CrossMark

Evolution of the Stellar Mass–Metallicity Relation. I. Galaxies in the $z \sim 0.4$ Cluster Cl0024

Nicha Leethochawalit[1] , Evan N. Kirby[1] , Sean M. Moran[2], Richard S. Ellis[3] , and Tommaso Treu[4]

[1] Cahill Center for Astronomy and Astrophysics, California Institute of Technology, MS 249-17, Pasadena, CA 91125, USA
[2] Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA
[3] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[4] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

Abstract

We present the stellar mass–stellar metallicity relationship (MZR) in the galaxy cluster Cl0024+1654 at $z \sim 0.4$ using full-spectrum stellar population synthesis modeling of individual quiescent galaxies. The lower limit of our stellar mass range is $M_* = 10^{9.7} M_\odot$, the lowest galaxy mass at which individual stellar metallicity has been measured beyond the local universe. We report a detection of an evolution of the stellar MZR with observed redshift at 0.037 ± 0.007 dex per Gyr, consistent with the predictions from hydrodynamical simulations. Additionally, we find that the evolution of the stellar MZR with observed redshift can be explained by an evolution of the stellar MZR with the formation time of galaxies, i.e., when the single stellar population (SSP)-equivalent ages of galaxies are taken into account. This behavior is consistent with stars forming out of gas that also has an MZR with a normalization that decreases with redshift. Lastly, we find that over the observed mass range, the MZR can be described by a linear function with a shallow slope ([Fe/H] $\propto (0.16 \pm 0.03) \log M_*$). The slope suggests that galaxy feedback, in terms of mass-loading factor, might be mass-independent over the observed mass and redshift range.

Key words: galaxies: abundances – galaxies: evolution – galaxies: stellar content

Supporting material: machine-readable table

1. Introduction

One of the best-known properties in observed galaxies is the tight correlation between galaxy stellar mass and gas-phase metallicity, i.e., the mass–metallicity relation (MZR). Several large galaxy surveys, such as the Sloan Digital Sky Survey (SDSS), have confirmed that galaxies at all redshifts with higher stellar masses retain more metals than galaxies with lower stellar masses (e.g., Tremonti et al. 2004; Sanders et al. 2015; Guo et al. 2016; Onodera et al. 2016). While the details of the evolution of the gas-phase MZR with redshift are still debated, mainly due to the different metallicity indicators (e.g., Steidel et al. 2014; Kewley et al. 2015; Bian et al. 2017; Strom et al. 2017), galaxies at higher redshifts generally follow the same trend as the local MZR but are somewhat offset toward lower metallicities (e.g., Erb et al. 2006; Maiolino et al. 2008; Steidel et al. 2014; Zahid et al. 2014).

Despite the well-established gas-phase MZR, our understanding of the amount of metals that stars incorporate from the gas is less secure. The gas-phase metallicity only indicates the amount of metals in the gas during the time of observation. The metallicity in stars indicates the metal content in stars at their formation. Therefore, measuring stellar metallicity from a galaxy's integrated stellar light can reveal the "star formation history-averaged" galactic metallicity. It is less susceptible than the gas to instantaneous fluctuations. Obtaining stellar metallicities over a range of galaxy masses, i.e., the stellar MZR, can provide insight into the chemical evolution of galaxies complementary to the gas-phase MZR. For example, Peng et al. (2015) compared the stellar MZR of local star-forming galaxies to that of quenched galaxies to study galaxy quenching mechanisms.

A few suites of cosmological hydrodynamical simulations have made quantitative predictions of the stellar MZR in the past few years. Ma et al. (2016) investigated the evolution of both gas-phase and stellar MZRs with redshift from a limited number of galaxies in the Feedback in Realistic Environment (FIRE) cosmological zoom-in simulations. These simulations suggest that the stellar MZR evolves monotonically, with an increase in stellar metallicity of ~ 0.3 dex at fixed stellar mass from $z = 1$ to $z = 0$. De Rossi et al. (2017) presented stellar MZRs at four redshifts derived from a different suite of cosmological hydrodynamical simulations, the Evolution and Assembly of Galaxies and their Environments (EAGLE). The derived stellar MZRs came from a larger number of galaxies than those in Ma et al.'s (2016) study but with coarser spatial resolution. De Rossi et al. (2017) predicted that the evolution of the stellar MZR is 0.2 dex from $z = 1$ to $z = 0$ at a stellar mass of $10^{9.5} M_\odot$, slightly smaller than predicted by Ma et al. (2016).

The classical approach to measuring ages and metallicities of stellar populations is to use spectrophotometric indices such as Lick indices (Faber 1973; Worthey 1994), where the equivalent widths of some spectral features expected to correlate with metal abundance or age are measured and compared to model predictions. Gallazzi et al. (2005) conducted one of the pioneering works using spectrophotometric indices to measure the stellar MZR from local $z \sim 0$ SDSS star-forming and quiescent galaxies. Gallazzi et al. (2014) extended their work to 77 galaxies at $z \sim 0.7$. They found an offset of the stellar MZR by 0.12 dex from $z = 0.7$ to $z = 0$ among the star-forming population but no significant evolution among the quiescent population.

However, the stellar metallicities of star-forming galaxies should be treated with caution since they tend to be more difficult to measure. Large uncertainties and biases could arise

from emission line subtractions, a lack of young stars in stellar libraries, and the fact that the most luminous stars are not necessarily the majority of the mass. The uncertainties of metallicities in star-forming galaxies measured by Gallazzi et al. (2006) are generally twice the uncertainties of the metallicities in quiescent galaxies, with a median of $\delta \log Z \sim 0.16$ for star-forming galaxies and 0.08 for quiescent galaxies. Moreover, using a similar sample of SDSS galaxies, Panter et al. (2008) found that star-forming galaxies have higher stellar metallicity than the whole sample while Gallazzi et al. (2005) and Peng et al. (2015) found the opposite. Adding further confusion, when metallicities were measured from equivalent widths of UV absorption lines, Sommariva et al. (2012) found that the stellar MZR of star-forming galaxies at $z = 3$ is consistent with that of local galaxies measured by Gallazzi et al. (2005), i.e., no significant evolution of the stellar MZR among the star-forming galaxies from $z = 3$ to $z = 0$. This contradicts the conclusion of Gallazzi et al. (2014). Nonetheless, the metallicities measured by Sommariva et al. (2012)—using light mainly produced by O stars—might trace different populations from those measured by Gallazzi et al. (2005)—using the light from stars of later spectral type.

If we focus on the quiescent galaxies whose ages and metallicities can be measured more reliably, we find that a number of recent works have been employing an alternative approach—a full-spectrum fitting technique—to determine the ages and metallicities of these quiescent stellar populations. The modeling of full optical–near-IR spectra of stellar systems has been advanced in the past decade (e.g., Cid Fernandes et al. 2005; Ocvirk et al. 2006; Conroy et al. 2009, 2014; Walcher et al. 2009; Conroy & van Dokkum 2012). The method is preferred over the use of spectrophotometric indices because it utilizes nearly all of the information from the collected light, resulting in smaller uncertainties.

Though the approach has been used to measure ages and metallicities of both local and high-redshift galaxies, the measurements of high-redshift galaxies are still mainly limited to stacks of spectra because of the lack of sufficient signal in individual spectra. Choi et al. (2014) measured the stellar MZR from stacked spectra of quiescent galaxies ranging from $z = 0$ to $z = 0.7$. The stellar MZRs show possible evidence for evolution with redshift. The MZR measured from stacked spectra can only reflect the median metallicities of the population. It cannot reveal the scatter of age or metallicity within the population. In fact, Choi et al. (2014) measured very different ages and metallicities of two individual galaxies at $z = 0.8$. The ages and metallicities of the individual galaxies were significantly different from the results from the stacked spectrum at the same redshift. Although the two galaxies were brightest cluster galaxies, the results suggest a possible large scatter within the population. Moreover, measuring galaxy properties in individual galaxies can reveal potentially important correlations between galaxy parameters such as Δ[Fe/H] and age or Δ[Fe/H] and [α/Fe], where Δ[Fe/H] is the deviation in an individual galaxy's metallicity from the MZR of the whole population.

Ultimately, the observations so far reveal no strong evidence for the evolution of the stellar MZR, though the simulations suggest otherwise. If evolution is present, conceivably it has not been detected because earlier observations have been limited to the most massive galaxies. De Rossi et al. (2017) predicted that the evolution of the stellar MZR is 0.2 dex from $z = 1$ to $z = 0$ for a mass of $10^{9.5} M_{\odot}$ but less than 0.05 dex at $10^{10.5} M_{\odot}$ over the same redshift span. Choi et al.'s (2014) sample at $z > 0.4$ is limited to massive galaxies ($M_* > 10^{10.8} M_{\odot}$). The passive galaxies at $z = 0.7$ in Gallazzi et al.'s (2014) sample are also massive ($M_* > 10^{10.5} M_{\odot}$) and have large uncertainties in [Fe/H] of the order of ~ 0.2 dex. Onodera et al. (2015) used Lick indices to measure the age and stellar metallicity of a stacked spectrum of passive galaxies at $z = 1.6$, the highest redshift at which stellar metallicity has been measured. The masses were also limited to $M_* > 10^{11} M_{\odot}$, where minimal evolution is expected. Since all of the observations at higher redshifts so far are limited to the massive end, it is not surprising that there is no statistically significant evidence for the evolution of the stellar MZR. In fact, the MZR of Gallazzi et al. (2014) tentatively suggests a stronger chemical evolution in lower-mass galaxies even among the quiescent populations.

This paper is the first in a series to present individual stellar metallicity measurements from $z = 0$ to $z = 1$. For the first time, we report a detection of $> 5\sigma$ in the evolution of the stellar MZR with redshift based on the stellar MZR of 62 early-type galaxies from the galaxy cluster Cl0024+1654 at $z \sim 0.4$. To our knowledge, this is the first attempt to measure metallicities using full-spectrum synthesis from individual spectra that extend to stellar masses as low as $10^{9.7} M_{\odot}$ beyond the local universe. The major advances in this paper that made the detection possible are (a) the measurement of ages and metallicities in individual galaxies at $z > 0$; (b) the extension of the mass range to low-mass galaxies; and (c) the use of full-spectrum synthesis in deriving the parameters. Aside from the evolution with observed redshift (at 0.037 ± 0.007 dex per Gyr), we also detect an even stronger evolution of the stellar MZR with the redshifts at which the galaxies formed (at 0.055 ± 0.006 dex per Gyr).

In Section 2, we describe the data and their completeness. In Section 3, we describe the method we used to measure ages and metallicities including how it performed as a function of signal-to-noise ratio and when the assumption of the single stellar population was dropped. Further comparisons between our measurements and the measurements from the literature can be found in the Appendix. In Section 4, we present the MZR derived from a subsample of local SDSS quiescent galaxies. In Section 5, we discuss our results and demonstrate that there is evolution in the stellar MZR with both observed redshift and formation redshift, i.e., when galaxy ages are taken into account. We also examine the relatively gentle slope found in the observed MZR and how it relates to the feedback strength of galaxies in the observed mass range. Lastly, we discuss the impact of cluster environments on our results in Section 6. Throughout this paper, we assume cosmological parameters $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$, and $H_0 = 70$ km s^{-1} Mpc^{-1}.

2. Data

We leverage a large imaging and spectroscopic survey of the $z \simeq 0.4$ galaxy cluster Cl0024+1654 (Treu et al. 2003; Moran et al. 2005, 2006, 2007a, 2007b). The study provides comprehensive Keck spectroscopy and *Hubble Space Telescope* (*HST*) imaging of Cl0024+17 members. Studying galaxy clusters offers the advantage of being able to obtain a large number of spectra in a few Keck pointings.

Moran et al. (2007b) provide complete details of the survey. In summary, the *HST* imaging of Cl0024+17 consists of 39

sparsely sampled WFPC2 images taken in the F814W filter at an exposure time of 4–4.4 ks each. The imaged field spans up to 5 Mpc from the cluster center, a significantly larger radius than the cluster's virial radius of 1.7 Mpc. Supplementing the *HST* imaging, the study also provides infrared imaging in the K_s-band and *J*-band from the WIRC camera on the Hale 200 inch telescope, optical imaging in the *BVRI* bands from the 3.6 m Canada–France–Hawaii Telescope, and both near- and far-ultraviolet (NUV/FUV) imaging on the *Galaxy Evolution Explorer* (*GALEX*) satellite. We used the photometric catalog from the survey's website,[5] which includes all optical and infrared photometric measurements, *HST* morphologies, and photometric/spectral redshifts. We downloaded the UV catalog separately through the Mikulski Archive for Space Telescopes (MAST). All UV galaxy images were visually examined with their optical counterparts to ensure the correct associations.

The survey used DEIMOS (Faber et al. 2003) on the Keck II telescope to obtain deep spectra of 300 member galaxies to $M_V = -18$. The observations took place between 2003 and 2005. The targets were selected with priority given to known cluster members up to $I = 22.5$ with classified *HST* morphologies. A total of 16 masks were observed with integration times of 2–4 hr each. Twelve masks were observed with the 900 line mm^{-1} grating from 2003 to 2004, while the rest were observed with the 600 line mm^{-1} grating in 2005 (Moran et al. 2007b). All slitlets were 1″ wide, yielding spectral resolutions of $R \sim 2000$–3000. All masks were centered at ~ 6200 Å, providing a spectral range of 3500 to 5500 Å in the rest frame. The spectroscopic sample is >65% complete for objects with $m_{\mathrm{F814W}} < 21.1$.

The DEIMOS spectra were reduced using the spec2d DEIMOS data reduction pipeline (Cooper et al. 2012; Newman et al. 2013) adapted by Kirby et al. (2015). Each spectrum was flat-fielded, wavelength-calibrated, sky-subtracted, and telluric-corrected.

We selected subsamples of quiescent galaxies with an average signal-to-noise ratio (S/N) greater than 8 Å^{-1} in the observed frame to ensure that we can break the degeneracy between age and metallicity (see Section 3.3). We define quiescent galaxies as those with rest-frame equivalent widths (EWs) of [O II] $\lambda 3727$ smaller than 5 Å and rest-frame FUV – *V* colors larger than 3. The EW limit is roughly equivalent to having a cut in specific star formation rate (sSFR)[6] at approximately $10^{-11} M_\odot$ yr^{-1}. The color cut is intended to further minimize contamination by star-forming galaxies. Specifically, the cut eliminates galaxies with star formation more recent than 10^7–10^8 yr (see Moran et al. 2006). The spectral S/N is estimated as the inverse of the average ratio of the absolute deviation of the observed spectrum to the best-fit spectrum of all pixels in the continuum region. We quote the S/N per angstrom rather than per pixel.

The [O II] EW, FUV – *V* color, and S/N criteria reduced the sample from the original total of 300 observed DEIMOS spectra to 62 quiescent spectra with sufficient S/N. Figure 1 shows the fraction of the sample selected from the parent sample as a function of F814W magnitude and stellar mass. Stellar masses and rest-frame colors were derived from available photometry using the SDSS KCORRECT software
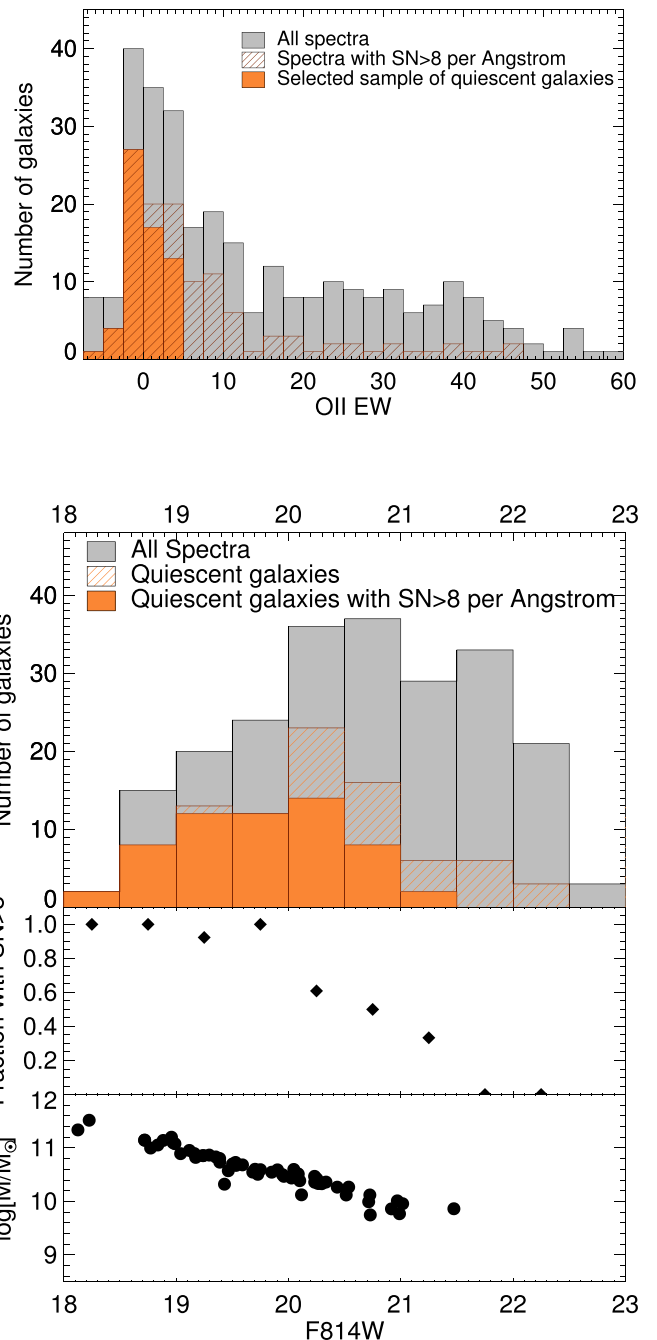

Figure 1. Characteristics of the observed spectra in the Cl0024+17 cluster. (Top) Histograms of [O II] $\lambda 3727$ rest-frame equivalent widths. In this study, we use samples of quiescent galaxies (EW < 5 Å and FUV – V > 3) with S/N > 8 Å^{-1}. (Bottom) Histograms of galaxies as a function of F814W magnitude. The bottom panel shows magnitude as a function of stellar mass. Our sample is ~50% complete at $M_* > 10^{9.7} M_\odot$.

version v4_3 (Blanton & Roweis 2007), which assumes population synthesis models of Bruzual & Charlot (2003) and the stellar initial mass function (IMF) of Chabrier (2003). The fraction of quiescent galaxies with S/N > 8 Å^{-1} is ~75% for F814W < 21.1. Since the parent survey is >65% complete in the same magnitude range for the DEIMOS spectra, our final sample is therefore ~50% complete for F814W < 21.1 or $M_* \gtrsim 10^{9.7} M_\odot$.

3. Model Fitting

In this section, we describe the full-spectrum fitting technique used to derive ages and metallicities. Section 3.1 describes the stellar population synthesis models used in this work. In Section 3.2, we detail our fitting technique. We test the accuracy of our measurements against our assumptions in Section 3.3.

3.1. Model Spectra

We derived stellar metallicities and ages of the galaxies using the full-spectrum fitting technique via stellar population synthesis (SPS). In the past decade, SPS models have been refined and often used as alternative tools to photometric indices to derive stellar population parameters such as stellar mass, star formation history (SFH), metallicity, and age (e.g., Schiavon et al. 2006; Walcher et al. 2009; Choi et al. 2014; Fumagalli et al. 2016). The commonly used SPS models include those of Bruzual & Charlot (2003, BC03), Fioc & Rocca-Volmerange (1999, PEGASE), Maraston (2005, M05), and Conroy et al. (2009, FSPS). The main advantage of using SPS over photometric indices is that it utilizes information from the whole spectrum simultaneously instead of using portions with the strongest stellar absorption features. Consequently, spectra with lower signal-to-noise ratios can be used to achieve the same level of precision.

We adopted the Flexible Stellar Population Synthesis (FSPS) models (Conroy et al. 2009) among the publicly available SPS models to derive the stellar population parameters from our sample. FSPS utilizes the most recent model of the Padova stellar evolution tracks (Marigo et al. 2008). The model includes treatments of thermally pulsating asymptotic giant branch stars (TP-AGBs) and an option to include horizontal-branch (HB) and blue straggler (BS) stars. The TP-AGBs are particularly important because leaving them out can result in systematic differences in age by a factor of 2 (Maraston 2005; Conroy et al. 2009). Lastly, the model offers flexibility in modeling the spectra that match the spectral resolution and the ranges of stellar parameters in our study.

We generated the templates of single stellar population (SSP) spectra using the FSPS code version 3.0. The SSP spectra were generated with the Kroupa IMF (Kroupa 2001) and the MILES spectral library (Sánchez-Blázquez et al. 2006). We chose the MILES spectral library because it spans a wide range in the stellar parameters, i.e., $\log g$ and [Fe/H], that are suitable for measuring galaxies at higher redshifts with lower metallicity. In addition, the library has a spectral resolution of 2.3 Å FWHM across the wavelength range 3525–7500 Å, comparable to that of our spectra. The spectra were interpolated from 22 modeled metallicities ranging from $\log Z = -1.98$ to 0.2, corresponding to the metallicity values of the Padova isochrones. The age ranges from 0.3 Myr to 14 Gyr. The rest of the parameters were set to the default mode in generating SSP spectra, which means that dust, blue HB stars, and BS stars were excluded. Because dust absorption mainly affects the continuum of the spectra but not the absorption lines, and because we remove the continuum from the observed spectra in our fitting procedure, we omitted dust for simplicity. The omission of blue HB and BS stars should not affect the derived ages and metallicities because these stars mainly contribute to ultraviolet wavelengths, which are not observed in our spectra.

3.2. Measurements of Metallicities and Ages

We iteratively fit each spectrum to FSPS models via χ^2 minimization. First, we created a mask for continuum normalization where all emission lines, strong absorption lines, and telluric regions are masked out. We continuum-normalized each galaxy spectrum with B-spline fitting with breakpoints at every $\sim$100 Å. We then created a different mask for the fitting procedure. In this mask, the Mg b triplet and emission lines (if [O II] was detected in emission) were masked out from the continuum-normalized spectra. The Mg b triplet was masked out so that the measured [Fe/H] would better reflect iron abundance rather than magnesium abundance. For each iteration, we simultaneously fit for four parameters: [Fe/H], age, velocity dispersion, and redshift. The priors were uniform for [Fe/H] and age. [Fe/H] was in the range [−1.98, 0.2], while the age was in the range from 0.3 Myr to the age of the universe at the galaxy's observed redshift. We set the prior for velocity dispersion according to the Faber–Jackson relation (Faber & Jackson 1976) between the velocity dispersion and the stellar mass measured by Dutton et al. (2011a), with a range of ±0.4 dex, which is large enough to capture any uncertainties and evolution with redshift (Dutton et al. 2011b).

The spectral fitting proceeded as follows. First, we used the IDL code MPFIT (Markwardt 2012) to fit the continuum-normalized observed spectrum with the SSP spectra. In this first fitting iteration, all model SSP spectra were continuum-normalized in the same manner as the observed spectrum using the same continuum mask. By fitting continuum-normalized spectra in the first iteration, we can bypass uncertainties from flux calibration and dust absorption, and uncertainties of the continuum flux in the modeled SSPs. This is also important because the MILES library provides spectra that have been normalized to unity.

For the rest of the iterations, we did not continuum-normalize any model spectra in order to minimize any alteration to the model spectra. Instead, we applied a "synthesized" continuum curve to the observed continuum-normalized spectrum. To do so, (1) we divided the observed, continuum-normalized spectrum by the best-fit SSP model spectrum from the previous iteration; (2) we fit the result from (1) with a B-spline using the same continuum mask; (3) we divided the observed continuum-normalized spectrum by the continuum curve from (2) to create an observed spectrum with a "synthesized" continuum; (4) we refit the resulting spectrum with the SSP model spectra—including the continuum shape—using the MPFIT code. The process (1)–(4) was repeated for 100 iterations, more than enough for the parameters to converge. We show examples of the observed spectra and their best-fit spectra in Figure 2.

3.3. Accuracy of Metallicities and Ages

In this section, we first examine the statistical uncertainties as a function of signal-to-noise ratio of the spectra and when the assumption of SSP is dropped. We then explore the systematic uncertainties that arise from the age–metallicity degeneracy. Lastly, we refer the reader to the Appendix for comparisons between our age and metallicity measurements and those in the literature.
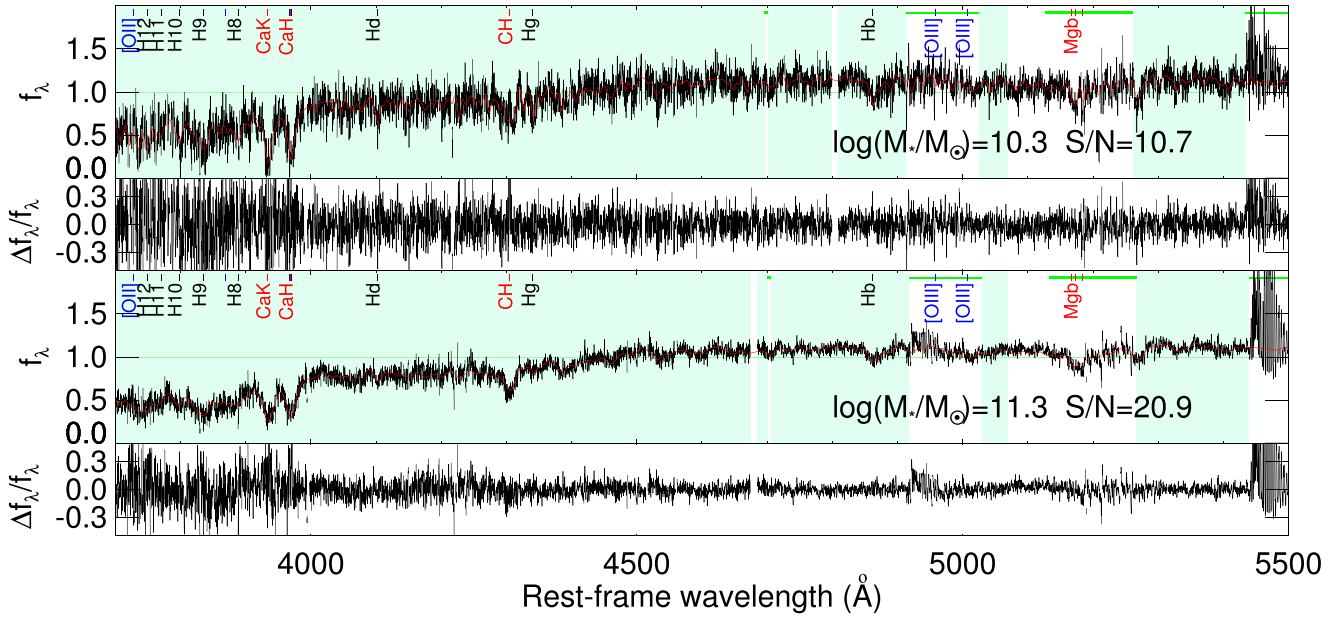
Figure 2. Examples of observed $z \sim 0.4$ spectra (black) and the corresponding best-fit models (red). The flux is continuum-normalized with an applied "synthesized" continuum. The model spectra are normalized by their median flux and smoothed to the instrumental resolution and best-fit velocity dispersion. The teal background shows the spectral regions used for spectrum modeling while the white background shows the spectral regions that are masked out. The green bars show regions with strong telluric absorption lines. We display the fractional residuals in the bottom panel of each spectrum. The measured metallicities and ages of the two galaxies are $[\mathrm{Fe/H}] = -0.14^{+0.13}_{-0.09}, -0.15^{+0.04}_{-0.06}$ and Age $= 3.2^{+1.1}_{-0.6}, 6.3^{+0.7}_{-0.8}$ Gyr, respectively. The uncertainties include the systematic uncertainties from the age–metallicity degeneracy.

3.3.1. Dependence on Signal-to-noise Ratios

To investigate how the observed S/Ns influence the uncertainties in the measured ages and metallicities, we tested our spectral fitting code on a set of mock DEIMOS spectra with different S/Ns. We first adopted the SSP assumption. We created a mock SSP spectrum from the FSPS code with $\log Z = -0.1$ dex and an age of 3 Gyr. These numbers were chosen to be representative of our $z \sim 0.4$ data. The spectrum was smoothed to have a velocity dispersion of FWHM $= 250$ km s^{-1} and the same spectral range and resolution as a typical DEIMOS spectrum. Gaussian noise was added to the spectrum to create 20 spectra for each S/N in the range from $\sim$3 to 30 Å^{-1}. We then multiplied the spectra with a telluric transmission curve and the DEIMOS instrumental throughput to mimic the observed spectra.

We found that given the SSP assumption, we can measure [Fe/H] and age well to $\sim$0.05 dex precision when the S/Ns of the spectra are >8 Å^{-1}, which is the minimum S/N in our sample. The uncertainties in [Fe/H] and age as a function of S/N are shown in Figure 3. The fluctuation in the measurements decreases rapidly as a function of S/N. However, once the S/N is higher than $\sim$8 Å^{-1}, the fluctuations decrease slowly. The level of the fluctuation does not change much when we change the age of the mock spectra from 3 to 8 Gyr. With uncertainties $\lesssim$0.05 dex, we conclude that we only minimally suffer from statistical uncertainties.

3.3.2. Validity of the SSP Assumption

Next, we test the validity of the SSP assumption. We created a set of mock spectra of a composite stellar population (CSP) with an exponentially declining star formation history $\psi(t') \propto \exp(-t'/\tau)$ with the following parameters: $\tau = 1$ Gyr, $\log(Z/Z_\odot) = -0.2$, and an elapsed time ranging from 1 to 8 Gyr since the onset of the star formation. We added Gaussian noise so that the S/N is 12 Å^{-1}, the average S/N of our spectra.
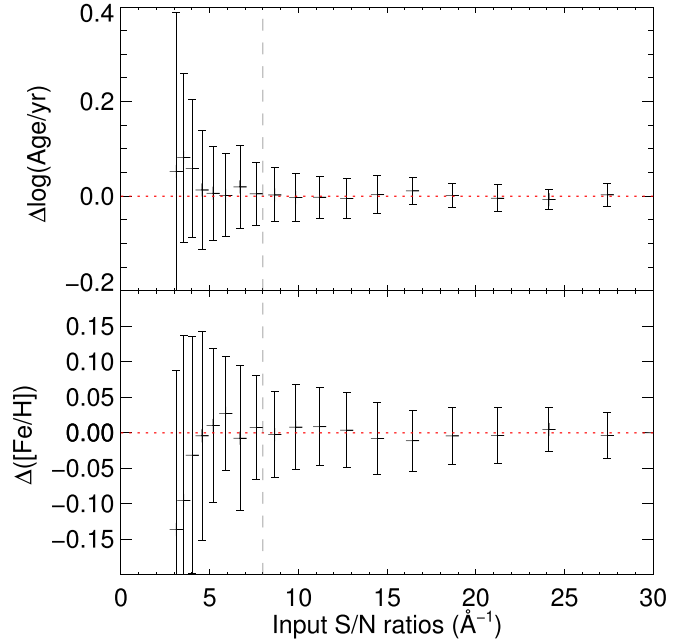
Figure 3. Accuracy in measurements of ages and metallicities as a function of S/N when assuming an SSP. Each data point shows the mean and standard deviation of 20 mock SSP spectra with a certain S/N. At S/N $\gtrsim 8$ Å^{-1}, we can recover the metallicities within 0.05 dex.

We applied telluric features, the instrumental throughput, and a smoothing kernel to the mock spectra in the same manner as in Section 3.3.1. We note that all stars in these spectra have the same metallicity and this should be interpreted as the population's light-weighted metallicity.

We then fit the CSP model spectra with SSP models. In general, we recovered the CSP age and metallicity within $\sim$0.1 dex precision when most of the star formation has been

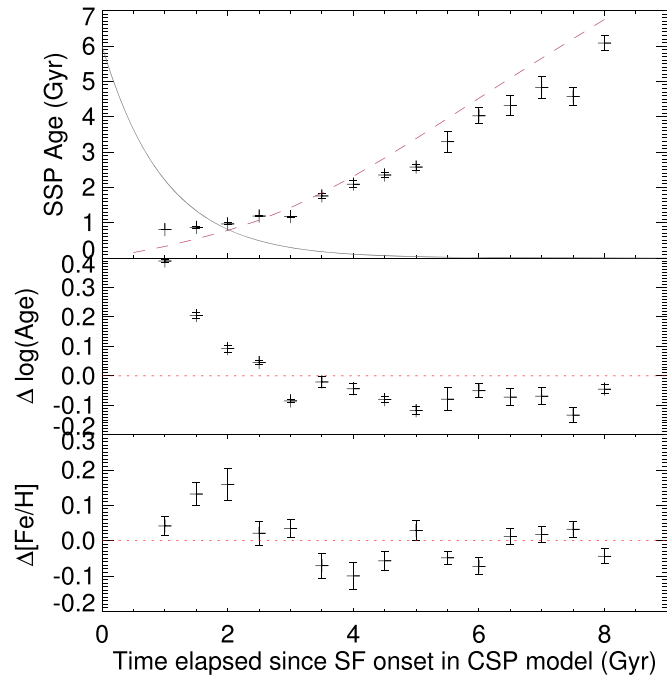**Figure 4.** Accuracy in measurements of ages and metallicities when SSP is not assumed. (Top) Measured SSP ages from the mock spectra of composite stellar populations. The spectra have $S/N = 12 \, \text{Å}^{-1}$ and the same star formation history but were observed at different amounts of time elapsed after the onset of the star formation. The solid curve shows the assumed exponentially declining star formation history. The red dashed line shows the light-weighted age calculated over the range 3700–5500 Å. (Middle and bottom) The deviation of the measured ages and metallicities from the input values. The middle panel shows the difference between the measured SSP ages and the red dashed line in the top panel.

quenched. The measured metallicities and ages are shown in Figure 4. The gray curve in the top panel shows the shape of the exponentially declining star formation rate. The red dashed line is plotted as a guide for the "light-weighted" age of the population at elapsed time t via

$$\text{Age}(t) \sim \frac{\int_0^t (t - t') L(t') \psi(t') dt'}{\int_0^t L(t') \psi(t') dt'}$$

where $L(t')$ is the integrated light in the wavelength range of rest-frame 3700 to 5500 Å produced by SSP stars of age t'.

Although the integrated light $L(t')$ accounts for the massive stars that died before reaching age t', the integration is over every star that has formed ($\psi(t')dt'$) and does not account for the stars that have died. Therefore we expect this "light-weighted" age to overestimate the time elapsed since the beginning of star formation, especially for older populations. The overestimate of age in old populations might partially explain the behavior shown in the middle plot. The SSP-equivalent ages that we measured are consistently younger than the respective "light-weighted" ages of the CSP population.

The measured metallicities fluctuate well within 0.1 dex of the true answer, and they do not seem to be affected by assuming an SSP rather than a CSP. We conclude that we can measure metallicity to a precision of ∼0.1 dex under the SSP assumption. On the other hand, we likely underestimated ages by an amount less than 0.1 dex as long as the majority of the star formation has been quenched, i.e., the light-weighted age is greater than ∼1.5 Gyr. Because our sample is comprised of

quiescent galaxies, we expect that our age and metallicity measurements are not greatly affected by the SSP assumption.

Our results regarding the SSP assumption are similar to the results of Trager & Somerville (2009). These authors concluded that an SSP-equivalent metallicity is an excellent tracer of the light- or mass-weighted metallicities, whereas an SSP-equivalent age generally biases toward values younger than the true mass- and light-weighted ages. A level of bias in ages of ∼−0.1 dex compared to the true light-weighted age is also similar to ours. Trager & Somerville argued that hot young stars contribute minimally to the metal lines but heavily to the Balmer lines relative to old stars.

3.3.3. Age–Metallicity Degeneracy

For each galaxy, we estimated the systematic uncertainty arising from the age–metallicity degeneracy. As shown in Figure 4, the statistical uncertainties obtained from MPFIT underestimate the level of total uncertainty. To estimate total uncertainty, we created a mock SSP spectrum with the same age and metallicity as each observed galaxy. Gaussian noise was added to the spectrum to reach the same S/N as the observed spectrum. We then compared the noised spectrum to a 100×120 grid of noiseless SSP spectra with a range of [0.5, 13] Gyr in age and [−0.8, 0.2] dex in [Fe/H], and calculated a χ^2 array for the noised mock spectrum. All SSP spectra were smoothed to achieve a velocity dispersion of $250 \, \text{km s}^{-1}$ FWHM convolving with the typical resolution of an SDSS or a DEIMOS spectrum.

The uncertainties in age and metallicity are calculated by marginalizing the 2D posterior probability distribution obtained from the χ^2 array. The uncertainties for upper (lower) limits are the differences between the values at 84th (16th) percentile and at the 50th percentile in the posterior probability distributions. Since we calculate these uncertainties specifically to the S/N of each galaxy, we take these uncertainties as the total uncertainty of each measurement. The uncertainties quoted in the subsequent text and figures refer to these systematic uncertainties.

In general, the uncertainties in ages and metallicities are neither Gaussian nor symmetric. Figure 5 shows contours of 1σ uncertainty based on the 2D posterior probability distribution functions of the age–metallicity degeneracy at $S/N = 10 \, \text{Å}^{-1}$ and $S/N = 25 \, \text{Å}^{-1}$. The average total uncertainties for upper and lower limits are +0.11 and −0.14 dex for [Fe/H] and +0.12 and −0.11 dex for age, respectively. We list all our measurements of ages and metallicities in Table 1.

4. Mass–Metallicity Relation of Local Galaxies

In this section, we report the MZR of SDSS local galaxies measured with our method. We will use this MZR as a reference to compare with the MZR of $z \sim 0.4$ galaxies in Section 5.

Figure 6 shows the stellar mass–stellar metallicity relation we measured for local quiescent SDSS galaxies. We selected a subsample of 155 quiescent galaxies from Gallazzi et al. (2005). The selection criteria and detailed comparisons to the age and metallicity measurements of Gallazzi et al. are described in the Appendix. In order to ensure that the stellar masses of both the $z \sim 0$ and $z \sim 0.4$ samples are on the same scale, we remeasured the stellar masses of the local galaxies with the KCORRECT code using the SDSS u, g, r, i, z
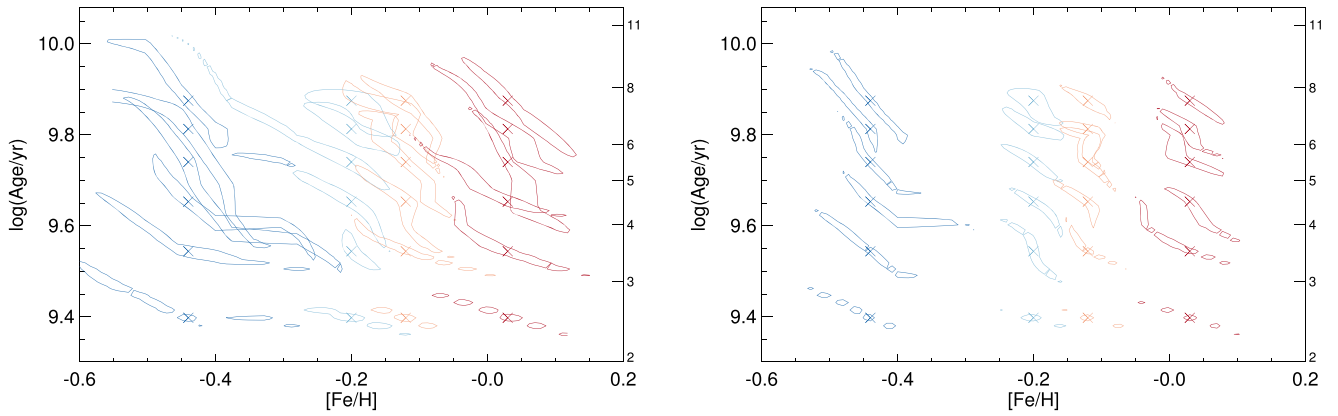
Figure 5. Uncertainties in the measured ages and metallicities according to the age–metallicity degeneracy. Each contour shows the 1σ range in age and metallicity. The true ages and metallicities are shown as cross marks. The S/N of each spectrum is 10 Å^{-1} (left) and 25 Å^{-1} (right).

photometry. The MZR of local galaxies shows the expected relation: metallicity increases with stellar mass. However, the "knee," or where the slope in the MZR changes, is less visible from our measurements than from Gallazzi et al. (2005). The MZR is generally consistent with the MZR of the complete galaxy population (both star-forming and quiescent galaxies) analyzed by Gallazzi et al. at stellar masses above $M_* \approx 10^{10.3} M_\odot$. However, at lower stellar masses, the MZR of the subsample of quiescent galaxies shows metallicities higher than those of Gallazzi et al..

We consider three functions to approximate the MZR. First, we try fitting with the three-parameter logarithmic function proposed by Moustakas et al. (2011) to describe the gas-phase MZR:

$$[\mathrm{Fe/H}] = [\mathrm{Fe/H}]_0 - \log\left[1 + M_*/M_0\right]^\gamma. \qquad (1)$$

We then fit with a quadratic function, another proposed form of gas-phase MZR (e.g., Tremonti et al. 2004), and a linear function. For each function, we fit with 1000 iterations of the Monte Carlo random sampling method because the uncertainties of our measurements are not Gaussian. In each iteration, we resample according to the probability distribution in [Fe/H] of each galaxy and find the best-fit function by minimizing the chi-square error statistic.

The MZR of the subsample of quiescent galaxies is best described by a linear fit with a modest slope. The best-fit logarithmic function (the bottommost dashed-dotted line in Figure 6) performs the worst in terms of χ^2, at approximately 1.6 times the minimum χ^2 of the two other functions. Both quadratic and linear functions fare equally well based on the minimum χ^2. However, according to the Akaike information criterion (Burnham & Anderson 2003), the linear function is a better choice because it is less complex and minimizes the loss of information. The slope of the best linear fit is $\sim0.16 \pm 0.02$ dex in [Fe/H] per logarithmic mass unit (per logmass)—consistent with the slope found by Gallazzi et al. (2006). The slope indicates that an increase in galaxy mass by a factor of 10 corresponds to an increase in metallicity by a factor of 1.4.

When compared with the MZR measured by Gallazzi et al. (2005), it is found that higher metallicities in the low-mass galaxies can directly result from the systematic difference in the metallicity measurements. We tried plotting the MZR using the stellar masses reported in Gallazzi et al. (2005), which were measured from spectral indices and z-band photometry as

opposed to the KCORRECT code used in this paper. Because the mass measurements from the two methods are generally consistent within 0.1 dex, the best-fit linear function does not depend on the difference in the mass measurement methods. The systematic difference in metallicity measurement is therefore the cause of the difference in the MZR at the lower-mass end.

As discussed in the Appendix, for the same galaxies with low metallicities, we measured their [Fe/H] to be higher than what Gallazzi et al. (2005) measured. Since galaxies with low metallicities are mainly less massive galaxies, this results in the MZR lying above that shown by Gallazzi et al. (2005). The MZR with higher metallicities at the lower-mass end found here is consistent with the MZRs found in stacks of local spectra in Choi et al. (2014) and in individual local early-type galaxies via integral field spectroscopy by Sybilska et al. (2017). The metallicities measured by Sybilska et al. (2017) were measured via spectroscopic indices using the SSP models from Vazdekis et al. (2015), which is also based on the MILES library. Because the measurements in Choi et al. (2014), Sybilska et al. (2017), and this work utilize the same stellar library, it is very likely that the choice of stellar library used caused the systematic differences in the metallicity measurements from those of Gallazzi et al. (2005).

Whether the higher metallicity at the lower-mass end of the MZR can be additionally caused by differences in SFH between the two samples, on top of the measurement methods, is still ambiguous. There is observational evidence that the stellar MZR of early-type galaxies might differ from that of late-type galaxies. Although categorizing galaxies based on their morphologies is not necessarily the same as categorizing based on their SFH, the two properties closely correlate with each other in both local and high-redshift galaxies (e.g., Kriek et al. 2009; Wuyts et al. 2011; Lee et al. 2013). Gallazzi et al. (2005) found that morphology, described by a concentration index, is responsible for a difference as large as ~0.6 dex in [Fe/H] in low-mass galaxies. The difference between the two populations in more massive galaxies is minimal.

On the other hand, theoretical work suggests that there should not be a significant difference between star-forming and passive galaxies. Okamoto et al. (2017) created separate MZRs for passive and star-forming galaxies based on the Illustris simulation (Nelson et al. 2015, Illustris-1) and the EAGLE simulation (McAlpine et al. 2016). Okamoto et al. (2017) showed that the difference between passive and star-forming

Table 1
Catalog of Measured Ages and Metallicities

No.	R.A.	Decl.	$\log M_*$	[Fe/H]	Age	S/N	No.	R.A.	Decl.	$\log M_*$	[Fe/H]	Age	S/N
1	00 25 51.07	+17 08 42.4	10.8	$+0.10^{+0.08}_{-0.10}$	$2.4^{+0.3}_{-0.2}$	16.0	32	00 26 31.73	+17 12 24.1	11.1	$-0.09^{+0.12}_{-0.07}$	$3.3^{+1.1}_{-0.8}$	15.4
2	00 25 54.52	+17 16 26.4	10.4	$-0.16^{+0.19}_{-0.17}$	$2.9^{+2.3}_{-1.0}$	9.0	33	00 26 31.81	+17 11 56.9	10.4	$-0.02^{+0.08}_{-0.08}$	$7.1^{+1.7}_{-1.0}$	8.0
3	00 25 57.73	+17 08 01.5	10.4	$-0.12^{+0.08}_{-0.09}$	$5.7^{+1.5}_{-2.0}$	8.4	34	00 26 32.50	+17 10 26.0	10.9	$-0.07^{+0.12}_{-0.20}$	$2.5^{+0.8}_{-0.3}$	10.6
4	00 26 04.44	+17 20 00.6	10.5	$-0.22^{+0.39}_{-0.17}$	$2.0^{+0.7}_{-0.8}$	8.6	35	00 26 32.71	+17 07 56.0	10.3	$-0.20^{+0.12}_{-0.27}$	$2.0^{+0.4}_{-0.4}$	10.0
5	00 26 05.80	+17 19 19.0	10.9	$-0.05^{+0.07}_{-0.10}$	$5.4^{+2.1}_{-2.6}$	11.9	36	00 26 33.54	+17 09 23.9	10.6	$-0.11^{+0.14}_{-0.60}$	$1.3^{+0.7}_{-0.1}$	19.3
6	00 26 05.84	+17 19 19.0	10.9	$-0.23^{+0.18}_{-0.20}$	$6.0^{+4.3}_{-2.2}$	11.5	37	00 26 33.60	+17 09 20.2	10.5	$-0.13^{+0.09}_{-0.07}$	$10.0^{+1.4}_{-2.2}$	8.1
7	00 26 06.95	+17 19 42.8	10.8	$-0.14^{+0.06}_{-0.05}$	$4.6^{+0.7}_{-1.0}$	25.7	38	00 26 33.66	+17 09 31.0	10.5	$-0.07^{+0.07}_{-0.07}$	$4.7^{+1.1}_{-1.7}$	12.3
8	00 26 08.85	+17 09 54.7	10.7	$-0.14^{+0.12}_{-0.20}$	$2.9^{+1.0}_{-0.4}$	18.3	39	00 26 33.81	+17 12 16.6	10.6	$-0.02^{+0.15}_{-0.07}$	$3.4^{+1.5}_{-0.6}$	8.4
9	00 26 09.66	+17 11 13.5	10.4	$-0.42^{+0.15}_{-0.15}$	$7.5^{+1.9}_{-3.2}$	8.1	40	00 26 34.35	+17 10 22.1	11.1	$-0.16^{+0.07}_{-0.07}$	$4.1^{+0.7}_{-1.0}$	27.4
10	00 26 13.92	+17 13 34.9	10.1	$-0.33^{+0.24}_{-0.17}$	$2.4^{+0.7}_{-0.4}$	8.1	41	00 26 34.59	+17 10 16.4	10.8	$-0.25^{+0.10}_{-0.15}$	$5.6^{+2.4}_{-1.7}$	17.9
11	00 26 15.16	+17 18 15.6	10.4	$-0.08^{+0.16}_{-0.31}$	$2.7^{+2.2}_{-0.5}$	10.8	42	00 26 34.98	+17 10 21.3	10.5	$+0.04^{+0.09}_{-0.06}$	$4.0^{+2.3}_{-1.0}$	16.4
12	00 26 18.45	+17 07 01.1	10.3	$-0.27^{+0.18}_{-0.12}$	$4.0^{+3.0}_{-1.0}$	9.0	43	00 26 35.70	+17 09 43.1	11.5	$+0.01^{+0.06}_{-0.05}$	$5.6^{+1.0}_{-1.1}$	27.8
13	00 26 21.49	+17 14 11.8	10.6	$-0.08^{+0.06}_{-0.08}$	$5.5^{+1.5}_{-2.0}$	13.3	44	00 26 36.77	+17 09 28.7	10.1	$-0.23^{+0.12}_{-0.15}$	$4.0^{+2.3}_{-0.6}$	10.5
14	00 26 22.90	+17 12 31.4	11.1	$+0.20^{+0.04}_{-0.09}$	$2.8^{+0.2}_{-0.1}$	23.7	45	00 26 37.27	+17 10 00.2	10.6	$+0.10^{+0.12}_{-0.12}$	$5.6^{+1.3}_{-1.7}$	15.3
15	00 26 22.91	+17 12 31.3	11.1	$-0.05^{+0.10}_{-0.08}$	$4.3^{+1.6}_{-2.0}$	9.9	46	00 26 37.52	+17 09 08.7	10.6	$-0.18^{+0.07}_{-0.07}$	$6.9^{+1.0}_{-0.9}$	16.8
16	00 26 24.82	+17 12 21.5	10.9	$-0.14^{+0.11}_{-0.30}$	$5.2^{+5.2}_{-2.0}$	12.3	47	00 26 37.90	+17 09 22.0	11.2	$-0.01^{+0.05}_{-0.06}$	$4.8^{+1.5}_{-0.8}$	26.3
17	00 26 26.11	+17 11 57.8	10.3	$-0.02^{+0.15}_{-0.15}$	$5.4^{+3.2}_{-3.6}$	8.3	48	00 26 37.91	+17 09 37.8	10.7	$-0.07^{+0.09}_{-0.05}$	$3.8^{+1.2}_{-0.8}$	14.5
18	00 26 27.12	+17 12 25.9	10.9	$-0.16^{+0.12}_{-0.15}$	$9.0^{+2.7}_{-3.1}$	12.1	49	00 26 38.41	+17 09 58.7	10.3	$-0.14^{+0.13}_{-0.09}$	$3.2^{+1.1}_{-0.6}$	10.7
19	00 26 27.98	+17 11 37.7	10.5	$-0.30^{+0.12}_{-0.22}$	$8.5^{+3.1}_{-2.5}$	8.2	50	00 26 38.65	+17 09 14.5	9.7	$-0.15^{+0.17}_{-0.28}$	$3.9^{+4.4}_{-2.9}$	10.9
20	00 26 29.11	+17 10 24.7	10.7	$-0.07^{+0.07}_{-0.04}$	$3.7^{+1.0}_{-0.7}$	18.0	51	00 26 38.80	+17 09 59.5	10.7	$-0.11^{+0.21}_{-0.18}$	$2.7^{+2.3}_{-1.0}$	24.5
21	00 26 29.50	+17 10 32.6	10.5	$-0.15^{+0.12}_{-0.14}$	$3.1^{+1.2}_{-0.5}$	18.3	52	00 26 40.13	+17 08 21.5	9.9	$-0.60^{+0.08}_{-0.09}$	$4.6^{+0.7}_{-0.9}$	8.7
22	00 26 29.92	+17 10 06.8	10.3	$-0.21^{+0.04}_{-0.21}$	$1.3^{+0.0}_{-0.1}$	16.4	53	00 26 41.16	+17 10 01.8	11.1	$+0.01^{+0.19}_{-0.25}$	$3.0^{+2.7}_{-1.4}$	16.0
23	00 26 30.08	+17 07 49.1	9.8	$-0.21^{+0.29}_{-0.26}$	$2.1^{+0.8}_{-0.5}$	7.9	54	00 26 43.23	+17 08 41.0	10.1	$-0.25^{+0.36}_{-0.18}$	$4.0^{+2.2}_{-5.1}$	11.5
24	00 26 30.76	+17 12 26.3	10.3	$-0.15^{+0.13}_{-0.22}$	$1.6^{+0.6}_{-0.3}$	13.2	55	00 26 43.70	+17 07 12.8	10.0	$-0.06^{+0.15}_{-0.17}$	$3.0^{+1.5}_{-0.4}$	8.2
25	00 26 31.00	+17 17 09.0	10.0	$+0.10^{+0.06}_{-0.10}$	$2.3^{+0.3}_{-0.2}$	17.7	56	00 26 44.68	+17 08 33.8	9.9	$-0.79^{+0.35}_{-0.13}$	$6.8^{+1.6}_{-0.9}$	8.3
26	00 26 31.04	+17 11 09.3	10.4	$-0.09^{+0.08}_{-0.22}$	$4.0^{+2.7}_{-1.7}$	12.5	57	00 26 48.22	+17 10 46.3	10.2	$-0.27^{+0.06}_{-0.16}$	$1.7^{+0.2}_{-0.1}$	8.3
27	00 26 31.22	+17 12 08.4	10.7	$+0.06^{+0.13}_{-0.08}$	$3.1^{+1.7}_{-1.1}$	21.3	58	00 26 48.30	+17 12 35.4	10.6	$-0.28^{+0.05}_{-0.17}$	$1.3^{+0.0}_{-0.2}$	16.2
28	00 26 31.40	+17 17 00.3	11.1	$-0.06^{+0.14}_{-0.17}$	$5.0^{+2.5}_{-3.3}$	13.5	59	00 26 51.84	+17 08 39.9	10.5	$-0.21^{+0.10}_{-0.18}$	$6.5^{+2.8}_{-1.6}$	15.7
29	00 26 31.41	+17 10 55.6	11.3	$-0.15^{+0.04}_{-0.06}$	$6.3^{+0.7}_{-0.8}$	20.9	60	00 26 54.38	+17 08 27.0	11.0	$+0.10^{+0.11}_{-0.09}$	$3.5^{+2.8}_{-1.4}$	24.8
30	00 26 31.41	+17 10 27.1	10.0	$-0.27^{+0.13}_{-0.16}$	$5.0^{+1.4}_{-3.5}$	8.0	61	00 27 07.76	+17 10 49.5	11.0	$+0.10^{+0.10}_{-0.13}$	$3.0^{+1.0}_{-0.3}$	24.8
31	00 26 31.56	+17 17 12.8	10.5	$-0.12^{+0.04}_{-0.06}$	$5.9^{+0.7}_{-1.3}$	22.5	62	00 27 25.16	+17 07 22.4	10.1	$-0.51^{+0.24}_{-0.18}$	$6.1^{+1.7}_{-3.4}$	12.0

Note. The units of M_*, age, and S/N are $M_\odot$, Gyr, and $\mathring{A}^{-1}$, respectively.

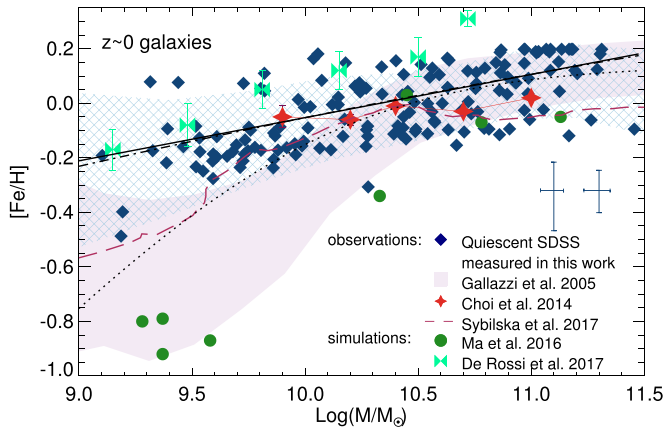(This table is available in machine-readable form.)

Figure 6. The stellar mass–metallicity relation of local quiescent galaxies. Each navy-blue diamond shows the metallicities measured in this work. The light-blue hatched strip shows the average and standard deviation of metallicities in each mass bin. The two error bars to the right of the plot show median uncertainties for galaxies with stellar mass lower (left) and higher (right) than $10^{10} M_\odot$. The black solid curve is the best-fit linear function while the black dashed-dotted and dotted lines are the best-fit quadratic and logarithmic (Equation (1)) functions, respectively. The light-purple solid strip shows the average relation from Gallazzi et al. (2005). The plot also shows local MZRs measured in other works (Choi et al. 2014; Sybilska et al. 2017) and the local MZRs found in the FIRE and EAGLE hydrodynamical simulations (Ma et al. 2016; De Rossi et al. 2017).

galaxies is ~0.05 dex or smaller at any fixed mass. Furthermore, they found that the shape of the MZR does not depend on galaxy type.

Regardless of the true shape of the MZR at smaller masses, we will use our measurements in Figure 6 as a point of reference when comparing with those at higher redshift. Comparing to our own measurement of the local galaxy populations reduces the systematic uncertainties introduced from measurements and sample selection.

5. Evolution in the Mass–Metallicity Relation

In this section we show the main result of this work. In Section 5.1, we report the MZR of individual $z \sim 0.4$ galaxies in which, for the first time, we detect an evolution in the stellar MZR with observed redshift. In Section 5.2, we explain the scatter in the MZR and report an even greater evolution of the MZR when the age of galaxies is taken into account. Lastly, in Section 5.3, we discuss the meaning of the MZR slope and what it reveals about feedback in galaxies.

5.1. Evolution of the MZR with Observed Redshift

We plot the stellar MZR of the $z \sim 0.4$ sample in Figure 7. The $z \sim 0.4$ MZR matches well with that measured by Choi et al. (2014) at the same redshift at $M_* > 10^{10.5} M_\odot$. However, in contrast to Choi et al. (2014), we measured the MZR of individual (rather than stacked) galaxies and extended the MZR relation almost 1 dex lower in stellar mass, to $10^{9.7} M_\odot$.

Based on the best-fit linear relations to the MZRs at $z \sim 0$ and $z \sim 0.4$, we find an evolution of the MZR with observed redshifts at a significance level greater than 5σ. As in Section 4, we fit a linear function to the measured $z \sim 0.4$ MZR via the Monte Carlo method. We then use the analysis of covariance to compare the best-fit linear functions of the MZR at $z \sim 0$ and $z \sim 0.4$. First, we check if the slopes of the two linear functions
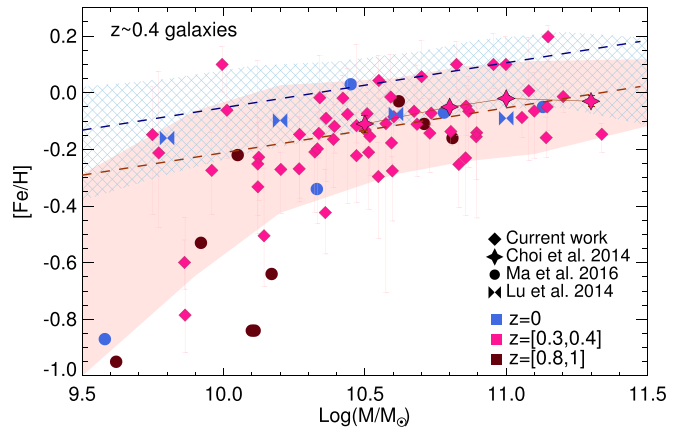

Figure 7. Stellar MZR relation of $z \sim 0.4$ quiescent galaxies. The solid orange and blue hatched strips show the average and standard deviation of metallicities in each mass bin of $z \sim 0.4$ and local galaxies (same as Figure 6). The upper dashed line and lower dashed line show the best-fit linear functions, where slopes were fixed to the common value, to the local and $z \sim 0.4$ galaxies, respectively. The best-fit parameters are shown in Equations (2) and (3). We also show the predicted MZR from the FIRE simulations (Ma et al. 2016) and from semi-analytic models with a constant mass-loading factor (Lu et al. 2014). The plot is color-coded by galaxy redshift.

are different. The best-fit slope for the $z \sim 0.4$ population is 0.15 ± 0.03 dex per logmass. This is the same as the slope of 0.16 ± 0.02 dex per logmass of the $z \sim 0$ population to within $\sim 1\sigma$ significance. Therefore, we conclude that the slopes of the MZRs at two redshifts are not significantly different. We then test for the evolution in the normalization values (the constant terms). To do so, we refit linear functions to the two MZRs using a common fixed slope, equal to the weighted-mean slope of 0.16 dex per logmass. The best-fit linear equations when the slopes are fixed are

$$\langle [\mathrm{Fe/H}] \rangle = (-0.05 \pm 0.01) + 0.16 \log\left(\frac{M_*}{10^{10} M_\odot}\right) \qquad (2)$$

for the MZR of $z \sim 0$ quiescent galaxies, and

$$\langle [\mathrm{Fe/H}] \rangle = (-0.21 \pm 0.02) + 0.16 \log\left(\frac{M_*}{10^{10} M_\odot}\right) \qquad (3)$$

for the MZR of $z \sim 0.4$ quiescent galaxies. The difference in the constant terms is 0.16 ± 0.03 dex, which is significant at a level greater than 5σ.

The observed metallicity evolution is larger than but consistent with the predictions from hydrodynamical simulations. For now, we focus on the shifts in the MZRs with redshift and ignore the slopes predicted in simulations when comparing to our observations. The observed metallicity evolution of 0.16 ± 0.03 dex from $z \sim 0.4$ to $z \sim 0$ translates to an increase in metallicity of 0.037 ± 0.007 dex per Gyr. Based on the FIRE hydrodynamical simulations, Ma et al. (2016) predicted that the stellar metallicity evolution from $z = 0.4$ to $z = 0$ should be 0.13 dex in [Fe/H]. The observed evolution is slightly larger than but consistent within 1σ with the evolution predicted from the FIRE hydrodynamical simulations. We found a larger discrepancy when comparing the observed evolution to the evolution predicted by De Rossi et al. (2017) based on the EAGLE hydrodynamical simulation. De Rossi et al. predicted evolution of 0.11 ± 0.09 dex from $z = 1$ to $z = 0$ in $10^{10} M_\odot$ galaxies (see their Figure 5), which

is $\sim 0.014 \pm 0.012$ dex per Gyr. Our observed evolution is therefore also greater, but due to the large uncertainty in the predicted evolution it is consistent within 2σ.

The observed evolution of stellar MZR likely emphasizes the importance of metal recycling in galaxies. Among the hydrodynamical simulations, the FIRE simulations predicted the strongest evolution of the stellar MZR with redshift (Ma et al. 2016), and this is the most consistent with our observations. As explained in Ma et al. (2016), this is because the galaxies in the FIRE simulations are able to retain more metals, resulting in a greater increase in metallicity with redshift. This is particularly true in galaxies with stellar mass above $10^{10} M_\odot$, where the retained fractions of metals in the halos are almost unity. The reason why different simulations achieve different metal retention fractions lies in the physical models on which the simulations are based. Many of the cosmological simulations (including the EAGLE simulations) adopt "sub-grid" empirical models of galactic winds and stellar feedback, where fractions of gas are forced to escape the galaxy as a result of energy injection from supernovae and stellar winds. In contrast, the FIRE simulations adopt a Lagrangian formulation of smooth particle hydrodynamics (Hopkins et al. 2014), where metallicities are derived from tracked individual star particles that can be ejected and, importantly, accreted back to the galaxy. As a result, the metal recycling effect is better captured in the FIRE simulations than in those that assume "sub-grid" models.

However, complex and more realistic simulations such as the FIRE simulations are computationally expensive and still limited in terms of sample size, which might explain the small discrepancy between the observed and predicted amount of evolution. The predicted evolution of 0.13 dex from $z = 0.4$ to $z = 0$ in Ma et al. (2016) came from fitting a linear function to the MZRs over a wide mass range, from $M_* \sim 10^4 M_\odot$ to $M_* \sim 10^{11.5} M_\odot$ at different redshifts. If we limit the mass range to above $10^{9.7} M_\odot$, as in our observations, there are only four simulated galaxies at $z = 0$ and eight at $z = 0.8$. With this limited mass range and sample size, the evolution of the MZR in the simulated galaxies is 0.2 ± 0.6 dex at $10^{10} M_\odot$ from $z = 0.8$ to $z = 0$ or 0.03 ± 0.09 dex per Gyr. Moreover, the FIRE simulations do not include feedback from possible active galactic nuclei (AGNs), which can potentially affect the metallicities in massive galaxies with $M_* > 10^{11} M_\odot$ (Ma et al. 2016). Although the evolution from $z = 0.8$ to $z = 0$ in the FIRE simulation is not significant, it is consistent with the better constrained values from our observations.

We note that the galaxies in both the EAGLE and the FIRE simulations are not necessarily passive, whereas galaxies in our sample are. The fact that the observed evolution of the MZR with redshift is consistent with the simulations does not have any implication on whether the two populations' metallicities are the same at any given redshift (the latter has been suggested by Okamoto et al. 2017). In this section, we compared only the magnitude of the change in metallicity with redshift, and not the metallicities themselves. In fact, none of the metallicity values are consistent. Different suites of simulations predict different MZR normalizations at each redshift, none of which are consistent with each other or with our observations (see Figures 6 and 7).

5.2. Can Galaxy Formation Time Explain the Evolution of the MZR with Observed Redshift?

We now measure the intrinsic scatter in the MZR and test for the correlation between the scatter and the galaxy's redshift of formation.

We repeat the linear fit to the stellar MZR with an additional parameter, an intrinsic variability σ_v, by minimizing the negative-log-likelihood

$$L = -\sum_i \log\left[P(\Delta[\text{Fe/H}]_i) * \mathcal{N}(0, \sigma_v^2)\right]|_{\Delta[\text{Fe/H}]_i = 0}$$

where $P(\Delta[\text{Fe/H}]_i)$ is the probability of the difference between the observed metallicity and the model linear equation $[\text{Fe/H}]_i - (a \log \frac{M_{*,i}}{10^{10} M_\odot} + b)$, which has the same shape as the probability of each observed metallicity $P([\text{Fe/H}]_i)$. This equation means that the probability of the deviation of each observed metallicity from the linear model is equal to the convolution between the probability function inherited from the age–metallicity degeneracy $P[\text{Fe/H}]_i$ and the Gaussian probability of an intrinsic scatter of size σ_v.

The intrinsic scatter in [Fe/H] is 0.07 ± 0.01 dex for both local galaxies and the $z \sim 0.4$ sample. The samples at both redshifts have consistent intrinsic scatter within uncertainties. The measured intrinsic scatters do not change when we leave the slope and intercept (a and b) as free parameters or fix them to the values in Equations (2) and (3). Interestingly, this level of scatter is slightly smaller than the intrinsic scatter of ~ 0.1 dex found in the gas-phase MZR (e.g., Yabe et al. 2012; Guo et al. 2016) for galaxies with $M_* \gtrsim 10^{9.5} M_\odot$ but comparable to the intrinsic metallicity scatter of ~ 0.05–0.08 dex in the fundamental metallicity relation or in the gas-phase MZR when the star formation rate is taken into account (e.g., Mannucci et al. 2010; Yates et al. 2012; Lilly et al. 2013). This is expected because the stellar metallicity is less affected by the current star formation rate.

To further investigate the source of the intrinsic scatter, we plot the MZRs derived from both local and $z \sim 0.4$ galaxies, color-coded by their formation redshifts, in Figure 8(a). Ideally, this should yield the MZR of the star-forming galaxies at each formation redshift, which should not, or at most weakly, depend on when the galaxies were observed, i.e., the observed redshifts. Remarkably, the figure shows that, at each fixed mass, the galaxies that formed earlier (red data points) generally have lower metallicities than galaxies that formed later (blue data points) do. This is as we expect from the evolution of the gas-phase MZR, if stars approximately adopt the metallicity from their birth clouds.

At this point, we conjecture that the stellar MZR does not depend solely on galaxy mass, but also on the redshift at which that galaxy formed (or when the majority of stars formed when the SSP is not assumed). If this is true, the dependence on galaxy formation time should be able to explain the observed evolution with observed redshift. To see this effect better, we plot the deviation of the measured metallicities from the best linear fit of $z \sim 0$ galaxies (Equation (2), the upper dashed line) as a function of the age of the universe at the formation of their stellar populations in Figure 8(c). By subtracting the mass–metallicity function from the observed metallicities, the figure shows the effect of the age of universe at galaxy formation on stellar metallicity when the dependence on mass is removed. We can clearly see that galaxies that formed earlier (older
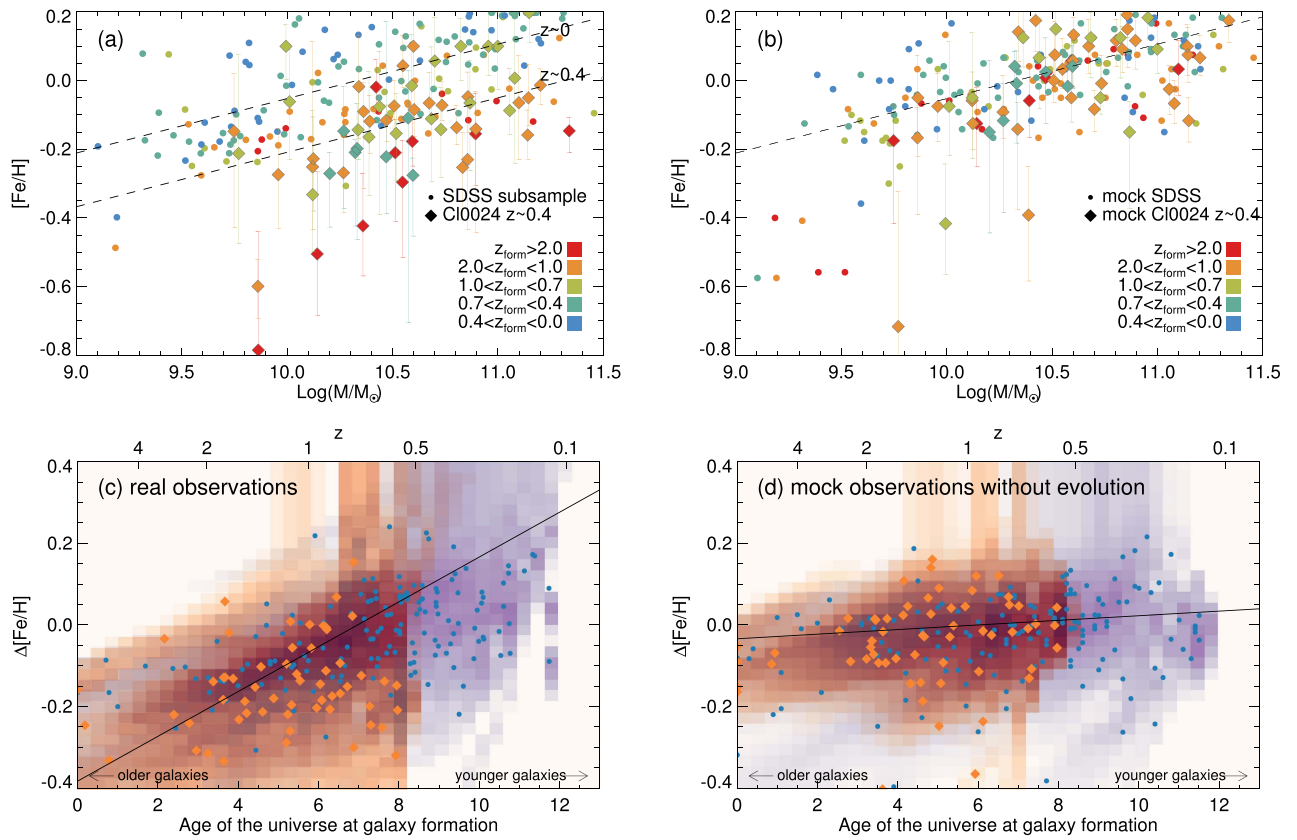
Figure 8. (a) Stellar MZRs of both local and $z \sim 0.4$ galaxies color-coded by their SSP-equivalent formation redshifts. The black dashed lines show the best-fit linear functions of the data at $z \sim 0$ and $z \sim 0.4$. (b) Mock observation when the "true" metallicity of each data point is assumed to be a linear function of its mass (the $z \sim 0$ black dashed line). The "observed" ages and metallicities of the mock galaxies were determined by the age–metallicity degeneracy. (c) and (d) The deviation of metallicities from the $z \sim 0$ best-fit linear function in observed data and mock data. The orange diamonds and the blue dots represent data points from $z \sim 0.4$ and $z \sim 0$, respectively. The underlying shaded purple (orange) colors are the co-added probability distribution of individual data points from $z \sim 0$ ($z \sim 0.4$) sample. The slope in the mock data (d) is caused by the age–metallicity degeneracy and is significantly smaller than the slope in the observation (c), which suggests an evolution of the MZR with formation redshift.

galaxies) are offset toward lower metallicities, while galaxies that formed later (younger galaxies) are offset toward higher metallicities.

However, the correlation with formation redshift (older galaxies have lower metallicities) is in the same direction as the age–metallicity degeneracy. Though some previous works have pointed out the anticorrelation between age and metallicity at a given mass, it was complicated by or was thought to be the result of the age–metallicity degeneracy (e.g., Jørgensen 1999; Gallazzi et al. 2005). To test whether the scatter in the MZR as a function of formation redshift is real or caused by the age–metallicity degeneracy, we created a set of mock observed ages and metallicities. We assume that the metallicities are solely determined by stellar mass according to the best-fit linear function found in the observed MZR at $z \sim 0$. If the trend with age is caused by the age–metallicity degeneracy, then we should obtain the same level of scatter in the MZR and its correlation with formation redshift after noise is added.

For each observed galaxy, we construct its twin mock galaxy. We took the measured mass of each observed galaxy and calculated its "true" metallicity from the linear function found in Section 5.1 (Equation (2)). We also took the measured age as the "true" age of the galaxy. The "true" observed spectrum was obtained from the FSPS according to its "true" age and metallicity, smoothed to the observed

velocity dispersion convolved with SDSS/DEIMOS instrumental dispersion. Gaussian noise was added at each pixel with the same flux uncertainty array as the observed spectrum. We calculated a χ^2 grid for each noised spectrum with noiseless SPS spectra of every possible combination of age and metallicity in the grid of 0.5 to 13 Gyr in age and -0.8 to 0.2 dex in metallicity. The grid spectra were smoothed to the same dispersion as the noised spectrum. The "observed" age and metallicity of each mock spectrum were then selected according to the probability of each cell in the χ^2 grid. The resulting mock MZR is shown in Figure 8(b).

We do not find the same level of separation of the mock MZR with galaxy formation redshift as in the observed MZR. The dots and diamonds of different colors in Figure 8(b) are visibly more mixed than those in Figure 8(a). We plot the deviation of the "measured" metallicities from the best linear fit to the $z \sim 0$ stellar MZR as a function of the age of the universe at their formation (Figure 8(d)). The slope in Figure 8(d) (mock observation) is purely caused by the age–metallicity degeneracy. If there were no such degeneracy, the deviation from the best-fit linear relation should scatter around Δ[Fe/H] = 0 at all ages. However, the degeneracy causes the data points to move slightly toward the lower left (more metal-poor and older) or upper right (more metal-rich and younger). Fitting a linear fit to the underlying probability distribution with

11

a maximum likelihood estimation,[7] we found a small positive slope of 0.005 ± 0.003 dex per Gyr in the relation in Figure 8(d).

The slope in the relation between the observed deviation from the best-fit line in the MZR and the formation redshift is steeper than that of the mock observation. The slope of the deviation in Figure 8(c) is 0.055 ± 0.006 dex per Gyr, significantly larger than the slope in Figure 8(d). When we fit linear functions to the $z \sim 0$ and the $z \sim 0.4$ samples individually, the best-fit slopes and intercepts are consistent within uncertainties. (The slopes are 0.058 ± 0.010 and 0.48 ± 0.07 dex per Gyr for the $z \sim 0$ and $z \sim 0.4$ samples, respectively.) This confirms that the evolution of the MZR with formation redshift is real. Galaxies with SSP-equivalent formation redshift at $z \sim 2$ have [Fe/H] an average of 0.4—0.5 dex lower than the metallicities of galaxies that just formed in the past 2 Gyr.

The evolution of the MZR with formation redshift suggests that the mass–metallicity relation is determined not only by galaxy masses but also by star formation histories. Moreover, this evolution of the MZR with formation redshift can consistently explain the evolution with observed redshift found in Section 5.1, which was 0.16 ± 0.03 dex from $z \sim 0.4$ to $z \sim 0$. The difference in the weighted mean formation time of the two populations is 2.7 ± 0.1 Gyr. If this difference is multiplied by the evolution with formation redshift of 0.055 ± 0.006 dex per formation Gyr, we would expect a difference of 0.15 ± 0.02 dex in [Fe/H] between the two populations, consistent with what we observed in the evolution with observed redshift. The gentler evolution of the MZR with observed redshift compared to its evolution with formation redshift is probably the result of shared histories prior to quenching that smear out the evolution.

Lastly, we note that the evolution of the MZR with formation redshift seen here is inconsistent with the observations in dwarf galaxies. Kirby et al. (2013) measured metallicities of dwarf galaxies based on measurements of individual stars. The authors established that a single MZR applies to all Local Group galaxies with $10^{3.5} < M_*/M_\odot < 10^9$ regardless of their star formation histories. Dwarf irregular galaxies with gas present today have the same MZR as dwarf spheroidal satellites with no gas present today. If this scenario held true in the more massive population, we would expect a tight MZR regardless of formation redshift.

5.3. On the Slopes of the MZR Relations

Another interesting feature that emerges from the mock MZR is the curve that bends toward lower metallicities at the low-mass end, even though we constructed the mock MZR from a linear relation (black dashed line in Figure 8(b)). The curve is similar to the observed MZR. This suggests that the age–metallicity degeneracy causes the tendency to

scatter toward lower metallicity in low-mass galaxies. An explanation can be found in the χ^2 contours in Figure 5. At [Fe/H] ~ -0.2, approximately where the change in slope occurs, the contours of equal probability can be asymmetric, biasing toward lower metallicities than the true values. In addition, the uncertainties at lower metallicities are generally larger than those at solar metallicities, causing larger scatter at the low-metallicity or low-mass end. This finding could suggest that the stellar MZR might in fact be a single power law with a similarly tight dispersion, at least over the observed mass range of $M_* \approx 10^{9.7} M_\odot$ to $10^{11.5} M_\odot$. To confirm this, individual spectra of low-mass galaxies with high S/N are required to secure low uncertainties in [Fe/H], which is beyond our current work but might be achieved by using gravitationally lensed galaxies, larger telescopes, or longer exposure times.

The shallow slopes of ~ 0.16 dex per logmass found in both local and $z \sim 0.4$ MZRs may give new insight into the strength of feedback in galaxies of the observed mass range. Strong star formation feedback generally results in a steep MZR (e.g., De Lucia & Borgani 2012; Lu et al. 2014). Lu et al. (2014) compared model predictions of both gas-phase and stellar MZRs from three independently developed semi-analytic models, namely the Croton model (Croton et al. 2006), the Somerville model (Somerville et al. 2012; Porter et al. 2014), and the Lu model (Lu et al. 2011). The authors found that the Croton model, which assumes a constant mass-loading factor, predicts the shallowest slope in the local stellar MZR, at ~ 0.17 dex per logmass over the mass range 10^8–$10^{11} M_\odot$. In contrast, the Lu model, in which the mass-loading factor is a strong function of halo circular velocity, predicts a very steep MZR, at ~ 0.6 dex per logmass over the same mass range. The slope of the MZR in our observations is consistent with the slope predicted by the Croton model.

Remarkably, our results agree with what Lu et al. (2014) found based on the gas-phase MZRs at $z \lesssim 1$. Among the three models considered, the Croton model also describes the observations of gas-phase MZRs the best. This might suggest that, over the observed range of mass and redshift, the amount of galaxy outflow is mainly a function of SFR and does not have a strong additional dependence on galaxy mass. This picture is closely related to the results from a simple closed-box model and from the FIRE hydrodynamical simulations, where stellar metallicity is a strong function of gas fraction within a galaxy halo (Ma et al. 2016).

However, the slope of the stellar MZR found in this work is not consistent with the slope of the stellar MZR found in dwarf galaxies. Dwarf galaxy satellites of the Milky Way exhibit a stellar MZR that is consistent with an unbroken power law (e.g., Grebel et al. 2003; Kirby et al. 2013). Based on measurements of metallicities of individual stars in Local Group dwarf galaxies with masses ranging from $M_* \sim 10^3$ to $10^9 M_\odot$, Kirby et al. (2013) measured the slope of the MZR to be 0.30 ± 0.02 dex in [Fe/H] per logmass. The slope of the dwarf galaxy MZR is significantly larger than the slope found in this work, which is based on individual integrated spectra of both local and $z \sim 0.4$ galaxies with stellar masses $\gtrsim 10^{9.5} M_\odot$. If both numbers are correct, there must be a change in slope around the transition mass.

The slopes of our observed stellar MZRs do not change significantly when we limit the sample to galaxies with lower masses, $M_* < 10^{10.5} M_\odot$. The slopes at this lower-mass end are 0.15 ± 0.04 and 0.33 ± 0.16 dex per log mass for $z \sim 0$ and

---

[7] Because the uncertainties of individual ages and metallicities are not Gaussian and correlate strongly with each other (see Figure 5), we cannot use linear-fit estimators that assume a Gaussian probability distribution. In this case, we use Markov chain Monte Carlo sampling to obtain the best linear fit that minimizes the negative likelihood, $-\sum \log(\oint P_i dl(a, b))$, where P_i is the probability distribution of individual measurements of Δ[Fe/H] and the age of the universe at galaxy formation. The integration is along the considered linear function with parameters a and b for the slope and intercept. The summation is over all data points. The best linear fit is the line that passes through the regions of highest probability. Note that the underlying distributions in Figures 8(c) and 8(d) are $\sum P_i$, which represent uncertainties of the data.

$z \sim 0.4$ samples respectively. Although the slope at the low-mass end of the $z \sim 0.4$ sample may seem to suggest a change in slope, it is still consistent with the slope we found for the whole sample in Section 5.1. Furthermore, based on what we found from the mock data (Figure 8(b)), the age–metallicity degeneracy can cause the metallicity at the lower-mass end to bias low and create a seemingly steeper slope.

Although Ma et al. (2016) fit a single power law to the MZRs of simulated galaxies over an entire 8 dex in mass, their FIRE hydrodynamical simulations indeed seem to show a change in slope around $M_* \sim 10^{8.5} \, M_\odot$ for the MZR of $z \sim$ 1.4–4 simulated galaxies. Dwarf galaxies exhibit a steeper slope than higher-mass galaxies (as shown in Ma et al.'s Figure 4). Unfortunately, the sample sizes of the simulated galaxies at $z = 0$ and $z = 0.8$ are also not large enough to exhibit a clear change in slope.

As discussed earlier, the slopes of the MZRs reflect the strength of mass-loading factors. If the change in slope is real, the mass at the change in slope can suggest a mass below which feedback starts to have additional dependence on other parameters. In fact, Lu et al. (2017) argued that in the low-mass regime, two different feedback mechanisms, i.e., ejective and preventive feedback, are needed to explain both the observed mass–metallicity relation and the stellar mass function.

6. Effect of Galaxy Environment

Observing galaxies in galaxy clusters has the benefit of allowing us to obtain multiple spectra in a few telescope pointings. However, we have so far ignored the fact that our samples at $z \sim 0.4$ are in a dense cluster environment and have treated them as if they were general early-type galaxies. Here we discuss the impact of environment on our results.

Not all properties of galaxies have been shown to correlate with environment. Group and cluster environments show a higher fraction of passive galaxies than the field environment (e.g., Gerke et al. 2007; Muzzin et al. 2012; Koyama et al. 2013). At $z = 0.4$, based on the Hyper Suprime-Cam survey, the red fraction in cluster environments (number of members >25) is about 40% higher in $M_* \sim 10^{9.5} M_\odot$ galaxies and about 20% higher in $M_* \sim 10^{10.75} M_\odot$ galaxies (Jian et al. 2018) than it is in field environments.

However, neither galaxy size nor the galaxy stellar mass function (GSMF) seems to depend on global environment predominantly. Morishita et al. (2017) found no significant differences in half-light radii between cluster and field systems in the *Hubble* Frontier Fields. The shapes of the GSMFs of the general field and clusters are also mostly indistinguishable. The main difference is among the galaxies with $M_* \gtrsim 10^{11} M_\odot$, where they are more enhanced in high-density environments (e.g., Calvi et al. 2013; Etherington et al. 2017; Malavasi et al. 2017). Therefore, the completeness of our sample relative to the cluster population should be more or less transferable to the completeness of the general population.

In terms of chemical composition, gas-phase metallicities in star-forming galaxies have been shown to have slight or no correlation with environment. More metal-rich galaxies, on average, reside in overdense regions (e.g., Cooper et al. 2008; Wu et al. 2017). Cooper et al. (2008) used strong emission lines to measure gas-phase metallicities of SDSS star-forming galaxies. The authors found that the offset in metallicity relative to the median gas-phase MZR as a function of galaxy overdensity is significant. However, the metallicity offset

between the least dense and densest environment considered in that study is less than 0.03 dex. For galaxies at higher redshift, Kacprzak et al. (2015) studied gas-phase metallicities of star-forming galaxies in a $z \sim 2$ galaxy cluster. They found no distinguishable difference between the gas-phase MZR of field and cluster galaxies to within 0.02 dex.

The effects of environment on age and metallicity are also likely to be minimal for early-type galaxies. In particular, Fitzpatrick & Graves (2015) found that SDSS quiescent early-type galaxies have slight variations in age with environment. Isolated galaxies have the youngest ages; brightest cluster galaxies are 0.02 dex older; and satellites are 0.04 dex older than the isolated galaxies. There is no significant variation in Fe enrichment. Furthermore, Harrison et al. (2011) measured ages and metallicities of early-type galaxies residing in four local galaxy clusters and their surroundings extending to 10 $R_{\rm vir}$. The ages and metallicities were measured via spectrophotometric indices. They found no dependence of age or metallicity on the locations of galaxies in the clusters, i.e., those in the clusters or in the clusters' outskirts. Harrison et al. concluded that galaxy mass plays a major role in determining stellar populations.

In conclusion, cluster environment can affect the chemical abundance in galaxies. However, the effect seems to be small and weaker in stellar metallicities than in gas-phase metallicities.

7. Summary

In this paper, we measured ages and metallicities of 62 individual quiescent galaxies in the $z \sim 0.4$ galaxy cluster Cl0024+17. The quiescent galaxies were selected based on the EW of the [O II] $\lambda 3727$ emission line and FUV – V color. The final sample spans the stellar mass range from $10^{9.7}$ to $10^{11.5} M_\odot$ with ~50% completeness for $M_* \gtrsim 10^{9.7} M_\odot$. We employed a full-spectrum fitting technique by adopting FSPS models (Conroy et al. 2009) and the assumption of single stellar populations (SSPs). We examined the accuracy of our fitting technique in several aspects, including varying the signal-to-noise ratios, testing the validity of the SSP assumption, and comparing with previous measurements in the literature. Our age and metallicity measurements have typical uncertainties of <0.15 dex. We also measured ages and metallicities from a subsample of local SDSS quiescent galaxies from Gallazzi et al. (2005) and constructed the MZR of the local quiescent galaxies based on our measurements. We used this local MZR to compare with the MZR of $z \sim 0.4$ galaxies. We find the following:

1. We considered three functions (logarithmic, quadratic, and linear) to fit the MZRs at both redshifts. We found that the linear function fits the observed MZRs the best.
2. We detect an evolution of the stellar MZR with observed redshift at $>5\sigma$. The evolution is 0.16 ± 0.03 dex from $z \sim 0.4$ to $z \sim 0$ or 0.037 ± 0.007 dex per Gyr. The observed evolution is greater but consistent within 1σ and 2σ uncertainties with predictions from the FIRE and the EAGLE hydrodynamical simulations. Our results may have emphasized the importance of recycling processes in hydrodynamical simulations.
3. The intrinsic scatter of the MZR is smaller than that of the gas-phase MZR but comparable to the scatter in the fundamental metallicity relation. The intrinsic scatter can be explained by an evolution of the MZR with galaxy age
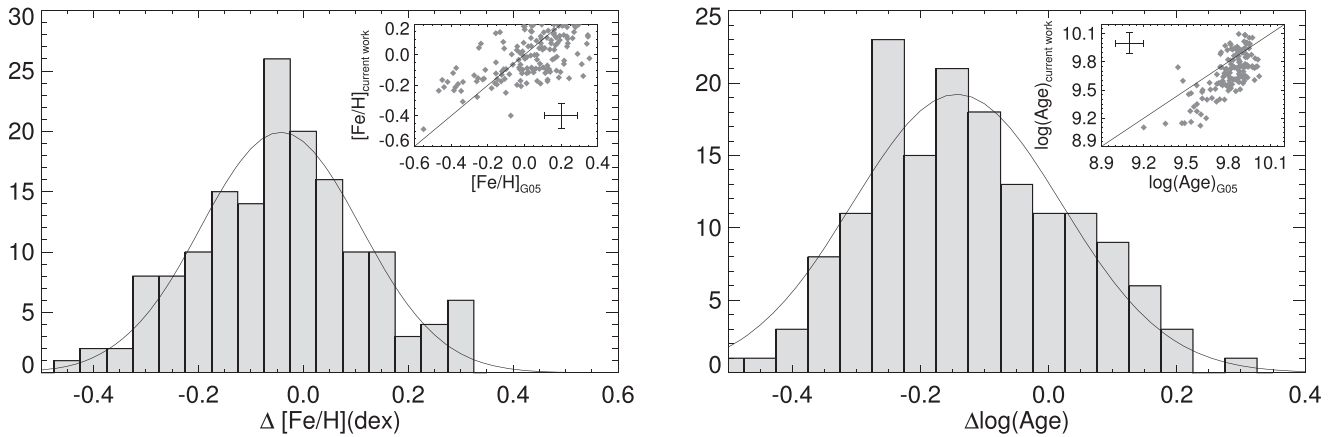
Figure 9. Comparison between the age and metallicities measurement in this work and Gallazzi et al. (2005) for the same subsample of SDSS quiescent galaxies.

or formation redshift. The MZR of galaxies that formed earlier is offset toward lower metallicity in the same manner as the evolution of the gas-phase MZR. The offsets are significant and not caused by the age–metallicity degeneracy. The evolution of the MZR with formation time is 0.055 ± 0.006 dex per Gyr, which is stronger than and can explain the evolution with observed redshift.

4. Based on the construction of a mock MZR from a linear relation, the age–metallicity degeneracy can cause the MZR at the low-metallicity end to be offset to lower metallicity, creating a downward curve sometimes seen in measurements of the MZR.

5. The slope of the MZR is ∼0.16 ± 0.03 dex per log mass. This slope is consistent with the slope predicted by a semi-analytic model (the Croton model, Croton et al. 2006) in Lu et al. (2014), which employs a mass-loading factor that is independent of galaxy mass. Our results suggest that galaxy feedback (in terms of mass-loading factor) might not have a strong additional dependence on galaxy mass over the observed mass and redshift range.

We will investigate the evolution of the MZR further in our future work using galaxies at higher observed redshifts. We will also measure α enhancements, as indicators of the star formation timescales of those population.

Appendix
Comparisons with Existing Measurements

We compare our measurements with literature measurements of the same galaxies (Gallazzi et al. 2005; Choi et al. 2014). We selected a subsample from 44,254 SDSS spectra from Gallazzi et al. (2005) that is comparable to our $z = 0.4$ sample based on emission line EWs and $U - B$ colors. Since the spectral ranges of the SDSS spectra extend beyond 8000 Å at the red ends but do not necessarily include the [O II] $\lambda3727$ emission lines at the blue ends, we instead used the criterion of rest-frame Hα

EW < 1 Å to define quiescent galaxies. The limit of Hα EW = 1 Å was chosen so that the SFR is comparable to the SFR when the [O II] EW is equal to 5 Å. These limits were based on the SFR calibrations from Kewley et al. (2004) assuming color $B - V = 2$, a typical color limit for quiescent galaxies (e.g., Schawinski et al. 2014) and no dust extinction. The color cut of $U - B > 1$ is to make sure that the contamination from star-forming galaxies is minimized (Mendez et al. 2011) in a similar manner to the color cut in our $z \sim 0.4$ sample. We selected all but at most eight random quiescent early-type galaxies from each bin of 0.1 dex in logarithmic mass spanning the stellar mass range from 10^9 to $10^{11.5} M_\odot$. This sums to a subsample of 155 quiescent galaxies. To be consistent with the observed $z \sim 0.4$ spectra, we limited the wavelength range of the SDSS spectra to 3700–5500 Å. We repeated the age and metallicity measurements in the same manner as in Section 3.

Our measured metallicities agree reasonably well with the values measured by Gallazzi et al. (2005). The results are shown in Figure 9. The differences in the [Fe/H] measurements follow a Gaussian distribution with a width of ∼0.15 dex, peaking at -0.04 dex. Though the width of the metallicity differences is comparable to the typical measurement uncertainty of 0.12 dex quoted by Gallazzi et al. (2005), there are some systematic differences in the measured [Fe/H]. The corner plot in Figure 9 shows that the metallicities we measured are slightly lower than those measured by Gallazzi et al. (2005) at high metallicities but the offsets reverse at lower metallicities. We argue that the main reasons for the discrepancy are the differences in the stellar libraries used in generating model spectra, which will be discussed below together with the uncertainties in age measurements.

The differences in age measurements show larger discrepancies than those in metallicities. The distribution of differences in age peaks at ∼-0.15 to -0.25 dex with a Gaussian width of 0.16 dex. The galaxies were generally younger than reported by Gallazzi et al. (2005) by about 0.2 dex. The discrepancy is likely dependent on SPS model. Gallazzi et al. (2005) computed five spectrophotometric indices from the BC03 stellar population synthesis, which is based on the STELIB spectral library (Le Borgne et al. 2003). The FSPS models used in the current work are based on the MILES spectral library (Sánchez-Blázquez et al. 2006). Although both spectral libraries are empirical, the STELIB library contains fewer stars—249 stellar spectra as compared to 945 spectra in
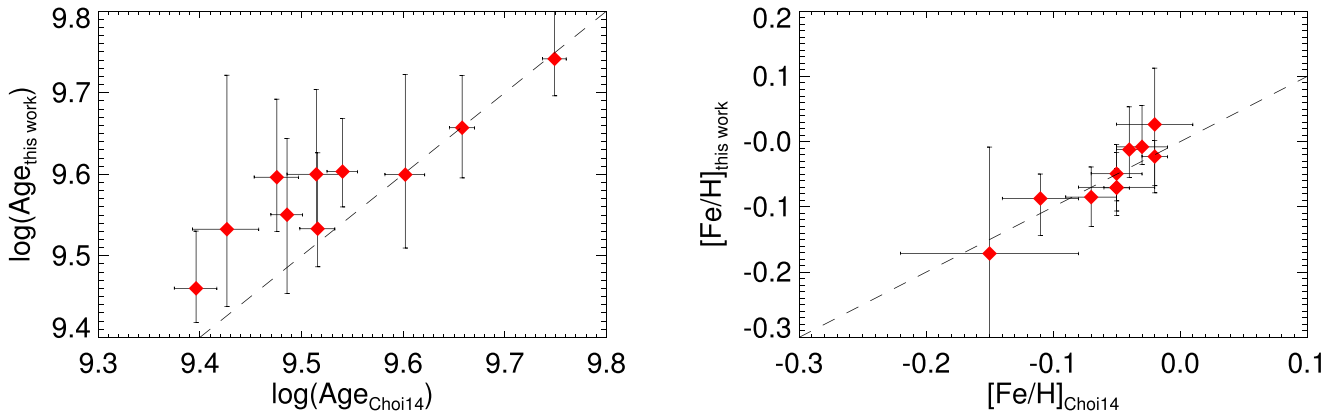
Figure 10. Comparison between the measurements of age and metallicities in this work and in Choi et al. (2014) for the same set of stacked spectra.

the MILES library. Very few stars in the STELIB library are at non-solar metallicities (Conroy & Gunn 2010).

Koleva et al. (2008) compared three spectral synthesis models of single stellar populations. In particular, the authors inverted the parameters from the SSP spectra produced by BC03 using a grid of models made with the Vazdekis/MILES (Vazdekis et al. 2010) and the Pegase–HR spectral synthesis model (Le Borgne et al. 2004). The two models are based on the MILES library and the ELODIE library (Prugniel et al. 2007), respectively. They found that at approximately solar metallicity, the ages retrieved by both Vazdekis/MILES and Pegase–HR are $\sim$0.2 dex younger than the input ages in BC03 when the input age is greater than $\sim$5 Gyr or $\log(\mathrm{Age/yr}) \gtrsim 9.7$ (see Koleva et al. 2008, Figure 2(a)). The trend reverses at $\sim$0.2–0.3 dex above the solar metallicity or 0.5 dex below the solar metallicity. Our finding is somewhat consistent with this result, in which we measured the ages to be $\sim$0.15–0.25 dex younger than measured by Gallazzi et al. (2005), as shown in Figure 9.

Lastly, we found good agreement between our measurements and those by Choi et al. (2014) in both ages and metallicities. We obtained a set of stacked spectra compiled by Choi et al. (2014).[8] These spectra were stacked from individual spectra of galaxies observed by the AGN and Galaxy Evolution Survey (AGES; Kochanek et al. 2012), in bins of redshifts from $z = 0.3$ to 0.7 and masses from $M_* = 10^{10.2}$ to $10^{11.3} M_\odot$. The spectral resolution is 6 Å, roughly double the resolution of our DEIMOS data. The wavelength coverage of the AGES spectra is 4000 to 5500 Å, which does not cover the full age-sensitive Balmer break region.

We found that our measurements of [Fe/H] are consistent with the values measured by Choi et al. (2014) (see Figure 10). For ages, all the values are consistent within 0.1 dex. However, there are greater discrepancies in the populations younger than 3.5 Gyr old, in the sense that we measured the ages to be slightly older than reported by Choi et al. (2014). The trend of this discrepancy is opposite to the discrepancy found when we compared our age measurements with Gallazzi et al. (2005).

Discrepancies between our work and Choi et al. (2014) likely arise from differing wavelength coverage and the spectral models used. Because the AGES spectra only cover from 4000 to 5000 Å, higher order Balmer lines, which contain age information, are not present in the spectra. Moreover, Choi et al. (2014) used the SPS model from Conroy & van Dokkum

(2012, CVD12). The main difference between the FSPS and the CVD12 model is that the latter allows abundances to be at non-solar ratios. Therefore, Choi et al. (2014) fit the spectra for abundances of individual elements including [Mg/Fe], [O/Fe], [C/Fe], [N/Fe], etc. Though it would be beneficial to measure individual elements in our observed spectra, using the solar metallicity models without adjusting individual elements can still provide a reasonable fit to the spectrum (Conroy & van Dokkum 2012). Besides, the CVD12 model is not yet readily applicable to our data at lower masses because the model is limited to a fairly small range around solar metallicity ([Fe/H] $\in (-0.4, 0.4)$) and age greater than 3 Gyr. In fact, the lower age limit at 3 Gyr in the CVD12 model might be responsible for the small age discrepancies found in Figure 10.

ORCID iDs

Nicha Leethochawalit ⬤ https://orcid.org/0000-0003-4570-3159

Evan N. Kirby ⬤ https://orcid.org/0000-0001-6196-5162

Richard S. Ellis ⬤ https://orcid.org/0000-0001-7782-7071

Tommaso Treu ⬤ https://orcid.org/0000-0002-8460-0390

References

Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212

Bian, F., Kewley, L. J., Dopita, M. A., & Blanc, G. A. 2017, ApJ, 834, 51

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Burnham, K., & Anderson, D. 2003, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (New York: Springer)

Calvi, R., Poggianti, B. M., Vulcani, B., & Fasano, G. 2013, MNRAS, 432, 3141

Chabrier, G. 2003, PASP, 115, 763

Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95

Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363

Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., & van Dokkum, P. G. 2012, ApJ, 760, 71

Cooper, M. C., Newman, J. A., Davis, M., Finkbeiner, D. P., & Gerke, B. F. 2012, spec2d: DEEP2 DEIMOS Spectral Pipeline, Astrophysics Source Code Library, ascl:1203.003

Cooper, M. C., Tremonti, C. A., Newman, J. A., & Zabludoff, A. I. 2008, MNRAS, 390, 245

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

De Lucia, G., & Borgani, S. 2012, MNRAS, 426, L61

De Rossi, M. E., Bower, R. G., Font, A. S., Schaye, J., & Theuns, T. 2017, MNRAS, 472, 3354

Dutton, A. A., Conroy, C., van den Bosch, F. C., et al. 2011a, MNRAS, 416, 322

---

[8] Kindly provided via private communication.

Dutton, A. A., van den Bosch, F. C., Faber, S. M., et al. 2011b, MNRAS, 410, 1660
Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813
Etherington, J., Thomas, D., Maraston, C., et al. 2017, MNRAS, 466, 228
Faber, S. M. 1973, ApJ, 179, 731
Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668
Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, Proc. SPIE, 4841, 1657
Fioc, M., & Rocca-Volmerange, B. 1999, arXiv:astro-ph/9912179
Fitzpatrick, P. J., & Graves, G. J. 2015, MNRAS, 447, 1383
Fumagalli, M., Franx, M., van Dokkum, P., et al. 2016, ApJ, 822, 1
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. 2006, MNRAS, 370, 1106
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Gerke, B. F., Newman, J. A., Faber, S. M., et al. 2007, MNRAS, 376, 1425
Grebel, E. K., Gallagher, J. S., III, & Harbeck, D. 2003, AJ, 125, 1926
Guo, Y., Koo, D. C., Lu, Y., et al. 2016, ApJ, 822, 103
Harrison, C. D., Colless, M., Kuntschner, H., et al. 2011, MNRAS, 413, 1036
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Jian, H.-Y., Lin, L., Oguri, M., et al. 2018, PASJ, 70, S23
Jørgensen, I. 1999, MNRAS, 306, 607
Kacprzak, G. G., Yuan, T., Nanayakkara, T., et al. 2015, ApJL, 802, L26
Kewley, L. J., Geller, M. J., & Jansen, R. A. 2004, AJ, 127, 2002
Kewley, L. J., Zahid, H. J., Geller, M. J., et al. 2015, ApJL, 812, L20
Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102
Kirby, E. N., Simon, J. D., & Cohen, J. G. 2015, ApJ, 810, 56
Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, ApJS, 200, 8
Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Koyama, Y., Smail, I., Kurk, J., et al. 2013, MNRAS, 434, 423
Kriek, M., van Dokkum, P. G., Franx, M., Illingworth, G. D., & Magee, D. K. 2009, ApJL, 705, L71
Kroupa, P. 2001, MNRAS, 322, 231
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
Le Borgne, J.-F., Bruzual, G., Pelló, R., et al. 2003, A&A, 402, 433
Lee, B., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 774, 47
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Lu, Y., Benson, A., Wetzel, A., et al. 2017, ApJ, 846, 66
Lu, Y., Mo, H. J., Weinberg, M. D., & Katz, N. 2011, MNRAS, 416, 1949
Lu, Y., Wechsler, R. H., Somerville, R. S., et al. 2014, ApJ, 795, 123
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Malavasi, N., Pozzetti, L., Cucciati, O., et al. 2017, MNRAS, 470, 1274
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Maraston, C. 2005, MNRAS, 362, 799
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Markwardt, C. 2012, MPFIT: Robust Non-Linear Least Squares Curve Fitting, Astrophysics Source Code Library, ascl:1208.019
McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, A&C, 15, 72
Mendez, A. J., Coil, A. L., Lotz, J., et al. 2011, ApJ, 736, 110
Moran, S. M., Ellis, R. S., Treu, T., et al. 2005, ApJ, 634, 977
Moran, S. M., Ellis, R. S., Treu, T., et al. 2006, ApJL, 641, L97
Moran, S. M., Loh, B. L., Ellis, R. S., et al. 2007a, ApJ, 665, 1067
Moran, S. M., Miller, N., Treu, T., Ellis, R. S., & Smith, G. P. 2007b, ApJ, 659, 1138
Morishita, T., Abramson, L. E., Treu, T., et al. 2017, ApJ, 835, 254
Moustakas, J., Zaritsky, D., Brown, M., et al. 2011, arXiv:1112.3300
Muzzin, A., Wilson, G., Yee, H. K. C., et al. 2012, ApJ, 746, 188
Nelson, D., Pillepich, A., Genel, S., et al. 2015, A&C, 13, 12
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006, MNRAS, 365, 46
Okamoto, T., Nagashima, M., Lacey, C. G., & Frenk, C. S. 2017, MNRAS, 464, 4866
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
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2008, MNRAS, 391, 1117
Peng, Y., Maiolino, R., & Cochrane, R. 2015, Natur, 521, 192
Porter, L. A., Somerville, R. S., Primack, J. R., et al. 2014, MNRAS, 445, 3092
Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, arXiv:astro-ph/0703658
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, ApJ, 799, 138
Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
Schiavon, R. P., Faber, S. M., Konidaris, N., et al. 2006, ApJL, 651, L93
Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, MNRAS, 423, 1992
Sommariva, V., Mannucci, F., Cresci, G., et al. 2012, A&A, 539, A136
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Sybilska, A., Lisker, T., Kuntschner, H., et al. 2017, MNRAS, 470, 815
Trager, S. C., & Somerville, R. S. 2009, MNRAS, 395, 608
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Ellis, R. S., Kneib, J.-P., et al. 2003, ApJ, 591, 53
Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, MNRAS, 449, 1177
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Walcher, C. J., Coelho, P., Gallazzi, A., & Charlot, S. 2009, MNRAS, 398, L44
Worthey, G. 1994, ApJS, 95, 107
Wu, P.-F., Zahid, H. J., Hwang, H. S., & Geller, M. J. 2017, MNRAS, 468, 1881
Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96
Yabe, K., Ohta, K., Iwamuro, F., et al. 2012, PASJ, 64, 60
Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215
Zahid, H. J., Dima, G. I., Kudritzki, R.-P., et al. 2014, ApJ, 791, 130