxies (where α-elements are enhanced), the high $\log Z_*/Z_\odot$ values may represent the α-enhancement, as is the case for the $z \gtrsim 1$ galaxy (Onodera et al. 2015; Kriek et al. 2016). However, due to the definition of the total metallicity used here and Gallazzi et al. (2005), $[\mathrm{Z/H}] \sim [\mathrm{Fe/H}] + 0.94\ [\alpha/\mathrm{Fe}]$ (Thomas et al. 2003), changing the template to the α-enhanced ones should result in minor differences in this comparison.

While keeping these systematics in mind, we find that most of our galaxies are already on the local relation, with a median measured for the entire mass range of $\log Z \sim 0.25$. The scatter around the median is revealed to be small ($\sigma_{\log Z} \sim 0.15$). This implies that chemical enrichment of at least some massive galaxies have already been completed within the first ∼3 Gyr of the universe. We revisit this in the following section.

Two galaxies fall below the median relation (IDs 00227 and 24569). The former, which was reported and discussed in Morishita et al. (2018), shows rather extended SFHs with small metallicity values over the entire history. According to its undisturbed morphology and gradual mass increase, seen in Figure 3, accretion of low-mass satellites (i.e., metal-poor) or late-time star formation triggered by the infall of pristine gas may explain the observed properties, rather than more dramatic episode involving, e.g., major merger. Detailed investigation of its inner structure and chemical composition at a higher angular resolution would provide further insight into its enrichment evolution (e.g., Abramson et al. 2018; Wang et al. 2018).

The other galaxy, on the other hand, shows a rapid assembly of mass. Since the galaxy formed a large fraction (∼50%) of its current mass at $z \gtrsim 6$, its low metallicity is consistent with the cosmic metal enrichment (e.g., Lehner et al. 2016), as well as an observed rapid decrease in gas-phase metallicity at a given mass (Troncoso et al. 2014; Onodera et al. 2016). Given the time left to $z \sim 0$, metal-poor galaxies like ID 24569 would possibly be enriched in metallicity by, e.g., mergers and recycled gas and may sneak into the local average population (see discussion in Morishita et al. 2018).

It is noted that the systematic uncertainty in α-enhancement would not explain the small value in $\log Z_*/Z_\odot$, as both iron and α-abundances need to be significantly low.

5.3. Redshift Evolution of Stellar-phase Metallicity

While our mass–metallicity relation indicates that galaxies are already enriched to the value at present day (Figure 5), their origin and observed scatter, especially those with small metallicity, are yet to be investigated. We investigate the redshift evolution of total metallicity as a population by considering the formation time (z_{form}), which is derived with the mass-weighted age and observed redshift, $T_{\mathrm{form}} = T(z_{\mathrm{form}}) = T(z_{\mathrm{obs}}) - T_*$ (i.e., lookback time to the half-mass time).

In Figure 6, we show the distribution of metallicity as a function of T_{form}. A clear correlation between T_{form} and observed metallicity is observed. A linear regression reveals a slope of $\log Z_*/dt \sim 0.20 \pm 0.08$ dex Gyr^{-1}, with a standard deviation of ∼0.15 dex. Our result shows the metallicity enrichment happening in this class of massive galaxies, whose metallicity already reaches the local value at $z \sim 3$, for the first time.

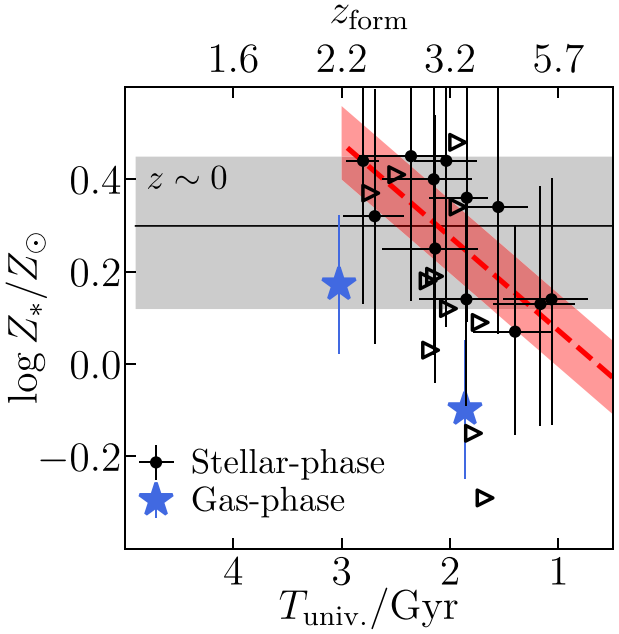

Figure 6. Observed stellar metallicity as a function of the formation time, or z_{form} (black circles/triangles for those with an upper limit in age estimate). The linear fit with a slope (0.20 ± 0.08 dex Gyr^{-1}) and standard deviation ($\sigma \sim 0.16$ dex) is shown (red dashed line and shaded region). The local metallicity value of massive galaxies ($\log M_*/M_\odot \sim 11.5$; Gallazzi et al. 2005) is shown for comparison (gray shaded region). Gas-phase metallicity measurements of massive galaxies ($\log M_*/M_\odot = 11.5$) at similar redshifts (Maiolino et al. 2008) are shown (blue stars), after being calibrated to the local Z_* measurement (Section 5.3).

One may suspect that this is an artifact from the age–metallicity degeneracy. While some galaxies show a weak correlation between the two parameters, the degree of correlation is much smaller than the quoted error bars in Figure 6. Our simulation test also revealed that the total metallicity/age are reproduced reliably enough for the observed trend (see Appendix A).

Interestingly, the linear regression suggests that the metallicity even exceeds the local value of the most massive galaxies in Gallazzi et al. (2005) by ∼0.1 dex. As noted before, our sample galaxies are biased to compact, high-density galaxies due to the high-S/N requirement for the SED fitting. While compact massive galaxies are rare at $z \sim 0$ (Taylor et al. 2010; but see also Poggianti et al. 2013), following events such as minor mergers/second bursts in the following ∼10 Gyr would resolve the tension. We discuss this in the following section.

In Figure 6, we also plot gas-phase metallicity measurements of star-forming galaxies in a similar redshift range for comparison. We use the formula derived in Maiolino et al. (2008) at the same mass ($\log M_*/M_\odot = 11.5$). We match the gas-phase metallicity measurement to the stellar metallicity at $z \sim 0$. While comparing the absolute values of two different metallicities is extremely challenging due to a number of uncertainties in each measurement (e.g., Sánchez et al. 2017; Bian et al. 2018), the matching process is reasonable for our purpose here, i.e., comparison of relative differences at $z \sim 0$ and 2.

It is interesting to find an offset of ∼0.2–0.3 dex between those metallicity measurements in our redshift range that may give us a clue to how massive quenched galaxies enrich

metallicity in this redshift. Assuming those star-forming galaxies are star-forming counterparts of our passive galaxies, the observed gap has to be resolved between the half-mass time (i.e., T_* ago) and the time when they are observed as quenched galaxies.

One possible explanation is the continuation of low-level star formation activity in a closed box, or "strangulation" (Larson et al. 1980; Peng et al. 2015). Peng et al. (2015) demonstrated in their chemical evolution model that the observed offset in metallicity of local massive star-forming and passive galaxies ($\sim$0.1 dex) can be explained by strangulation. Rapid cessation of star formation by AGN/stellar feedback would instead reproduce a similar metallicity for the two populations. Recent observations, in fact, revealed low star formation activity (e.g., Belli et al. 2017; Gobat et al. 2017), as well as remaining gas (Gobat et al. 2018, but also Sargent et al. 2015; Bezanson et al. 2019), in massive quiescent galaxies at $z \sim 2$, implying the continuation of star formation after they have formed a large amount of stars, or quenched. In addition, the closed-box enrichment seems to be a good agreement with the observed individual SFHs (Figure 3), independently supporting our speculation here.

However, more detailed chemical modelings would be required to reach a conclusion. For example, the observed gap may also be attributed to the dilution of gas-phase metallicity by infalling pristine gas, while there is no gas infall in quenched galaxies due to virial shock heating (e.g., Birnboim & Dekel 2003). If this is the case, it is suggested that galaxy quenching may be largely caused by termination of gas infall (e.g., Feldmann & Mayer 2015), while it is not likely that cutting the gas supply would result in extended SFHs as we observe here. A sophisticated chemical modeling with a panchromatic data set, including gas-mass measurements, would be required for further understanding.

5.4. Following Evolution to z $\sim$ 0

We have found that our galaxies are already enriched in metallicity, located on the local mass–metallicity relation. Given the amount of mass and its quiescence, it is within our interests to investigate how these galaxies will evolve to the local population. While it is challenging to predict their exact descendant population at low redshift (as described in the introduction), it is still worth describing their possible paths and mechanisms.

In particular, many members of our sample are compact, high-density galaxies ($\langle r_{\rm eff}\rangle \sim 2$ kpc; possibly due to the selection bias toward high S/Ns). Compact galaxies at these redshifts are often debated in terms of size evolution, where the observed size is $\sim$3–5 times smaller than galaxies at $z \sim 0$ with similar masses (Trujillo et al. 2007; Morishita et al. 2014; van der Wel et al. 2014). While there is still much debate as to whether (all of) these galaxies would follow such a significant size evolution (Newman et al. 2012; Nipoti et al. 2012; Poggianti et al. 2013; Belli et al. 2017), minor merger is a popular mechanism that can efficiently increase their sizes (e.g., Hopkins et al. 2009; Naab et al. 2009; Oser et al. 2010; van Dokkum et al. 2015; Morishita & Ichikawa 2016).

The scenario appears to consistently work for our result of metallicity, where accretion of low-mass galaxies (which are less metal-enriched, as expected from the mass–metallicity relation) would dilute the system's metallicity to the consistent value. For example, approximately five minor mergers, with 1/10 the mass of the host and metallicity inferred from the relation at $z \sim 2$, would lower the host metallicity by $\sim$0.1 dex, being consistent with the local value. The metallicity gradient observed at $z \sim 0$ (e.g., González Delgado et al. 2014; Martín-Navarro et al. 2018) is independent evidence that such high-z metal-rich galaxies would become cores while the accreted component locates the outer part of local massive galaxies. The integrated metallicity is instead an average value of the whole system; thus, metallicities observed in the local relation should be lower than those observed at higher redshifts.

Infall of metal-poor gas associated with minor merging satellites (e.g., Torrey et al. 2012) or direct infall from the cosmic web (Dekel & Birnboim 2006) would also dilute the system's total metallicity by inducing the second burst. While it is not clear if the scenario reproduces the observed metallicity gradient at $z \sim 0$, there is a large fraction of early-type galaxies that show evidence of ongoing star formation at intermediate redshift (e.g., Treu et al. 2005a; Kaviraj et al. 2011). Spatially resolved studies of such second burst galaxies will shed light on how these different scenarios contribute to the evolutionary path of massive galaxies at high redshift to the local counterpart.

6. Summary

We reconstructed SFHs of 24 massive, passively evolving galaxies at $z \sim 2$. Our new SED modeling with `gsf` simultaneously fit slitless spectroscopic and photometric data taken from multiple surveys, with no functional assumption for SFHs. Our main findings are as follows.

1. Our massive galaxies have already formed >50% of their current mass by $\sim$1.5 Gyr prior to the epoch of observation, with a downsizing trend where more massive galaxies evolve earlier.
2. The SFHs reconstructed by `gsf` show a more extended feature than what is obtained with a τ-model fitting for most of the sample galaxies, indicating low-level star formation activity until recently, rather than abrupt cessation.
3. The stellar-phase metallicities of most of our galaxies are already compatible with local values, indicating a rapid metallicity enrichment associated with the early stellar-mass formation.
4. By using the reconstructed SFHs and inferred metallicity, we revealed a rapid metallicity enrichment of this class of massive galaxies at a rate of $\sim$0.2 dex Gyr^{-1} in $\log Z_*/Z_\odot$ from $z \sim 5.5$ to 2.2.
5. While systematic uncertainties remain, the observed gap between the stellar- and gas-phase metallicities can be explained by continuation of a low level of star formation in quiescent galaxies and/or dilution of gas-phase metallicity due to the inflow of pristine gas to star-forming galaxies. The former scenario is consistent with the finding from individual SFHs.

We thank the anonymous referee for reading the manuscript carefully and providing constructive comments. We thank Marco Chiaberge, Colin Norman, Kartheik Iyer, Sara Ellison, and Susan Kassin for fruitful discussion. We thank Benedikt Diemer for providing star formation histories extracted from the Illustris simulation. We thank Pascal Oesch for providing HDUV data prior to the public release. Support for GLASS (HST-GO-13459) was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy,

Inc., under NASA contract NAS 5-26555. Support for this work is provided by NASA through a Spitzer award issued by JPL/Caltech, HST-AR-13235 and HST-GO-13177. M.T. acknowledges the support provided by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) through project No. CE170100013.

Software: Astropy (Muna et al. 2016), *emcee* (Foreman-Mackey et al. 2013), lmfit (Newville et al. 2017), Grizli (Brammer 2018), python-fsps (Conroy et al. 2009; Conroy & Gunn 2010; Foreman-Mackey et al. 2014).

Appendix A
Mock Simulation of SED Fitting

We test the fidelity of galaxy SFHs and other parameters with our SED fitting method.

A1. Simulation Setup

In this study, to explore the parameter spaces we are interested in (i.e., quenched galaxies at $z \sim 2$), we set parameters as follows: redshift $z \in$ [1.6: 2.5]; mass-weighted age T_*/Gyr$\in$ [0.6: 2.2] peaked at $\sim$1.5 Gyr; dust attenuation A_V/mag $\in$ [0: 2.5] with a flat distribution; and metallicity $\log Z_*/Z_\odot \in [-0.5: 0.5]$ peaked at $\log Z_*/Z_\odot = 0.15$ dex (top panels of Figure 7). The amplitudes of each template (a_i; i.e., SFHs) are randomly assigned. The input SFHs are shown in Figure 8.

The mock SEDs are generated via FSPS (Conroy et al. 2009; Conroy & Gunn 2010) with the assigned parameters. While we provide SFHs at the same time resolution as the fitting templates, this turns out only a small effect, thanks to a sufficient number of age bins (see also Appendix C for the result with higher-resolution SFHs).

Broadband photometry is then extracted by convolving the mock SEDs with filter response curves. For grism spectra, we convolve the mock template with the observed line spread

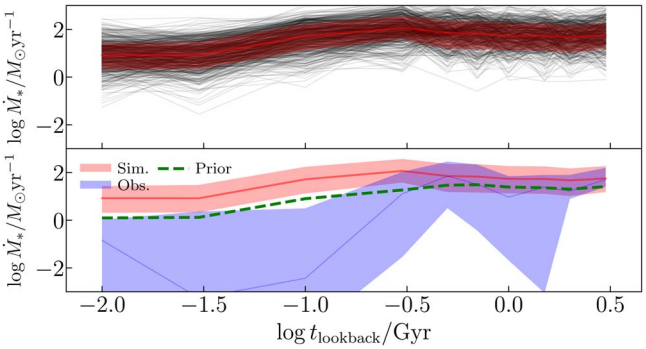

Figure 8. (Top) Input SFHs (thin lines) used for a simulation in Appendix A, with 16th/50th/84th percentiles (red shaded region). (Bottom) Median input SFH compared with the median of the observed SFH (blue). Since the simulated SFHs are randomly generated, it does not follow the observed decline at<0.3 Gyr. The flat prior used in `gsf` is shown (dashed line). Its effect should be minimal due to the wide constraint range set in this study ($\pm$3 dex), as seen in the posterior SFHs deviating from the prior.

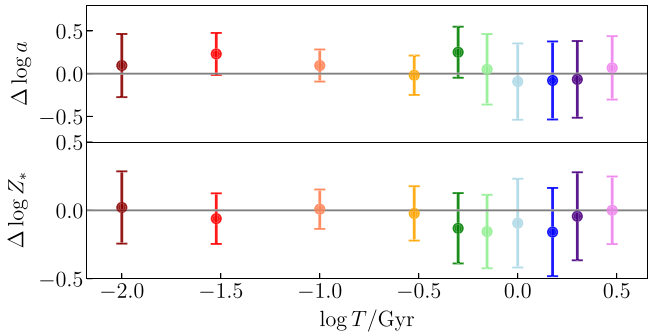

Figure 9. Mean offsets for amplitude ($\log a_{i,\mathrm{out}} - \log a_{i,\mathrm{in}}$; top panel) and metallicity ($\log Z_{i,\mathrm{out}} - \log Z_{i,\mathrm{in}}$; bottom panel) with standard deviation at each age pixel from the mock test. The measured standard deviation is added in quadrature to the observed uncertainty (i.e., star formation and metallicity histories).

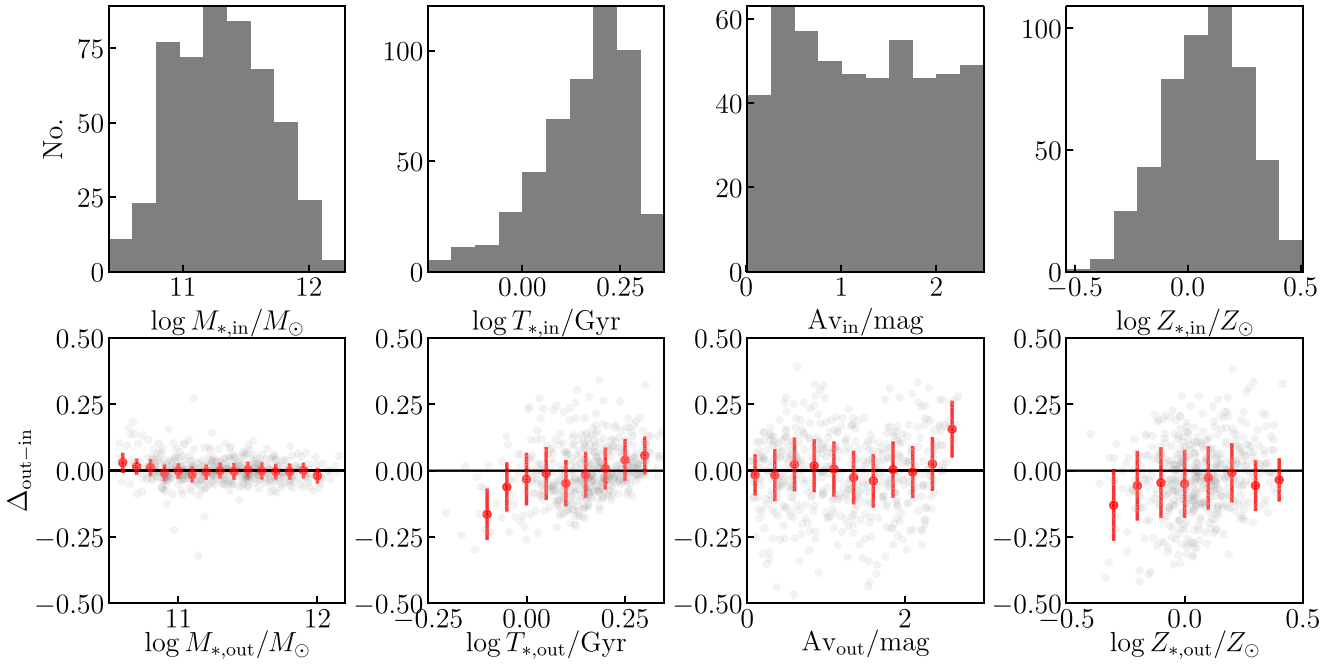

Figure 7. Top: parameter distributions of 600 mock galaxies. Note that age is distributed in a linear scale (peaked at $\sim$1.5 Gyr) but shown in a log scale here. Bottom: offsets of input and output values ($\Delta_{\mathrm{out-in}}$; solid lines represent the zero-point). Median offsets and 16th/84th percentile ranges at each output bin are shown (red circles and error bars). Stellar mass, dust attenuation, and mass-weighted metallicity show good agreement for the parameter ranges of our galaxies. While a weak positive correlation is seen in mass-weighted age, the measurement is not biased, and scatters are small ($\sigma \sim 0.1$) at the median value of our sample ($\log T_*$/ Gyr ~ 0.2).

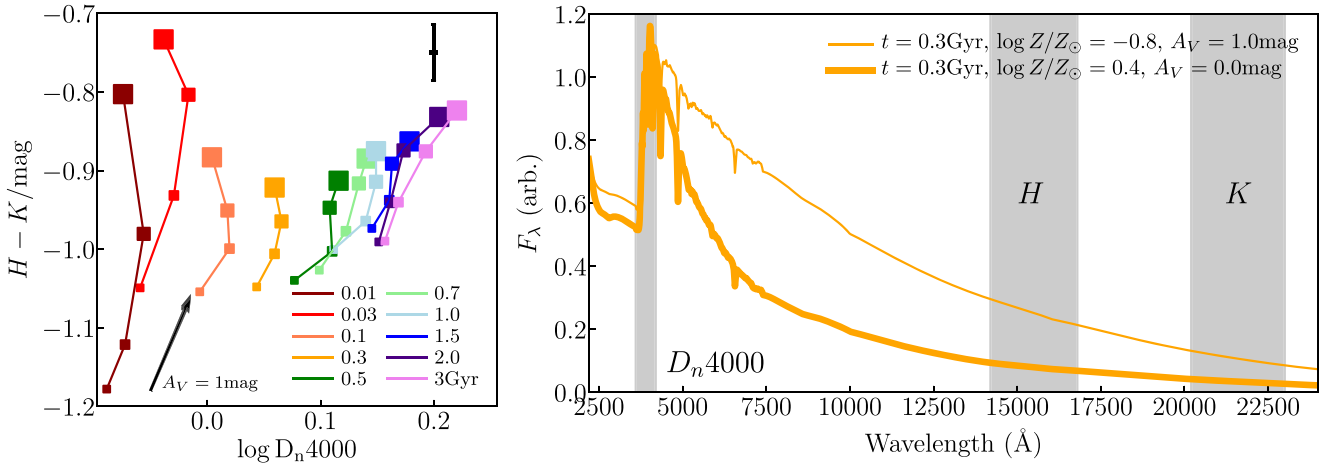

Figure 10. (Left) Rest-frame NIR color ($H - K$) as a function of D_n 4000 for each age (lines with different colors) and metallicity ($\log Z_*/Z_\odot = -0.8, -0.4, 0, 0.4$ for small to large symbols). Both parameters are derived from templates that are convolved to a comparable resolution of retrieved grism data ($R \sim 150$), though this is a minimal effect here. Age and metallicity are nearly orthogonal in most of the parameter space, except for the old and solar/supersolar metallicity population at the top right. Typical uncertainties in $H - K$ color and D_n 4000 are shown at the top right. (Right) Example of two spectral templates with similar color and D_n 4000 as in the left panel. Despite the similarity, these two templates are distinguishable with photometric data at the optical-to-NIR wavelength range.

function (modeled with a Gaussian), which takes into account the morphology/instrumental convolution. The error of each spectral element and broadband photometry is randomly assigned based on the observed uncertainty of 24 galaxies in this study (∼5–20 pixel^{-1} at $3700 < \lambda_{\rm rest}/\text{Å} < 4200$; Table 1).

No emission lines are added, since our focus is the quenched/old galaxy population. In total, 600 mock sets of G102/G141 spectra+broadband photometry are prepared from the template generated with random sets of parameters. We follow the same fitting method as in the main text.

A2. Result of Global Parameters

Figure 7 shows the offset of output and input values ($\Delta y = y_{\rm out} - y_{\rm in}$) as a function of output value for the major parameters: stellar mass, mass-weighted age, dust attenuation, and metallicity of the mock galaxies. By taking output values (rather than input ones) in the x-axis, it is possible to infer the false-positive fraction at a given output (i.e., observed) value and also implement the scatter to the observed values for a more comprehensive estimate of the uncertainty.

We find excellent agreement in stellar mass with a scatter of ∼0.05 dex and moderate agreement in mass-weighted age, dust attenuation, and mass-weighted metallicity with scatters of ∼0.11, 0.14, and 0.13 around the median values. Median offsets are small for most of the parameter ranges. Mass-weighted age shows a negative slope, underestimating for ∼0.15 dex at $\log T \lesssim -0.05$. However, most of our sample galaxies dominate higher values, with a median of $\log T_*/\,\text{Gyr} \sim 0.2$ (Figure 4), where the bias in parameters is small; thus, we do not correct the offset for our galaxies in the main text.

A3. Result of SFHs

In Figure 9, we summarize the offsets of the output and input values for amplitude and metallicity at each age pixel to show the fidelity of SFHs and metallicity enrichment histories from all mock galaxies used here. While the offset and scatter may depend on parameter sets with different combinations, this suggests that SFHs can be determined with ∼0.5 dex accuracy.

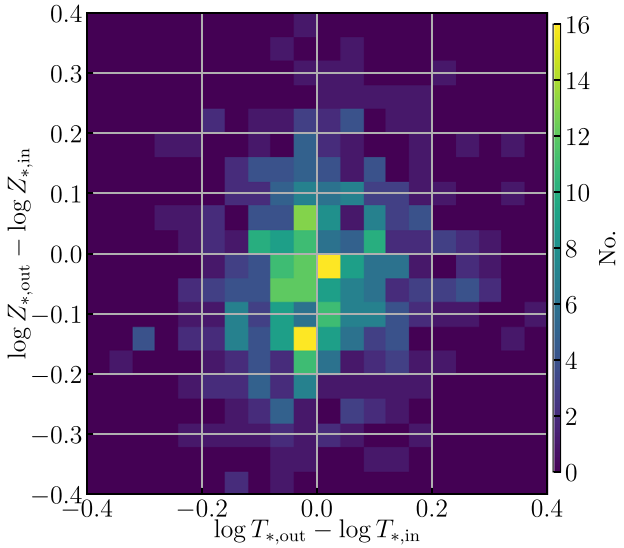

Figure 11. Distribution of offset in age ($\log T_{*,\rm out} - \log T_{*,\rm in}$; x-axis) and metallicity ($\log Z_{*,\rm out} - \log Z_{*,\rm in}$; y-axis) from our mock test. Colors represent the number of galaxies in each grid. The distribution is almost symmetric around the zero-point, whereas it would follow a negative slope in case of the age–metallicity degeneracy.

Metallicity shows a large scatter in reproduced values, with a standard deviation of ∼0.3 dex. Given the parameter range assigned for metallicity ($\in[-0.8: 0.6]$), we conclude that determination of metallicity at each age pixel is challenging with the current data set. It is noted that this does not mean that reproduced total metallicity has comparable uncertainty, since part of the scatter can be attributed to the age pixel where the total contribution of light is small. Total metallicity, which is light- or mass-weighted, should remain less scattered ($\lesssim$0.2 dex), as shown in the main text and Section A.2.

A4. Age–Metallicity Degeneracy

Degeneracy between age and metallicity is notoriously known as one of the difficult aspects when modeling accurate SEDs from photometric data (Worthey 1994). The degeneracy is, however, resolved once one obtains information at both the

optical and NIR wavelength ranges simultaneously (Figure 2; also de Jong 1996; Smail et al. 2001; Choi et al. 2016).

To see if this is the case for our data set, we first show in the left panel of Figure 10 a rest-frame NIR color– D_n 4000 diagram, both of which are available with our data in this study. Rest-frame NIR color ($H-K$) and the strength of 4000 Å(D_n 4000 Balogh et al. 1999) are calculated with templates used for fitting, which are convolved to a comparable resolution of grism data ($R\sim 150$), including the convolution effect by source morphology. As we see in the figure, age and metallicity are nearly orthogonal in most of the parameter range, meaning the age and metallicity can be well separated from those measurements. The only exception is for old ($\gtrsim$1.5 Gyr) and solar/supersolar metallicity, where the relation of the two measurements becomes less orthogonal. Due to this, the metallicities of our fitting typically have larger uncertainties for old populations.

One may notice that some parameters are not distinguished, especially by the reddening effect of dust (e.g., $t=0.3$ Gyr with $\log Z_*/Z_\odot=0.4$, $A_V=0$ versus $t=0.3$ Gyr with $\log Z_*/Z_\odot=-0.8$, $A_V=1.0$ mag). However, we stress that our SED fitting is not relying on any specific colors or indicators but rather on all spectrophotometric information over the wide wavelength range. As an example, we show two spectral templates in the right panel of Figure 10 that are located at a similar position in the D_n 4000–color space. Despite this, the two templates are clearly distinguishable at the rest-frame optical-to-NIR wavelength range, where sufficient photometric data points are available in this study. Also, the photometric error of each flux measurement may affect the SED parameters, but the uncertainty is properly implemented in our fitting framework using MCMC and reflected in the uncertainty range of the posterior.

We also investigate the age–metallicity degeneracy with our mock data set. In Figure 11, we show the distribution of the offset in total mass-weighted age and metallicity for our mock galaxies. The distribution is symmetric in both axes, whereas the distribution will follow a negative slope if these parameters are degenerated. The distribution is scattered for $\sim$0.2 dex, which is consistent with those found in Section A2.

From both tests here, we conclude that the data set used in this study can resolve the age–metallicity degeneracy for our moderately old galaxies ($\lesssim$2 Gyr), but star formation and metallicity histories become less certain beyond $T_{\rm lookback}\gtrsim 2$ Gyr.

Appendix B
Comparison of SED Parameters Obtained with Functional SFHs

In Section 4.3, we see that our reconstructed SFHs capture the detail features of individual galaxy SFHs that are often missed with functional forms. While the deviation is clear in these comparisons, it is yet to be investigated how a different assumption of SFH results in SED parameters. Here we compare the best-fit parameters between two types of SFHs.

In Figure 12, we compare the goodness of fit, χ^2/ν, and major parameters from SED fitting, i.e., stellar mass, light-weighted age and metallicity, and dust attenuation. These values are compared between those reproduced by `gsf` (main text) and functional ones. First, the goodness of fit is better with `gsf` for most of our galaxies, which is reasonable given the flexibility of its modeling. We also find that the τ model ($\propto\exp[-t/\tau]$) is more sensitive to the light from young stellar populations, where the model systematically underestimates the system's ages for $\sim$0.3 Gyr, on average (top panels). The discrepancy propagates to other parameters, where we find overestimated dust attenuation ($\sim$0.3 mag) and largely underestimated metallicity.

The discrepancy in age and dust becomes slightly smaller when the delayed-τ model ($\propto t\exp[-t/\tau]$) is used (bottom panels of Figure 12). This is shown in the goodness of fit, where the delayed-τ model results in smaller values of χ^2/ν (see also Pacifici et al. 2016). This is partly attributed to its rising slope in SFHs, which makes the age slightly older and cancels out the discrepancy in other parameters. A large discrepancy in metallicity, however, still remains.

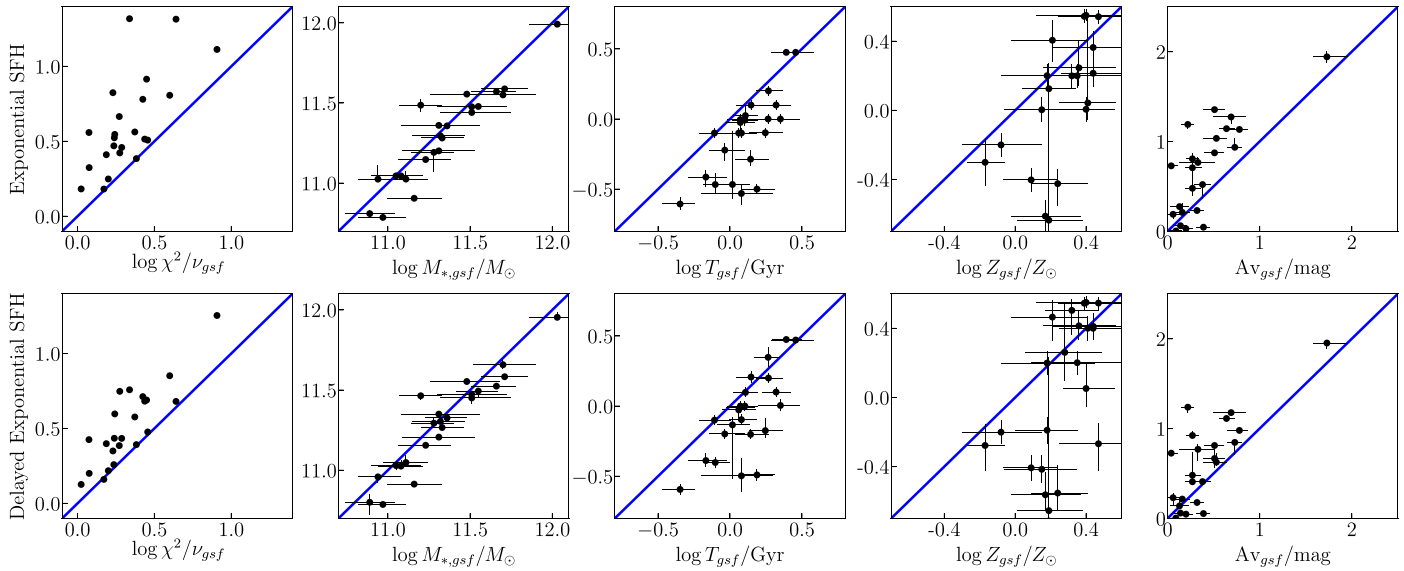

Figure 12. Comparison of χ^2/ν and SED parameters for 24 galaxies in this study derived from different SFHs, with `gsf` (x-axis) and functional forms for SFHs (y-axis). The exponential (top) and delayed (bottom) models are examined. The goodness of fit is better with `gsf` for most of our sample. While both SFHs reproduce the stellar mass in good agreement, the functional SFHs overestimate dust attenuation and underestimate age and metallicity. It is noted that the age and metallicity of `gsf` are light-weighted values to match those derived with functional form SFHs, while those in the main text are mass-weighted values.

Table 2
Summary of Physical Parameters with Different SFHs

Obj. ID	gsf[a]			Exponential Model					Delayed Exponential Model				
	χ^2/ν	$\log Z_L$[b] ($Z_\odot$)	$\log T_L$[b] (Gyr)	χ^2/ν	$\log M_*$ ($M_\odot$)	$\log Z_L$ ($Z_\odot$)	$\log T_L$ (Gyr)	A_V (mag)	χ^2/ν	$\log M_*$ ($M_\odot$)	$\log Z_L$ ($Z_\odot$)	$\log T_L$ (Gyr)	A_V (mag)
00141	2.41	$0.45^{+0.17}_{-0.19}$	$-0.34^{+0.09}_{-0.13}$	2.43	$11.49^{+0.03}_{-0.04}$	$0.54^{+0.04}_{-0.04}$	$-0.61^{+0.04}_{-0.04}$	$1.94^{+0.06}_{-0.07}$	2.47	$11.46^{+0.01}_{-0.02}$	$0.55^{+0.04}_{-0.04}$	$-0.59^{+0.04}_{-0.04}$	$1.95^{+0.03}_{-0.06}$
00227	1.75	$-0.17^{+0.12}_{-0.10}$	$-0.18^{+0.17}_{-0.14}$	3.52	$11.02^{+0.09}_{-0.01}$	$-0.30^{+0.14}_{-0.14}$	$-0.41^{+0.05}_{-0.07}$	$0.93^{+0.10}_{-0.01}$	3.95	$10.96^{+0.01}_{-0.03}$	$-0.28^{+0.12}_{-0.15}$	$-0.39^{+0.05}_{-0.04}$	$0.85^{+0.01}_{-0.14}$
06215	1.59	$0.41^{+0.14}_{-0.15}$	$-0.10^{+0.10}_{-0.09}$	1.77	$11.48^{+0.00}_{-0.01}$	$0.01^{+0.07}_{-0.07}$	$-0.47^{+0.08}_{-0.06}$	$1.14^{+0.01}_{-0.04}$	1.65	$11.49^{+0.00}_{-0.04}$	$0.05^{+0.15}_{-0.10}$	$-0.40^{+0.04}_{-0.04}$	$1.11^{+0.01}_{-0.04}$
07276	1.54	$0.13^{+0.21}_{-0.18}$	$0.15^{+0.11}_{-0.10}$	2.58	$11.57^{+0.00}_{-0.00}$	$0.00^{+0.07}_{-0.07}$	$0.10^{+0.03}_{-0.03}$	$0.23^{+0.01}_{-0.01}$	2.50	$11.52^{+0.00}_{-0.00}$	$-0.42^{+0.08}_{-0.08}$	$0.21^{+0.04}_{-0.04}$	$0.18^{+0.01}_{-0.01}$
09704	1.74	$0.40^{+0.17}_{-0.16}$	$0.15^{+0.13}_{-0.12}$	3.34	$11.59^{+0.00}_{-0.02}$	$0.54^{+0.04}_{-0.04}$	$-0.29^{+0.06}_{-0.04}$	$1.27^{+0.01}_{-0.08}$	2.72	$11.58^{+0.00}_{-0.00}$	$0.55^{+0.04}_{-0.04}$	$-0.20^{+0.04}_{-0.04}$	$1.18^{+0.01}_{-0.01}$
11470	1.88	$0.33^{+0.14}_{-0.14}$	$-0.11^{+0.11}_{-0.11}$	2.65	$11.05^{+0.00}_{-0.00}$	$0.20^{+0.07}_{-0.07}$	$-0.10^{+0.03}_{-0.03}$	$0.27^{+0.00}_{-0.00}$	5.59	$11.03^{+0.03}_{-0.00}$	$0.50^{+0.06}_{-0.14}$	$-0.10^{+0.03}_{-0.03}$	$0.14^{+0.09}_{-0.01}$
17599	1.49	$0.27^{+0.21}_{-0.22}$	$0.14^{+0.10}_{-0.09}$	3.69	$11.13^{+0.03}_{-0.08}$	$0.03^{+0.47}_{-0.10}$	$-0.03^{+0.13}_{-0.10}$	$0.52^{+0.11}_{-0.29}$	2.74	$11.11^{+0.10}_{-0.04}$	$0.52^{+0.06}_{-0.53}$	$-0.07^{+0.08}_{-0.05}$	$0.50^{+0.15}_{-0.17}$
17735	1.70	$0.40^{+0.15}_{-0.14}$	$-0.02^{+0.13}_{-0.12}$	6.68	$11.28^{+0.01}_{-0.00}$	$0.04^{+0.50}_{-0.10}$	$-0.22^{+0.05}_{-0.07}$	$0.87^{+0.00}_{-0.04}$	2.24	$11.27^{+0.00}_{-0.00}$	$0.40^{+0.07}_{-0.07}$	$-0.20^{+0.04}_{-0.04}$	$0.81^{+0.00}_{-0.00}$
19341	4.37	$0.08^{+0.17}_{-0.16}$	$0.19^{+0.13}_{-0.10}$	20.66	$10.79^{+0.00}_{-0.01}$	$-0.40^{+0.07}_{-0.07}$	$-0.50^{+0.04}_{-0.03}$	$0.73^{+0.01}_{-0.01}$	4.79	$10.79^{+0.00}_{-0.01}$	$-0.41^{+0.07}_{-0.07}$	$-0.49^{+0.04}_{-0.04}$	$0.73^{+0.01}_{-0.01}$
19850	1.87	$0.27^{+0.22}_{-0.20}$	$0.06^{+0.13}_{-0.12}$	4.63	$11.05^{+0.00}_{-0.00}$	$-0.75^{+0.03}_{-0.03}$	$-0.10^{+0.03}_{-0.03}$	$1.03^{+0.01}_{-0.01}$	2.43	$11.03^{+0.03}_{-0.03}$	$0.26^{+0.28}_{-0.17}$	$-0.03^{+0.05}_{-0.08}$	$0.63^{+0.10}_{-0.04}$
22774	1.19	$0.17^{+0.18}_{-0.19}$	$0.09^{+0.10}_{-0.09}$	2.11	$11.36^{+0.00}_{-0.00}$	$-0.75^{+0.03}_{-0.03}$	$-0.10^{+0.03}_{-0.03}$	$1.13^{+0.01}_{-0.01}$	1.59	$11.33^{+0.00}_{-0.00}$	$-0.19^{+0.08}_{-0.08}$	$-0.10^{+0.04}_{-0.04}$	$0.98^{+0.03}_{-0.03}$
23006	1.06	$0.18^{+0.17}_{-0.17}$	$0.08^{+0.10}_{-0.09}$	1.52	$11.03^{+0.05}_{-0.01}$	$-0.64^{+0.09}_{-0.10}$	$-0.00^{+0.04}_{-0.04}$	$0.19^{+0.02}_{-0.05}$	1.34	$11.05^{+0.05}_{-0.02}$	$-0.66^{+0.11}_{-0.10}$	$0.00^{+0.03}_{-0.04}$	$0.23^{+0.06}_{-0.08}$
23073	1.19	$0.35^{+0.21}_{-0.20}$	$0.07^{+0.10}_{-0.10}$	3.62	$11.29^{+0.03}_{-0.00}$	$0.25^{+0.16}_{-0.10}$	$-0.02^{+0.05}_{-0.08}$	$0.21^{+0.10}_{-0.01}$	2.66	$11.30^{+0.00}_{-0.00}$	$0.41^{+0.08}_{-0.08}$	$-0.01^{+0.04}_{-0.04}$	$0.22^{+0.01}_{-0.01}$
23249	1.48	$0.20^{+0.20}_{-0.22}$	$0.08^{+0.22}_{-0.27}$	1.52	$10.81^{+0.02}_{-0.01}$	$0.41^{+0.12}_{-0.12}$	$-0.53^{+0.06}_{-0.08}$	$0.77^{+0.06}_{-0.02}$	1.45	$10.80^{+0.05}_{-0.08}$	$0.47^{+0.10}_{-0.14}$	$-0.50^{+0.12}_{-0.11}$	$0.77^{+0.06}_{-0.13}$
24033	2.18	$0.46^{+0.17}_{-0.20}$	$0.11^{+0.10}_{-0.09}$	20.81	$11.15^{+0.00}_{-0.00}$	$0.21^{+0.08}_{-0.08}$	$-0.01^{+0.04}_{-0.04}$	$0.48^{+0.00}_{-0.01}$	5.73	$11.16^{+0.00}_{-0.00}$	$0.40^{+0.07}_{-0.07}$	$0.00^{+0.03}_{-0.03}$	$0.48^{+0.00}_{-0.00}$
24569	2.66	$-0.08^{+0.24}_{-0.22}$	$0.32^{+0.09}_{-0.09}$	6.03	$10.91^{+0.01}_{-0.00}$	$-0.20^{+0.07}_{-0.07}$	$0.10^{+0.03}_{-0.03}$	$0.03^{+0.03}_{-0.01}$	5.16	$10.91^{+0.01}_{-0.01}$	$-0.20^{+0.07}_{-0.07}$	$0.10^{+0.03}_{-0.03}$	$0.05^{+0.03}_{-0.02}$
31397	2.81	$0.43^{+0.23}_{-0.22}$	$0.10^{+0.09}_{-0.09}$	8.22	$11.48^{+0.00}_{-0.04}$	$0.36^{+0.09}_{-0.14}$	$0.02^{+0.08}_{-0.05}$	$0.52^{+0.01}_{-0.16}$	4.89	$11.48^{+0.00}_{-0.00}$	$0.41^{+0.07}_{-0.08}$	$0.10^{+0.04}_{-0.04}$	$0.41^{+0.01}_{-0.01}$
33780	1.72	$0.48^{+0.19}_{-0.21}$	$0.27^{+0.08}_{-0.09}$	2.95	$11.55^{+0.00}_{-0.00}$	$-0.75^{+0.03}_{-0.03}$	$0.00^{+0.03}_{-0.03}$	$1.35^{+0.01}_{-0.01}$	1.81	$11.66^{+0.03}_{-0.03}$	$-0.27^{+0.12}_{-0.16}$	$0.35^{+0.07}_{-0.07}$	$0.67^{+0.13}_{-0.10}$
39364	3.98	$0.17^{+0.19}_{-0.19}$	$0.36^{+0.13}_{-0.16}$	6.42	$11.44^{+0.00}_{-0.00}$	$-0.61^{+0.09}_{-0.09}$	$0.00^{+0.03}_{-0.03}$	$1.19^{+0.00}_{-0.05}$	7.11	$11.45^{+0.00}_{-0.04}$	$-0.56^{+0.13}_{-0.09}$	$0.00^{+0.04}_{-0.04}$	$1.24^{+0.00}_{-0.06}$
41021	2.86	$0.20^{+0.15}_{-0.15}$	$0.03^{+0.11}_{-0.13}$	3.22	$11.19^{+0.01}_{-0.12}$	$0.13^{+0.15}_{-0.86}$	$-0.47^{+0.38}_{-0.10}$	$0.71^{+0.05}_{-0.31}$	3.00	$11.29^{+0.03}_{-0.00}$	$-0.72^{+0.95}_{-0.06}$	$-0.13^{+0.06}_{-0.38}$	$0.41^{+0.32}_{-0.01}$
41148	2.36	$0.36^{+0.26}_{-0.25}$	$0.39^{+0.10}_{-0.10}$	3.65	$11.56^{+0.00}_{-0.00}$	$0.20^{+0.07}_{-0.07}$	$0.48^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	3.77	$11.55^{+0.00}_{-0.00}$	$0.20^{+0.07}_{-0.07}$	$0.47^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
41520	1.94	$0.19^{+0.24}_{-0.27}$	$0.45^{+0.12}_{-0.18}$	2.88	$11.36^{+0.00}_{-0.00}$	$0.20^{+0.07}_{-0.07}$	$0.48^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	2.71	$11.35^{+0.01}_{-0.00}$	$0.20^{+0.07}_{-0.07}$	$0.47^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$
43005	2.73	$0.38^{+0.29}_{-0.26}$	$0.28^{+0.10}_{-0.09}$	3.27	$11.20^{+0.00}_{-0.00}$	$0.55^{+0.03}_{-0.03}$	$0.20^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$	4.81	$11.21^{+0.00}_{-0.00}$	$0.55^{+0.03}_{-0.03}$	$0.20^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$
43114	8.08	$0.24^{+0.16}_{-0.16}$	$0.24^{+0.12}_{-0.12}$	12.98	$11.99^{+0.02}_{-0.00}$	$-0.42^{+0.08}_{-0.13}$	$-0.10^{+0.03}_{-0.04}$	$0.81^{+0.06}_{-0.00}$	17.93	$11.95^{+0.03}_{-0.01}$	$-0.55^{+0.16}_{-0.10}$	$-0.17^{+0.08}_{-0.05}$	$0.92^{+0.04}_{-0.01}$

Notes.
[a] Stellar mass and dust attenuation for gsf are listed in Table 1.
[b] Light-weighted metallicity and age, to match those derived with functional SFHs.

Despite this, it is interesting that the reproduced stellar masses are very consistent with each other, with only ∼0.1 dex scatter in $\log M_*$ among our sample. While it is challenging to comprehensively understand this agreement due to a large number of parameters here, this is partly due to the sufficient wavelength coverage of the rest-frame near-IR region, where the light from low-mass stars dominates and is less sensitive to other parameters (e.g., Bell et al. 2003). The contribution by other parameters (i.e., age/metallicity/dust) is canceled out due to partial degeneracy at a given form of SFH. However, this does not necessarily mean that the typical error in stellar mass remains comparably small. As shown in the main text and Appendix A, the stellar-mass measurement with `gsf` is accompanied by an ∼0.2 dex uncertainty that mainly originates from, e.g., systematics in estimating accurate SFHs and metallicity enrichment histories. (Therefore, the estimated uncertainty in stellar mass with functional form SFHs is much smaller here.) For this reason, we conclude that the stellar-mass measurement remains at least at ∼0.15 dex accuracy for our galaxies, and possibly for other types of galaxies, since the uncertainty comes from its assumption of SFHs and SED modeling. The best-fit parameters derived with the two functional SFHs are summarized in Table 2.

Appendix C
Simulation with Realistic SFHs

While our test with randomly generated SFHs provides a general idea of the goodness of `gsf`, there is still concern for how a specific type of SFH affects the output results. In particular, the random SFHs do not fully investigate the SFHs of quenched galaxies, i.e., the target galaxies of our study. Upon such a demand, here we repeat a similar fitting analysis as in Appendix A but with SFHs taken from a cosmological simulation, which gives us an idea of how well quenched galaxies are reconstructed within our framework.

In Figure 13, we compare the input and output histories for 10 galaxies. The set of galaxies is selected from the Illustris simulation (Nelson et al. 2015), with a similar mass to our galaxies ($\log M_*/M_\odot \sim 11$) and quenched at the time of observation (SFR $< 1\ M_\odot\ \mathrm{yr}^{-1}$ at $z \sim 2$). The SFHs and metallicity enrichment histories are provided to the FSPS to synthesize the SEDs. From the generated SEDs, we extract fluxes corresponding to our grism elements (convolved with the morphology of one of our sources) and broadband filters. We then add noise with a conservative value of $\langle \mathrm{S/N} \rangle = 10$ at 4200–5000 Å (and 15 as a supplemental test). We set the observed redshift uniformly to $z = 2$ and $A_V = 0.5$ mag for the sake of simplicity, but this hardly affects the conclusion here.

In general, the posterior captures the feature of SFHs—the peak time of SFR and its length—and gives fairly good estimates of SED parameters: stellar mass, mass-weighted age and metallicity, and dust attenuation (bottom panel of Figure 13). There is a trend that `gsf` underestimates star formation in the oldest bin. This leads to underestimation in the mass-weighted age for old galaxies with $\log T_{*,\mathrm{input}}/\mathrm{Gyr} > 0.2$, but this is only for a small amount (<0.1 dex). The offset seen in output metallicity is also small (∼0.1 dex) and hardly changes our conclusion in the main text, as the measurements in the main text quote much larger uncertainties from the analysis in Appendix A.

One caveat is that dust attenuation is overestimated for ∼0.2 mag (and stellar mass is overestimated for ∼0.1 dex accordingly). However, this should be considered as a result of this specific type of SFH (and SED), as we see fairly good reproduction of the parameter for random SFHs in Figure 7. The offset becomes smaller by increasing the input S/N to 15 but is not completely dismissed.

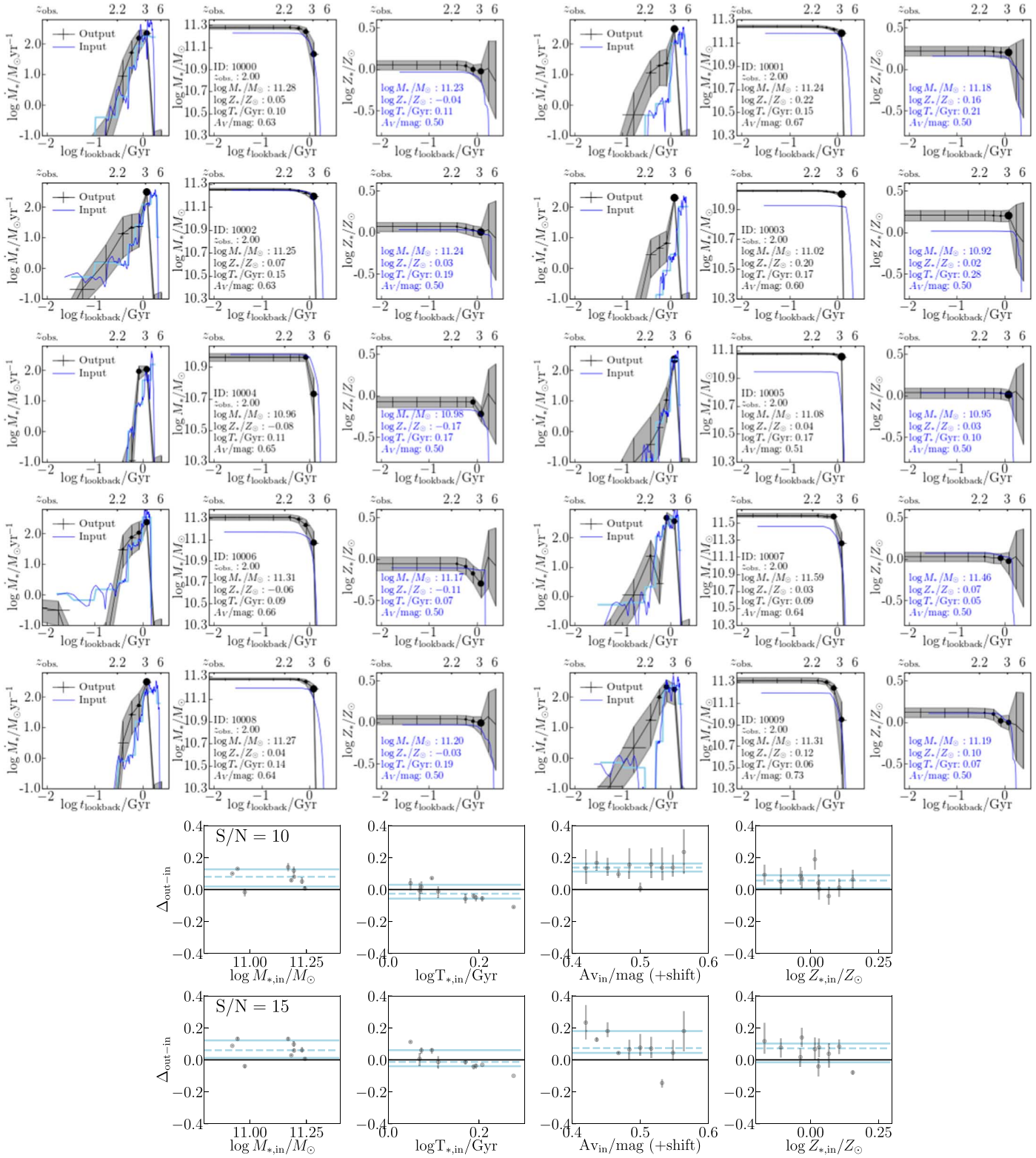

Figure 13. (Top) Results of the star formation reconstruction with `gsf` for 10 quenching galaxies. The input star formation and metallicity histories (dark blue lines) are selected from the Illustris simulation within our observed mass range ($\log M_*/M_\odot \sim 11$) and quenched (SFR $< 1\ M_\odot\ {\rm yr}^{-1}$) at $z \sim 2$. The SFHs binned to the fitting template resolution are also shown (light blue) for the comparison with output SFHs. The input and output values for each parameter are shown in the panels (blue and black text, respectively). (Bottom) Summary results of output parameters ($\langle$ S/N$\rangle$ = 10 and 15). There is a negative trend in mass-weighted age. This is caused by underestimation of star formation in the oldest bin, but the net effect is small ($\Delta T \sim 0.1$ Gyr). The dust attenuation shows an offset ($\Delta \sim 0.2$ mag), but this is rather dependent on a specific type of SFH (see Figure 7).

ORCID iDs

T. Morishita https://orcid.org/0000-0002-8512-1404
L. E. Abramson https://orcid.org/0000-0002-8860-1032
T. Treu https://orcid.org/0000-0002-8460-0390
G. B. Brammer https://orcid.org/0000-0003-2680-005X
T. Jones https://orcid.org/0000-0001-5860-3419
P. Kelly https://orcid.org/0000-0003-3142-997X
M. Trenti https://orcid.org/0000-0001-9391-305X
B. Vulcani https://orcid.org/0000-0003-0980-1499
X. Wang https://orcid.org/0000-0002-9373-3865

References

Abramson, L. E., Newman, A. B., Treu, T., et al. 2018, AJ, 156, 29
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54
Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289
Belli, S., Newman, A. B., & Ellis, R. S. 2014, ApJ, 783, 117
Belli, S., Newman, A. B., & Ellis, R. S. 2017, ApJ, 834, 18
Belli, S., Newman, A. B., & Ellis, R. S. 2019, ApJ, 874, 17
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bezanson, R., Spilker, J., Williams, C. C., et al. 2019, ApJL, 873, L19
Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175
Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349
Boquien, M., Buat, V., & Perret, V. 2014, A&A, 571, A72
Brammer, G. 2018, gbrammer/grizli: Preliminary release, Zenodo, doi:10.5281/zenodo.1146905
Brammer, G. B., Marchesini, D., Labbé, I., et al. 2016, ApJS, 226, 6
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2018, arXiv:1811.03635
Chauke, P., van der Wel, A., Pacifici, C., et al. 2018, ApJ, 861, 13
Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Cid Fernandes, R. 2018, arXiv:1807.10423
Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363
Cimatti, A., Daddi, E., Renzini, A., et al. 2004, Natur, 430, 184
Cole, S., Norberg, P., Baugh, C. M., et al. 2001, MNRAS, 326, 255
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, ApJ, 626, 680
de Jong, R. S. 1996, A&A, 313, 377
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Diemer, B., Sparre, M., Abramson, L. E., & Torrey, P. 2017, ApJ, 839, 26
Domínguez Sánchez, H., Pérez-González, P. G., Esquej, P., et al. 2016, MNRAS, 457, 3743
Dressler, A., & Gunn, J. E. 1983, ApJ, 270, 7
Dressler, A., Kelson, D. D., Abramson, L. E., et al. 2016, ApJ, 833, 251
Dressler, A., Kelson, D. D., & Abramson, L. E. 2018, arXiv:1805.04110
Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813
Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2019, ApJ, 870, 133
Feldmann, R., & Mayer, L. 2015, MNRAS, 446, 1939
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, Zenodo, doi:10.5281/zenodo.12157
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748
Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, ApJ, 788, 72
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, Natur, 544, 71
Gobat, R., Daddi, E., Magdis, G., et al. 2018, NatAs, 2, 239
Gobat, R., Daddi, E., Strazzullo, V., et al. 2017, A&A, 599, A95
González Delgado, R. M., Cid Fernandes, R., García-Benito, R., et al. 2014, ApJL, 791, L16
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Hamann, F., & Ferland, G. 1999, ARA&A, 37, 487
Harikane, Y., Ouchi, M., Ono, Y., et al. 2019, arXiv:1902.09555
Heavens, A., Panter, B., Jimenez, R., & Dunlop, J. 2004, Natur, 428, 625
Hopkins, P. F., Hernquist, L., Cox, T. J., Keres, D., & Wuyts, S. 2009, ApJ, 691, 1424
Iyer, K., & Gawiser, E. 2017, ApJ, 838, 127
Jimenez, R., Bernardi, M., Haiman, Z., Panter, B., & Heavens, A. F. 2007, ApJ, 669, 947
Juneau, S., Bournaud, F., Charlot, S., et al. 2014, ApJ, 788, 88
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Kaviraj, S., Tan, K.-M., Ellis, R. S., & Silk, J. 2011, MNRAS, 411, 2148
Kelly, P. L., Brammer, G., Selsing, J., et al. 2016, ApJ, 831, 205
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Sci, 347, 1123
Kelson, D. D., Williams, R. J., Dressler, A., et al. 2014, ApJ, 783, 110
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, Natur, 540, 248
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692
Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, ApJ, 856, 15
Lehner, N., O'Meara, J. M., Howk, J. C., Prochaska, J. X., & Fumagalli, M. 2016, ApJ, 833, 283
Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2018, arXiv:1811.03637
Lonoce, I., Longhetti, M., Maraston, C., et al. 2015, MNRAS, 454, 3912
Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, ApJ, 742, 103
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Man, A., & Belli, S. 2018, NatAs, 2, 695
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2015, ApJ, 801, 133
Martín-Navarro, I., Vazdekis, A., Falcón-Barroso, J., et al. 2018, MNRAS, 475, 3700
McDermid, R. M., Alatalo, K., Blitz, L., et al. 2015, MNRAS, 448, 3484
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Morishita, T., Abramson, L. E., Treu, T., et al. 2017, ApJ, 835, 254
Morishita, T., Abramson, L. E., Treu, T., et al. 2018, ApJL, 856, L4
Morishita, T., & Ichikawa, T. 2016, ApJ, 816, 87
Morishita, T., Ichikawa, T., & Kajisawa, M. 2014, ApJ, 785, 18
Morishita, T., Ichikawa, T., Noguchi, M., et al. 2015, ApJ, 805, 34
Muna, D., Alexander, M., Allen, A., et al. 2016, arXiv:1610.03159
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJL, 699, L178
Nelson, D., Pillepich, A., Genel, S., et al. 2015, A&C, 13, 12
Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
Newville, M., Nelson, A., Ingargiola, A., et al. 2017, Zenodo, doi:10.5281/zenodo.802298
Nipoti, C., Treu, T., Leauthaud, A., et al. 2012, MNRAS, 422, 1714
Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129
Oesch, P. A., Montes, M., Reddy, N., et al. 2018, arXiv:1806.01853
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42
Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, ApJ, 808, 161
Onodera, M., Renzini, A., Carollo, M., et al. 2012, ApJ, 755, 26
Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, ApJ, 725, 2312
Osterbrock, D. E. 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Mill Valley, CA: University Science Books)
Pacifici, C., Kassin, S. A., Weiner, B. J., et al. 2016, ApJ, 832, 79
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, MNRAS, 378, 1550
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2008, MNRAS, 391, 1117
Peng, Y., Maiolino, R., & Cochrane, R. 2015, Natur, 521, 192
Pirzkal, N., Malhotra, S., Ryan, R. E., et al. 2017, ApJ, 846, 84
Poggianti, B. M., Calvi, R., Bindoni, D., et al. 2013, ApJ, 762, 77
Postman, M., Coe, D., Benítez, N., et al. 2012, ApJS, 199, 25
Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, Natur, 496, 329
Salpeter, E. E. 1955, ApJ, 121, 161

Sánchez, S. F., Barrera-Ballesteros, J. K., Sánchez-Menguiano, L., et al. 2017, MNRAS, 469, 2121
Sargent, M. T., Daddi, E., Bournaud, F., et al. 2015, ApJL, 806, L20
Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, ApJL, 782, L36
Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018, arXiv:1807.02523
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Smail, I., Kuntschner, H., Kodama, T., et al. 2001, MNRAS, 323, 839
Snyder, G. F., Lotz, J. M., Rodriguez-Gomez, V., et al. 2017, MNRAS, 468, 207
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJL, 783, L14
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Taylor, E. N., Franx, M., Glazebrook, K., et al. 2010, ApJ, 720, 723
Teplitz, H. I., Rafelski, M., Kurczynski, P., et al. 2013, AJ, 146, 159
Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 339, 897
Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673
Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, MNRAS, 404, 1775
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68
Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, MNRAS, 381, 1252
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
Torrey, P., Cox, T. J., Kewley, L., & Hernquist, L. 2012, ApJ, 746, 108
Torrey, P., Wellons, S., Ma, C.-P., Hopkins, P. F., & Vogelsberger, M. 2017, MNRAS, 467, 4872
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 119, 1645
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Ellis, R. S., Liao, T. X., et al. 2005a, ApJ, 633, 174
Treu, T., Ellis, R. S., Liao, T. X., & van Dokkum, P. G. 2005b, ApJL, 622, L5
Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114
Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, A58
Trujillo, I., Conselice, C. J., Bundy, K., et al. 2007, MNRAS, 382, 109
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Dokkum, P., Brammer, G., Momcheva, I., et al. 2013a, arXiv:1305.2140
van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJL, 677, L5
van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013b, ApJL, 771, L35
van Dokkum, P. G., Nelson, E. J., Franx, M., et al. 2015, ApJ, 813, 23
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Walcher, C. J., Coelho, P., Gallazzi, A., & Charlot, S. 2009, MNRAS, 398, L44
Walcher, C. J., Coelho, P. R. T., Gallazzi, A., et al. 2015, A&A, 582, A46
Wang, X., Jones, T. A., Treu, T., et al. 2017, ApJ, 837, 89
Wang, X., Jones, T. A., Treu, T., et al. 2018, arXiv:1808.08800
Wellons, S., Torrey, P., Ma, C.-P., et al. 2015, MNRAS, 449, 361
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Worthey, G. 1994, ApJS, 95, 107
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114
Yabe, K., Ohta, K., Iwamuro, F., et al. 2014, MNRAS, 437, 3647
Yi, S., Demarque, P., & Oemler, A., Jr. 1997, ApJ, 486, 201
Younger, J. D., Fazio, G. G., Huang, J.-S., et al. 2007, ApJ, 671, 1531
Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, ApJ, 792, 75
Zitrin, A., Labbé, I., Belli, S., et al. 2015, ApJL, 810, L12