

A SPECTROSCOPIC SAMPLE OF MASSIVE, EVOLVED $z \sim 2$ GALAXIES: IMPLICATIONS FOR THE EVOLUTION OF THE MASS-SIZE RELATION¹

J.-K. KROGAGER^{2,3}, A. W. ZIRM², S. TOFT², A. MAN², G. BRAMMER^{3,4}

Submitted to ApJ September 6, 2013

ABSTRACT

We present deep, near-infrared *HST*/WFC3 grism spectroscopy and imaging for a sample of 16 galaxies at $z \approx 2$ in the COSMOS field selected by the presence of the 4000 Å break. This sample significantly increases the number of spectroscopically confirmed evolved galaxies at this redshift with accurate structural measurements. Moreover, this sample is the first representative sample of spectroscopically confirmed galaxies at $z \sim 2$. By combining the grism observations with photometry in 30 bands, we derive accurate constraints on their redshifts, stellar masses, ages, dust extinction and formation redshifts. We fit the rest-frame optical surface brightness profiles, and show that these are well described by compact, high- n Sérsic models. We show that the slope and scatter of the $z \sim 2$ mass-size relation of quiescent galaxies is consistent with the local relation, and confirm previous findings that the sizes for a given mass are smaller by a factor of two to three. Finally, we show that the observed evolution of the mass-size relation of quiescent galaxies between $z = 2$ and 0 can be explained by quenching of increasingly larger star-forming galaxies, at a rate dictated by the increase in the number density of quiescent galaxies with decreasing redshift. However, we find that the scatter in the mass-size relation should increase in the quenching-driven scenario in contrast to what is seen in the data. This suggests that merging is not needed to explain the evolution of the mean mass-size relation of massive galaxies, but may still be required to tighten its scatter, and explain the size growth of individual $z = 2$ galaxies quiescent galaxies.

Subject headings: galaxies: formation — galaxies: high-redshift — cosmology: observations

1. INTRODUCTION

Over the past decade, studies of the $z \sim 2$ Universe have been revolutionized by the availability of deep near-infrared (NIR) imaging surveys. One of the primary early results was the discovery of a population of optically-faint, massive galaxies which are missed in optical (rest-frame UV) surveys (Franx et al. 2003; Daddi et al. 2004; Wuyts et al. 2007). Large photometric surveys have since shown that at $z = 2$, roughly half of the most massive ($\log M/M_{\odot} > 11$) galaxies are dusty and star-forming, and half are old, quiescent systems (e.g. Franx et al. 2008; Toft et al. 2009; Williams et al. 2010; Brammer et al. 2011), a result that has been confirmed through low resolution spectroscopy of a small sample of the brightest examples (Kriek et al. 2008, 2009a,b).

Using high-resolution NIR imaging, it was shown that most of the quiescent galaxies at $z > 2$ have effective radii, r_e , a factor of 2 – 6 smaller than local elliptical galaxies with the same stellar masses (e.g., Daddi et al. 2005; Trujillo et al. 2006; Zirm et al. 2007; Toft et al. 2007; van Dokkum et al. 2008; Szomoru et al. 2010; Casata et al. 2011). Their inferred stellar mass densities (within r_e) therefore greatly exceed those of local galaxies at the same stellar mass. However, recent studies show that if one compares the stellar densities within

the inner 1 kpc the discrepancy is much less pronounced (Bezanson et al. 2009; Patel et al. 2013). The discovery that the inner regions of these massive galaxies correspond well with their local counterparts supports the so-called *inside-out* scenario, in which galaxies form at high redshift as compact galaxies presumably from a gas rich merger funneling the gas to the center and igniting a massive, compact star burst (e.g., Hopkins et al. 2006; Wuyts et al. 2010). These resulting compact stellar cores subsequently grow by adding mass to their outer regions. How this size growth is accomplished is the big question; A cascade of merger events with smaller systems, known as minor merging, is a plausible explanation as simulations have shown that it is possible to obtain the needed mass increase in the outer regions while leaving the central core mostly intact (Oser et al. 2012). However, observations of the merger rate of massive galaxies between $z = 2$ and 0 do not find as many mergers as required to account for the observed size evolution (Man et al. 2012; Newman et al. 2012).

Recently, studies of high-redshift galaxies have suggested that their structure may differ from that of local elliptical galaxies when quantified using a Sérsic profile. The high- z galaxies show lower Sérsic indices ($n \sim 2$ on average) than the local population of ellipticals ($n \sim 4$). This has motivated suggestions that the high- z population might be more disc-like and hence might contain a faint, extended stellar component which would be undetected in present observations due to cosmological surface-brightness dimming (van der Wel et al. 2011), but deeper and higher resolution imaging, along with image stacking, has confirmed that the massive, red galaxies indeed are compact, and has failed to detect any extended

¹ Based on observations carried out under programs #12177, 12328 with the Wide Field Camera 3 installed on the *Hubble Space Telescope*.

² Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen O

³ European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Vitacura, Santiago, Chile

⁴ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD21210, USA

stellar haloes around these compact cores (van Dokkum et al. 2008, 2010).

Now, with the advent of the next generation of NIR spectrographs on 8-m class telescopes, we can study the stellar populations via continuum detections and absorption line indices (Toft et al. 2012; Onodera et al. 2012; van de Sande et al. 2013, Zirm et al. in prep); The quiescent galaxies can be further sub-divided into post-starbursts (those that show strong Balmer absorption lines) and more evolved systems with metal absorption lines. However, even with state-of-the-art instrumentation, target samples are limited to the rare and bright examples.

Grism spectroscopy from space with *Hubble Space Telescope* (*HST*) allows us to obtain redshifts for fainter, less massive examples of $z \sim 2$ galaxies. While these data have poor spectral resolution, they do not suffer from the strong atmospheric emission lines, poor transmission and bright background that limit ground-based observations. A near-infrared spectroscopy survey, 3D-HST, has recently been carried out using the Wide Field Camera 3 (WFC3) onboard the *HST*. The survey provides imaging in the F140W-band and grism observations in the G141 grism. In total the survey will provide rest-frame optical spectra of ~ 7000 galaxies in the redshift range from $z = 1 - 3.5$. Moreover, the pointings cover approximately three quarters of the deep NIR survey, CANDELS (Grogin et al. 2011; Koekemoer et al. 2011). The combination of imaging and spectroscopic data from 3D-HST and CANDELS allows for powerful analysis of the redshift $1 < z < 3.5$ Universe. For more details about the 3D-HST survey, see Brammer et al. (2012).

We have searched the public 3D-HST data in the COSMOS field to identify a sample of galaxies with indications of a strong 4000 Å break, redshifted to the wavelength covered by the grism observations (corresponding to $1.86 < z < 2.75$). Our selection is motivated by the correlation between population age and the strength of the 4000 Å break, allowing us to select a population of evolved, massive galaxies. The presence of the break also serves as a direct indicator that enables us to derive accurate spectroscopic redshifts which in turn allow for stronger constraints on parameters of the stellar populations than what is possible with broad-band photometry alone. Until now, spectroscopic samples of quiescent, high-redshift galaxies with structural parameter data are sparse; van Dokkum et al. (2008) presented a sample of nine galaxies at $z \sim 2$, recently Gobat et al. (2013) presented five quiescent galaxies from a proto-cluster at $z = 2$, and at slightly lower redshifts Onodera et al. (2012) presented sample of 18 quiescent galaxies at $z \sim 1.6$. Samples of $z \sim 2$ quiescent galaxies with measured velocity dispersions and dynamical masses are even smaller; so far only four examples have been published (van Dokkum et al. 2009; Onodera et al. 2010; van de Sande et al. 2011; Toft et al. 2012). With our selection, we increase the sample size of $z \sim 2$ galaxies with spectroscopic redshifts significantly by adding 16 galaxies, and with these data, we are able to go deeper allowing us to get a more representative sample. By inferring sizes, redshifts, and stellar population parameters including age, star-formation rate, and mass, we are able to populate the mass-size relation using a mass-complete, quiescent sample of galaxies at $z \sim 2$. This provides

strong constraints on what drives the size evolution of the massive galaxies. We explore different physical explanations for the apparent size growth. Specifically, we create a simplistic model to investigate the effect of "dilution", i.e., addition of newly quenched, larger galaxies to the mass-size relation, a mechanism proposed by previous studies (Cassata et al. 2011; Trujillo et al. 2012; Poggianti et al. 2013) and recently investigated in detail out to redshift $z \sim 1$ by Carollo et al. (2013).

The paper is organized as follows: In section 2 we present the data used in our analysis, in section 3 we describe the selection of our sample before presenting the results of our analysis in section 4, in section 5 we investigate the mass-size relation and describe our model for size evolution driven by quenching, and finally we discuss the implications in section 6.

Throughout this paper, we assume a flat cosmology with $\Omega_\Lambda = 0.73$, $\Omega_m = 0.27$ and a Hubble constant of $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2. DATA

The analysis is based on public grism spectroscopy data from the 3D-HST survey from which we have used 25 pointings in the COSMOS field. We have combined the spectroscopic data with photometric data in 30 bands covering 0.15–24 μm from the latest K_s -selected catalog by Muzzin et al. (2013).

The 25 pointings in COSMOS are covered by imaging in the F140W filter and by NIR spectroscopy using the G141 grism providing wavelength coverage from 1.1 μm to 1.6 μm with a spectral resolution of $R \sim 300$ (for a point source) with a sampling of 46.5 Å per pixel. Since these are slitless spectroscopic data the effective resolution depends on the size and morphology of the dispersed source. Furthermore, we have used the two epochs of WFC3/F160W (H_{160}) images from the public CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) survey to constrain the structural parameters of our sample sources.

2.1. Data reduction

Each pointing was observed in a four-point dither pattern with half-pixel offsets in order to increase the resolution of the final image. Both the undispersed, direct images and the dispersed grism images were observed with this pattern for a total exposure of around 800 sec and 4700 sec for undispersed and dispersed, respectively.

The data sets were reduced using the publicly available pipeline *threedhst*⁵ (Brammer et al. 2012). The pipeline handles the combination and reduction of the dithered exposures, source identification using SExtractor, and extraction of the individual spectra. Since we are dealing with slitless spectroscopy some sources will have spatially overlapping spectra. This is handled in the pipeline and each extracted spectrum is provided with an estimate of the amount of contamination from nearby sources. For our analysis, we have subtracted the contaminating flux from the total extracted flux.

We have used the standard extraction parameters in the pipeline except for the final pixel scale used in the call to the IRAF-task MULTIDRIZZLE, where we chose

⁵ <http://code.google.com/p/threedhst>

$0''.09 \text{ px}^{-1}$ instead of $0''.06 \text{ px}^{-1}$. This was chosen to reduce the noise in the extracted spectra. For further details about the observations and data reduction see Brammer et al. (2012). We used a detection threshold of 4σ to identify sources in the F140W images.

After the initial reduction we encountered some issues with the background not being flat. We were not able to correct this gradient sufficiently to recover a completely flat background, which meant that some spectra were disregarded due to background issues. However, when we increased the pixel size from $0''.06 \text{ px}^{-1}$ to $0''.09 \text{ px}^{-1}$ the noise decreased and the background subtraction was performed more successfully. In the process of selecting our sample we removed two sources due to background-subtraction issues. In these cases there were discontinuities in the background, that we could not correct for.

3. SAMPLE SELECTION

In total we ended up with 10 239 extracted spectra. Many of these were very low signal-to-noise ratio (SNR) spectra ($\text{SNR} < 1.0$, averaged over the entire spectrum), corrupted extractions of objects near the CCD edge, or low-redshift objects. Our first selection criterion was therefore to quantify the significance of each spectrum using the method described by Pirzkal et al. (2004). They define the *net significance* of a spectrum, N , as the maximum value of the cumulative sum of the sorted signal-to-noise spectrum.

In order to cut down the sample size we invoked a few quality cuts. For a source to be accepted in our sample, its net significance had to be larger than 200. This corresponds roughly to a cut in terms of H -band magnitude, $H \lesssim 24.5$. We required that at least 80per cent of the pixels were well-defined, i.e., non-zero and non-negative. In some cases where the objects were located close to the edge of the CCD some light was dispersed out over the edges, and hence the spectral range was reduced in those cases. By only allowing spectra with more than 80per cent well-defined pixels, we ensured that our targets were fully covered in the wavelength range from 1.1–1.6 μm . Moreover, we computed the integrated amount of contaminating flux and compared this to the integrated total flux and removed sources for which the contamination was higher than 50per cent. We then matched our extracted spectra by coordinates to the photometry from the K_s -band selected catalog. Photometric redshifts for targets in the catalog were determined with the EAZY code (Brammer et al. 2008).

Our main selection criterion was to look for the 4000 \AA break in the spectra. We here followed the definition of $D(4000)$ from Bruzual A. (1983) using the broad 200 \AA bins to measure the blue and red continuum on either side of the break (see also Hamilton 1985). We used this broad definition due to the low resolution and low SNR. In order to have sufficient quality data in the blue part of the spectrum, where the targets are typically fainter, we removed candidates with a SNR per pixel averaged over the whole spectrum of less than 2. When searching for detections of the break, i.e., more than 1σ detections, we implemented a third control bin red-wards of the red continuum bin in order to sort out broadened emission lines and spurious jumps in the data.

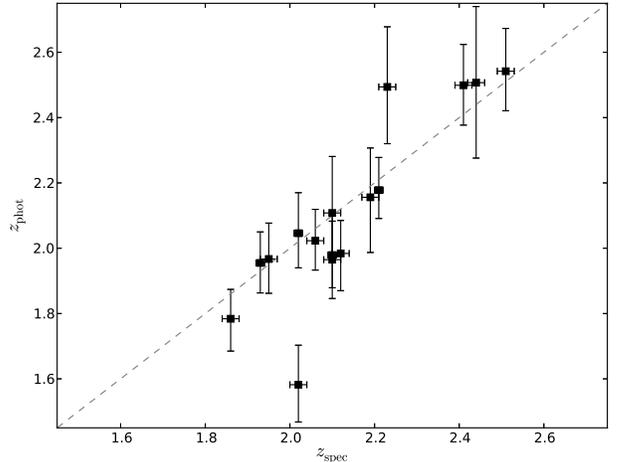


Figure 1. Spectroscopic redshift vs. photometric redshift for our sample of galaxies. The dashed line indicates the one-to-one relation. Photometric redshifts have been obtained using the code EAZY.

The galaxies in the sample with photometric redshifts $z_{\text{phot}} < 1.5$ were then thrown away since we were looking for galaxies with breaks within our spectral range, corresponding to redshifts in the range of $1.86 < z < 2.75$. We made the cut in redshift at 1.5 in order not to discard galaxies with underestimated photometric redshifts. The sources that showed a break in the spectrum were then compared to their broad band spectral energy distributions (SEDs) by scaling the spectra to the J -band, which is fully covered by $G141$. We scaled the spectra to correct for possible over-subtraction of contamination and to account for the loss of flux due to the limitations of the spectral extraction aperture. Candidates with a significant discrepancy (more than 1σ) between the H -band flux and the flux in the spectrum at the corresponding wavelengths were discarded. This discrepancy either stems from unaccounted for contamination or an uneven background subtraction. Finally, we checked how well the contamination (if any) had been subtracted. We discarded the most heavily contaminated spectra and the spectra where the contamination had been subtracted incorrectly leaving gaps and holes in the extracted spectra. This was done by visual inspection as not only the amount of contamination is important, but also the shape of the contaminating flux. In some cases the contaminating flux can enhance or even create a break in the spectrum, and this is difficult to quantify in a comparable way for all targets.

The properties of the final sample consisting of 16 galaxies are summarized in Table 1.

4. RESULTS

4.1. Spectral fitting

All galaxies in our sample were fitted by the FAST code (Kriek et al. 2009b). The code performs template fitting combining the photometric data with our grism spectra using exponentially declining star-formation histories, stellar population synthesis models by Bruzual & Charlot (2003) and a Chabrier (2003) initial mass function. Before fitting the spectra we binned them into 20 bins with bin-sizes of $\sim 250 \text{ \AA}$. We did this to avoid being affected by morphological broadening which arises

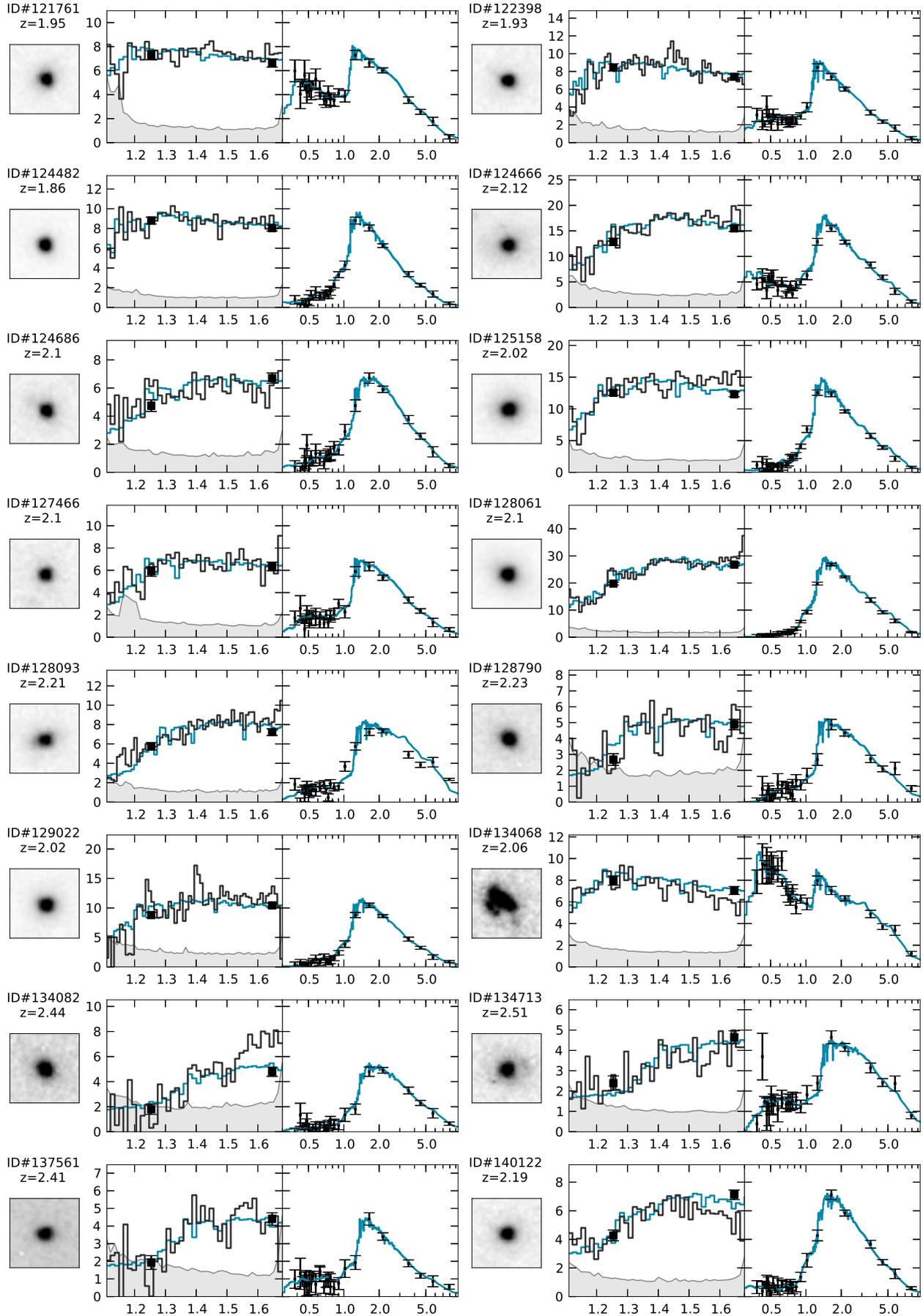


Figure 2. (left) 2.4×2.4 arcsec 2 H_{160} -band image, (middle) 1D extracted grism spectrum in black and error spectrum in grey, and (right) photometric SED. Both the middle and right panels show wavelength in units of μm vs. f_λ in units of $10^{-19} \text{ erg}^{-1} \text{ s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$. The blue line over-plotted in the two last panels indicate the best-fit model from FAST convolved and rebinned to match the grism spectra.

Table 1

Description of the full sample of galaxies. The photometric redshifts were calculated using the EAZY code.

ID	RA (deg)	DEC (deg)	z_{phot}	H ₁₆₀ (AB mag)	F140W (AB mag)	Grism ID
121761	150.1172352	2.2239839	1.97 ^{+0.11} _{-0.11}	21.96 ± 0.05	22.25 ± 0.01	ibhm42.243
122398	150.1538874	2.2324278	1.96 ^{+0.09} _{-0.09}	21.84 ± 0.05	22.37 ± 0.01	ibhm30.211
124482	150.0784925	2.2590418	1.78 ^{+0.09} _{-0.10}	21.75 ± 0.04	22.02 ± 0.01	ibhm33.118
124666	150.0657038	2.2610559	1.98 ^{+0.10} _{-0.11}	21.04 ± 0.04	22.10 ± 0.01	ibhm33.161
124686	150.0640338	2.2611897	2.11 ^{+0.17} _{-0.13}	21.95 ± 0.06	22.60 ± 0.02	ibhm33.160
125158	150.1047042	2.2671692	1.58 ^{+0.12} _{-0.11}	21.29 ± 0.03	21.79 ± 0.01	ibhm40.040
127466	150.1553232	2.2948989	1.97 ^{+0.14} _{-0.12}	22.01 ± 0.07	22.83 ± 0.02	ibhm51.200
128061	150.0738293	2.2979853	1.98 ^{+0.10} _{-0.10}	20.44 ± 0.02	21.30 ± 0.01	ibhm54.240
128093	150.0745036	2.3020139	2.18 ^{+0.10} _{-0.09}	21.87 ± 0.05	22.45 ± 0.02	ibhm54.256
128790	150.0995922	2.3118099	2.49 ^{+0.18} _{-0.17}	22.29 ± 0.07	22.82 ± 0.02	ibhm52.155
129022	150.0960034	2.3134756	2.05 ^{+0.12} _{-0.11}	21.47 ± 0.04	22.21 ± 0.01	ibhm52.157
134068	150.1635069	2.3724493	2.02 ^{+0.10} _{-0.09}	21.89 ± 0.06	22.27 ± 0.02	ibhm46.116
134082	150.1110445	2.3732350	2.51 ^{+0.23} _{-0.23}	22.31 ± 0.08	23.23 ± 0.03	ibhm53.075
134713	150.1871391	2.3801981	2.54 ^{+0.13} _{-0.12}	22.34 ± 0.07	23.06 ± 0.02	ibhm46.250
137561	150.0681893	2.4155696	2.50 ^{+0.12} _{-0.12}	22.40 ± 0.09	23.28 ± 0.02	ibhm35.010
140122	150.0796756	2.4496598	2.16 ^{+0.15} _{-0.17}	21.88 ± 0.05	22.57 ± 0.02	ibhm35.195

Table 2

Parameters from stellar population fitting to photometry and spectral data using the FAST code.

ID	z_{spec}	$\log(M_*)$ (M_\odot)	$\log(\text{Age})$ (yr)	Z	A_v (mag)	$\log(\text{sSFR})$ (yr^{-1})	$\log(\tau)$ (yr)	$\log(\Sigma_{50})$ $M_\odot \text{ kpc}^{-2}$	D(4000)
121761*	1.95	10.75 ^{+0.15} _{-0.11}	8.70 ^{+0.35} _{-0.33}	0.004 ^{+0.019} _{-0.000}	1.50 ^{+0.14} _{-0.46}	-9.02 ^{+0.11} _{-0.46}	8.40 ^{+0.63} _{-0.56}	9.78 ± 0.15	1.24 ± 0.15
122398	1.93	10.81 ^{+0.03} _{-0.15}	8.85 ^{+0.23} _{-0.41}	0.020	0.70 ^{+0.57} _{-0.52}	≤ -10.1	≤ 8.4	9.65 ± 0.12	1.76 ± 0.34
124482	1.86	10.74 ^{+0.05} _{-0.13}	9.05 ^{+0.21} _{-0.18}	0.020 ^{+0.017} _{-0.013}	0.10 ^{+0.38} _{-0.10}	≤ -11.6	≤ 8.3	10.17 ± 0.14	1.84 ± 0.30
124666*	2.12	11.17 ^{+0.13} _{-0.00}	9.00 ^{+0.26} _{-0.02}	0.050 ^{+0.000} _{-0.043}	0.00 ^{+0.60} _{-0.00}	-10.72 ^{+0.37} _{-0.00}	8.20 ^{+0.31} _{-0.04}	9.23 ± 0.09	1.31 ± 0.13
124686	2.10	10.99 ^{+0.10} _{-0.13}	9.25 ^{+0.15} _{-0.36}	0.004 ^{+0.006} _{-0.000}	0.60 ^{+0.58} _{-0.20}	≤ -10.8	≤ 8.6	9.72 ± 0.14	1.28 ± 0.15
125158	2.02	10.97 ^{+0.03} _{-0.02}	8.90 ^{+0.15} _{-0.06}	0.050 ^{+0.000} _{-0.030}	0.00 ^{+0.05} _{-0.00}	≤ -12.0	7.90 ^{+0.10} _{-0.64}	9.24 ± 0.06	1.76 ± 0.22
127466	2.10	10.92 ^{+0.03} _{-0.16}	8.95 ^{+0.30} _{-0.58}	0.004	1.00 ^{+0.56} _{-0.83}	≤ -10.0	≤ 8.5	10.02 ± 0.14	1.33 ± 0.14
128061	2.10	11.42 ^{+0.15} _{-0.01}	8.95 ^{+0.30} _{-0.01}	0.050 ^{+0.000} _{-0.030}	0.00 ^{+0.10} _{-0.00}	≤ -12.1	≤ 8.2	9.73 ± 0.09	1.66 ± 0.09
128093	2.21	11.19 ^{+0.04} _{-0.03}	8.75 ^{+0.11} _{-0.56}	0.050 ^{+0.000} _{-0.044}	1.20 ^{+1.02} _{-0.20}	≤ -10.0	≤ 8.1	10.07 ± 0.06	1.20 ± 0.10
128790	2.23	10.95 ^{+0.07} _{-0.12}	8.80 ^{+0.40} _{-0.33}	0.008	1.20 ^{+0.48} _{-0.60}	≤ -10.4	≤ 8.4	10.07 ± 0.12	1.73 ± 0.36
129022	2.02	11.04 ^{+0.07} _{-0.06}	9.00 ^{+0.27} _{-0.12}	0.020 ^{+0.016} _{-0.013}	0.40 ^{+0.35} _{-0.40}	≤ -11.3	≤ 8.4	9.96 ± 0.08	2.86 ± 0.78
134068*	2.06	10.90 ^{+0.08} _{-0.07}	8.95 ^{+0.21} _{-0.21}	0.050 ^{+0.000} _{-0.007}	1.20 ^{+0.14} _{-0.08}	-8.93 ^{+0.16} _{-0.06}	9.10 ^{+0.90} _{-0.32}	-	1.17 ± 0.12
134082	2.44	11.03 ^{+0.21} _{-0.09}	9.10 ^{+0.30} _{-0.22}	0.020	0.40 ^{+0.71} _{-0.40}	≤ -10.9	≤ 8.6	9.99 ± 0.18	1.73 ± 0.32
134713*	2.51	11.15 ^{+0.05} _{-0.22}	8.95 ^{+0.31} _{-0.53}	0.020	1.30 ^{+0.66} _{-0.64}	-9.70 ^{+0.62} _{-0.46}	8.40 ^{+0.38} _{-0.56}	9.18 ± 0.17	1.38 ± 0.16
137561	2.41	10.78 ^{+0.04} _{-0.13}	9.00 ^{+0.24} _{-0.19}	0.020	0.30 ^{+0.37} _{-0.30}	-10.72 ^{+0.09} _{-0.30}	8.20 ^{+0.28} _{-0.27}	10.33 ± 0.17	1.86 ± 0.35
140122	2.19	11.05 ^{+0.09} _{-0.19}	9.30 ^{+0.15} _{-0.30}	0.020	0.00 ^{+0.47} _{-0.00}	-11.61 ^{+0.58} _{-0.09}	8.40 ^{+0.21} _{-0.31}	9.83 ± 0.16	1.48 ± 0.18
Mean	2.1 ± 0.2	11.0 ± 0.2	9.0 ± 0.2	0.024 ± 0.016	0.6 ± 0.5	-	-	-	-

The ID of star-forming galaxies with constrained specific star formation rate (sSFR) are marked with *. Metallicities that are quoted without uncertainties were unconstrained in the fit, thus we only give the best-fit value.

due to the fact that we are using slitless spectroscopy on extended objects. The resulting effective resolution is $R \sim 50$.

We fitted the galaxies two times using FAST: The first time we allowed the redshifts to vary and kept the parameter grid coarse. We did this to get a description of the model spectrum for each target, which we then used to improve the best fitting redshift. By shifting the best fitting model in redshift with respect to the observed spectrum, we were able to obtain a spectro-photometric redshift with an error of $\sigma_z/(z+1) = 0.01$, determined by the break position or other visible features. Figure 1 shows the spectroscopic redshifts versus the photometric redshifts. The spectroscopic redshifts agree well with the photometric redshifts from EAZY, only one target is significantly off the one-to-one relation (ID #125158).

In the second fit we fixed the redshift to the spectrophotometric redshift determined above and refined the parameter grid in terms of age, which was constrained to be less than the age of the Universe at the given redshift, star formation time-scale τ , and dust extinction, A_V . All fits were performed with variable metallicity among four discrete values: $Z = 0.004, 0.008, 0.02, 0.05$. The spectroscopic redshift, parameters from the fits and the measurements of $D(4000)$ are summarized in Table 2. Figure 2 shows the individual spectra and SEDs along with their best fitting template. In Fig. 3, we show the distribution of stellar ages, masses, circularized effective radii, and formation redshifts. In each panel, we show an estimate of the probability density (indicated by the solid line), which we calculated from the observed distribution of parameters smoothed by a Gaussian kernel (for details, see Bashtannyk & Hyndman 2001). The density estimate helps to show the distribution in a way that is independent of binning.

Our sample is quite homogeneous in terms of age and mass with an average age and stellar mass of, respectively, 1 Gyr and $10^{11} M_\odot$. The size distribution shows hints of bi-modality, which is most likely caused by the few star-forming galaxies in our sample that are expected to, and indeed, have larger sizes at a given redshift (e.g., Newman et al. 2012). In Table 2, we give the best-fit metallicities from FAST. Although in many cases the metallicity is unconstrained in the range from $Z = 0.004 - 0.050$, we find that all galaxies except two have metallicities consistent with solar ($Z = 0.020$). This is also reflected in the average metallicity given in Table 2; $\langle Z \rangle = 0.024$. As mentioned, only two galaxies have constraints on Z that are inconsistent with solar; #124686 has sub-solar metallicity, $Z < 0.01$, and #134068 has super-solar metallicity, $Z > 0.04$.

4.2. Structural Fitting

We have obtained the structural parameters for the galaxies with GALFIT (Peng et al. 2002) using a single Sérsic component. This provides us with the parameters Sérsic n , half-light major axis a_e in pixels, and axis ratio b/a . The sizes quoted in table 3 are circularized ($r_e = a_e \sqrt{b/a}$) half-light radii in kpc, throughout the rest of the paper, circularized radii will be used. We used 8×8 arcsec² cutouts in the fit and adjacent objects were fitted simultaneously by either a Sérsic profile or as point sources. In two cases the fit was not able to con-

Table 3
Structural parameters from GALFIT.

ID	Sérsic n	r_e (kpc)	b/a
121761	5.8 ± 1.6	1.2 ± 0.2	0.81 ± 0.07
122398	6.3 ± 2.2	1.5 ± 0.2	0.67 ± 0.08
124482	2.8 ± 0.6	0.8 ± 0.2	0.79 ± 0.08
124666	1.00*	3.7 ± 0.4	0.72 ± 0.04
124686	7.2 ± 2.1	1.7 ± 0.2	0.85 ± 0.08
125158	5.0 ± 1.1	2.9 ± 0.2	0.79 ± 0.06
127466	7.8 ± 2.1	1.1 ± 0.2	0.76 ± 0.15
128061	5.2 ± 1.2	2.8 ± 0.2	0.85 ± 0.05
128093	4.00*	1.4 ± 0.1	0.55 ± 0.06
128790	2.0 ± 0.1	1.1 ± 0.1	0.86 ± 0.07
129022	3.5 ± 0.8	1.4 ± 0.1	0.83 ± 0.08
134068	–	–	–
134082	1.5 ± 0.2	1.3 ± 0.2	0.81 ± 0.05
134713	5.0 ± 1.1	3.9 ± 0.5	0.87 ± 0.05
137561	1.5 ± 0.4	0.7 ± 0.1	0.52 ± 0.12
140122	5.3 ± 1.6	1.6 ± 0.2	0.75 ± 0.09
Stack	6.3 ± 0.7	1.6 ± 0.1	0.8 ± 0.1

* Sources where n was fixed to get the fit to converge.

verge with the nearby objects being fit simultaneously (objects #124666 and #128061). We therefore used the SExtractor segmentation map to mask out the nearby objects. For each source we simulated the PSF at every position of the dither pattern using TINYTIM (Krist et al. 2011). We then combined these "raw" PSFs in the same way as the data images using the MULTIDRIZZLE algorithm. The PSF for some of the sources were not able to be simulated in this way because the object was located at the edge of the CCD in one or more exposures. For these sources we used a PSF from a target with similar CCD coordinates, i.e., within 100 pixels. In two cases (see the caption of table. 3) the fit did not converge when we allowed all parameters to vary. We therefore fixed n at either 1, 2, 3 or 4 and picked the best-fitting model out of these four.

In order to estimate the effect of the chosen PSF on the parameters we fitted all sources with all the available PSFs. This gave us a measure of the robustness of the fit. The parameters from the fits are summarized in Table 3. The quoted parameters and their errors are given as the best fit and standard deviation of all the different fits for each source. We have also computed the stellar mass density within the half-light radius from the fit given by:

$$\Sigma_{50} = \frac{0.5M_\star}{\pi r_e^2}. \quad (1)$$

The densities are listed in Table 2.

4.3. Stacking of Data

We now investigate the sample in more detail by stacking the spectra and the H_{160} images in order to look for weak features in the sample, e.g., faint outskirts of the galaxies missed in the individual Sérsic fits.

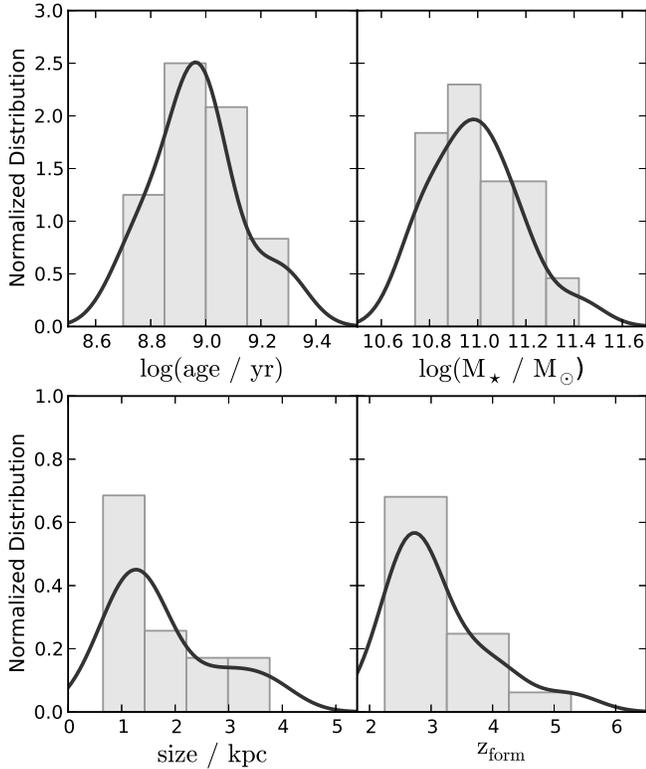


Figure 3. Histograms of the population parameters for our sample. The solid line shows the kernel density estimate of the given parameter. The top row shows the logarithm of stellar ages in units of years and the logarithm of stellar masses in units of M_\odot . The bottom row shows circularized effective radii in units of kpc and formation redshifts.

4.3.1. Spectral Stacking

For the spectral stacking, we have divided the sample into two sub-samples: star-forming (SFG) and quiescent (QG) galaxies. The SFGs are defined as having a constrained specific star formation rate (sSFR) from the fit larger than $\log(\text{sSFR} / \text{yr}^{-1}) > -10.7$. The quiescent galaxies constitute the rest of the sample. Two objects in our sample (#124666 and #137561) have sSFRs from the fit that are right on the border between SFG and QG. In these cases we have looked at their confidence intervals to decide in which category they most likely fall; #124666 is a SFG and #137561 is most likely a QG. We have excluded two objects from the stack; Object #128093 was excluded due to the poorer photometry, which impacts both the sSFR and the redshift precision, object #129022 was excluded due to the irregularities in its spectrum. We have stacked the spectra by interpolating the rest frame spectra onto a common wavelength grid, which corresponds to the rest frame pixel size (15 Å) at the mean redshift of the stack ($z = 2.1$). We then combined the spectra by median combination in order to decrease the influence of outliers in the stack. The two stacks along with the full stack of both subsamples are shown in Fig. 4. In the quiescent stack, we see tentative indications of absorption from the Balmer $H\gamma$ line, but no signs of $H\delta$ in absorption. Both lines are expected to be present in stellar populations where the last burst of star formation ended around 1 – 2 Gyr ago. The fact that we do not see $H\delta$ in absorption can be explained by

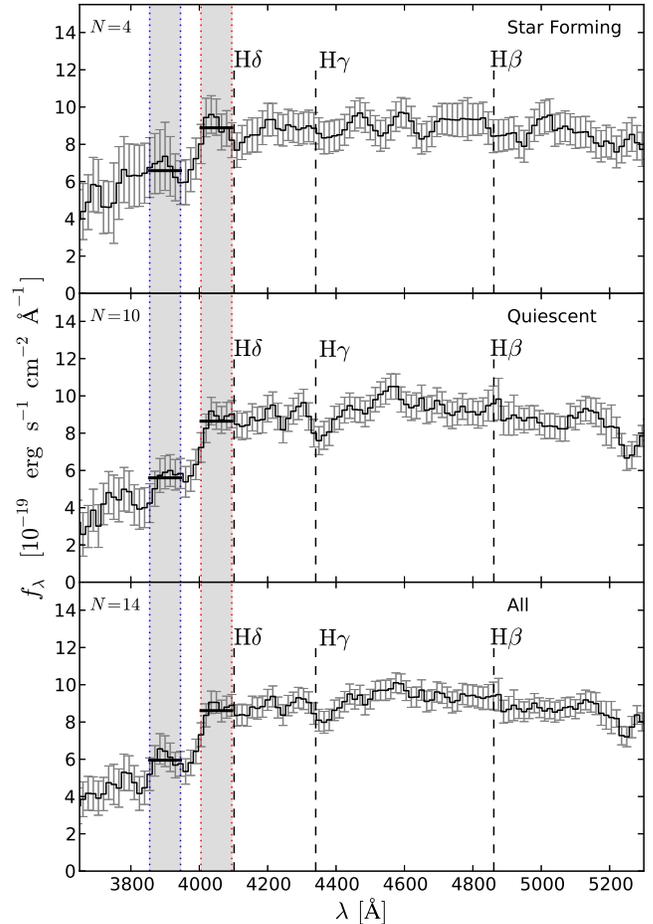


Figure 4. Stacks of the spectra from our sample divided into Star Forming (top), Quiescent (middle), and all (bottom) galaxies. See the text for definition of the sub-samples. Each figure shows the rest frame stacked spectrum. The position of the three Balmer lines, $H\beta$, $H\gamma$ and $H\delta$, are indicated by the dashed lines, and the continuum bins used for calculating $D(4000)$ are indicated by the shaded regions with the continuum level in each bin shown as the thick black horizontal line. In the upper left corner the number of galaxies in each stack is indicated.

the low resolution and the poor sampling of the spectra as this will blend together the $D(4000)$ feature and the $H\delta$ line. We clearly see a strong break at 4000 Å for the QGs, $D(4000) = 1.54 \pm 0.01$, indicative of an evolved stellar population. The stack confirms the homogeneity seen in the derived stellar ages: 0.6–2 Gyr. On the contrary, the star forming stack shows a shallower break, $D_n(4000) = 1.35 \pm 0.03$, and tentative signs of $H\delta$ in absorption, but no signs of $H\gamma$ nor $H\beta$. This may be caused by the mix of an evolved, underlying population with a younger, star-forming population. The individual SEDs of the four SFGs show signs of these mixed populations, see Fig. 2.

4.3.2. Stacking of H_{160} -Images

In order to characterize our sample in terms of structural parameters we also stacked the individual H_{160} -images of the quiescent galaxies. We masked out any nearby objects using SExtractor segmentation maps with a low detection threshold of 1.5σ to ensure that faint objects did not enter the stack. We then aligned all the

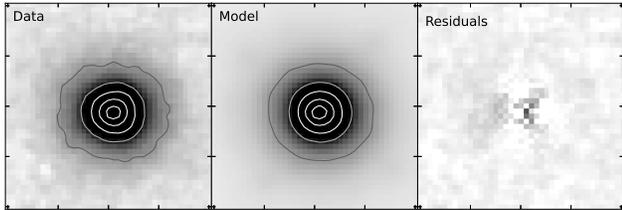


Figure 5. Stack of H_{160} images of the quiescent galaxies in the sample. Each panel shows a 2.5×2.5 arcsec² cutout. The panels show left to right; the stacked data, the model, and the residuals from GALFIT.

images and stacked them normalizing each source by the mean flux in the sample.

We then followed the same method as in our previous analysis, fitting the stack with all available PSFs and then estimating the errors on the parameters from all the individual fits. The parameters for the stack are indicated in Table 3. We found that the circularized effective radius for the stack was $r_e = 1.6 \pm 0.1$ kpc. We furthermore found a very high Sérsic index ($n = 6.3$) and no indications of faint outskirts in the stacked images in agreement with other studies (e.g., van Dokkum et al. 2008, 2010; Szomoru et al. 2012). The stacked image and the GALFIT model and residuals are shown in Fig. 5.

4.4. Mass Completeness

We assessed the completeness of our sample by comparing to the recent work of Ilbert et al. (2013) who investigated the mass function from UltraVISTA data. In Fig. 6, we show the data from stellar masses of our sample in the grey histogram where the error-bars represent the Poisson error of the number in each bin. The black line is the mass distribution of galaxies in the entire COSMOS field with specific star formation rates $\log(\text{sSFR}/\text{yr}^{-1}) < -10.0$, which correspond well with the sSFRs of our sample. The mass distribution from the entire COSMOS field has only been rescaled to match the area of COSMOS that is covered by 3D-HST ($\sim 2\%$). The solid blue line is the Schechter function from Ilbert et al. (2013) for quiescent galaxies scaled to match the COSMOS distribution, and the grey dashed line shows the mass distribution density estimate of our sample.

Due to the low number of galaxies in our sample it is difficult to assess the completeness in a quantitative manner. However, the agreement of both the observed mass distribution in COSMOS and the UltraVISTA mass function with our data for stellar masses above $10^{11} M_\odot$ is reassuring that our sample is reasonably representative of the massive, quiescent galaxies around $z \approx 2$.

5. THE MASS-SIZE RELATION

We have used our sample of spectroscopically confirmed galaxies at redshift $z \approx 2$ to investigate the mass-size relation at high redshift. We parameterize the relation for quiescent galaxies following Newman et al. (2012) and others:

$$r_e = \gamma \left(\frac{M_\star}{10^{11} M_\odot} \right)^\beta = \gamma M_{11}^\beta. \quad (2)$$

We fit the relation to the data using χ^2 minimization without taking the errors into account since the scatter dominates the relation. In the minimization we vary

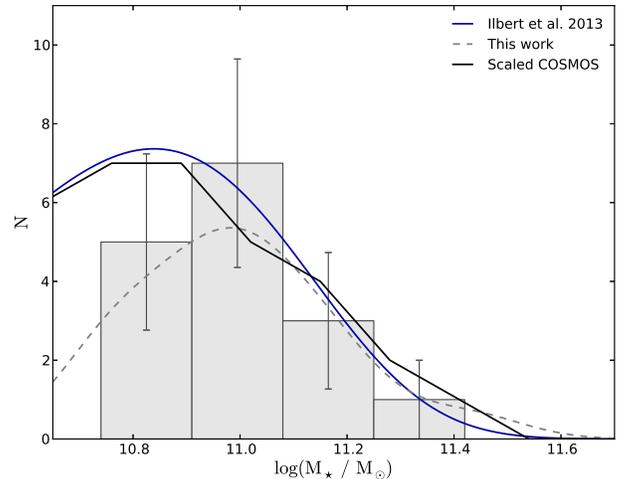


Figure 6. Distribution of stellar mass in our sample represented by the histograms and the kernel density estimate shown in the grey, dashed line. We compare to the stellar mass function from Ilbert et al. (2013) (blue line) and to the distribution of quiescent galaxies ($\text{sSFR} < 0.1 \text{ Gyr}^{-1}$) from the COSMOS field scaled to the area that is covered by the 3D-HST survey ($\sim 2\%$).

the two fit parameters; slope (β) and mass-normalized size (γ). In Sect. 4.4, we have estimated that our sample is complete only for masses larger than $\sim 10^{11} M_\odot$, and hence we only fit the relation for the galaxies in our sample that fulfill $\log(M_\star) > 10.9$. Furthermore, as this relation is only defined for quiescent galaxies we disregard the two star-forming galaxies above the mass-limit. The best-fitting values to our quiescent galaxies are: $\beta = 0.51 \pm 0.32$ and $\log(\gamma/\text{kpc}) = 0.17 \pm 0.05$ with a scatter of $\sigma_{\log r_e} = 0.12$ dex. The slope of our best fit is poorly constrained due to the low number of data points; however, the best-fitting value is in good agreement with the local slope found by various authors, e.g., Shen et al. (2003) find $\beta_{z=0} = 0.56$, see also Guo et al. (2009) and Newman et al. (2012).

In Fig. 7, we show our sample of quiescent galaxies above the mass-limit of $\log(M/M_\odot) > 10.9$ in red squares. The blue stars show the two star-forming galaxies above the mass-limit and the grey points with error bars show the sample below the mass-limit (dashed vertical line). We compare our data to local SDSS data with Sérsic index $n > 2.5$ (light grey, underlying distribution) and local early type galaxies with kinematical data (slow and fast rotators in dark grey circles and black triangles, respectively) from ATLAS^{3D} (Cappellari et al. 2011). In order to compare our high redshift sample to the that of the ATLAS^{3D}-team we fit the power-law relation given above to their data using the same mass-limit as for our data. From the best fit to the combined sample of fast and slow rotators, we find the following slope of $\beta_0 = 0.56 \pm 0.04$, a mass-normalized size of $\log(\gamma_0/\text{kpc}) = 0.61 \pm 0.01$, and a scatter of $\sigma_0 = 0.12$. For this analysis, we have used the tabulated values from Cappellari et al. (2013). Specifically, we note that we used the $\log(r_{1/2})$ to infer the sizes.

It is clearly visible that the quiescent galaxies from this work are smaller than local quiescent galaxies for a given mass. Moreover, the figure shows that the various samples of local galaxies infer slightly different normalizations of the relation. The relation derived from the

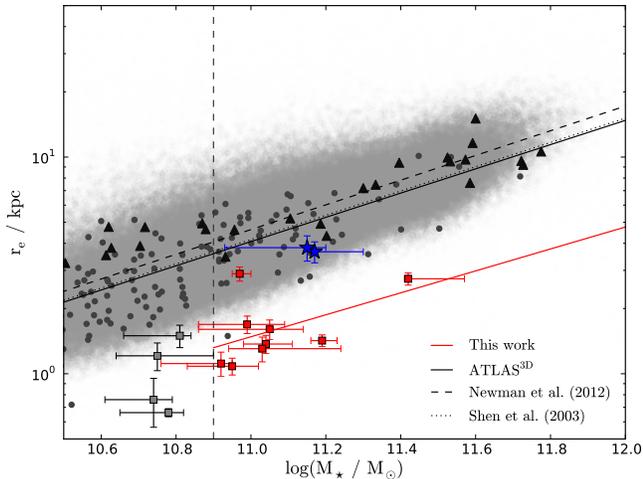


Figure 7. Mass-size relation using circularized effective radii. Red squares and blue stars indicate quiescent and star-forming galaxies, respectively, in our sample with $\log(M_*/M_\odot) > 10.9$. Grey squares with error-bars show our sample below the mass-limit. The solid red line is the best fit to our quiescent sample with a 1σ scatter of $\sigma_{\log r} = 0.12$. Data for local galaxies from SDSS are shown by the light grey, underlying points, while the dark grey circles and black, large triangles show, respectively, fast and slow rotators from the ATLAS^{3D} survey of local early type galaxies (masses and sizes are extracted from Cappellari et al. 2013). The dotted and dashed, black lines indicate the local mass-size relation defined for early type galaxies by Shen et al. (2003) and Newman et al. (2012), respectively. The solid, black line is the best fit to the ATLAS^{3D} points with a 1σ scatter of $\sigma_{\log r} = 0.12$.

ATLAS^{3D} data is in perfect agreement with the relation derived by Shen et al. (2003). Only the scatter is slightly smaller compared to the Shen et al. study, most probably due to the smaller sample size. We note that the scatter in our sample is most certainly underestimated due to the small sample size. However, by testing this with a simple calculation where we evaluate the scatter of a known log-normal distribution as function of sample size, we find that the scatter is at most underestimated by 0.04 dex. Even with a correction of 0.04 dex the scatter in our sample is still consistent with the locally observed scatter of 0.16 dex from Shen et al. (2003).

5.0.1. Passive Evolution

Next we investigate the evolution of our sample of $z \sim 2$ quiescent galaxies to lower redshifts, by comparing them to a spectroscopic sample of the brightest, most massive quiescent galaxies at $\langle z \rangle = 1.6$ (Onodera et al. 2012). In Fig. 8, we show sizes and masses as functions of stellar age for quiescent (black squares) and the star-forming (grey stars) $z \sim 2$ galaxies, and $z \sim 1.6$ quiescent galaxies (red circles). If the average size of quenched galaxies increases with time due to dilution, a correlation between the ages and sizes of quenched galaxies would be expected, due to the addition of larger, newly quenched (and therefore younger) galaxies to the quenched population. We do not find evidence for such correlation neither within our sample nor when comparing the two samples. However, this may simply be because of the relatively small dynamical range in ages and sizes probed by the samples and due to the large uncertainties on the ages.

The $z \sim 1.6$ galaxies are older than the $z \sim 2$ galaxies by roughly the cosmic time passed between the two red-

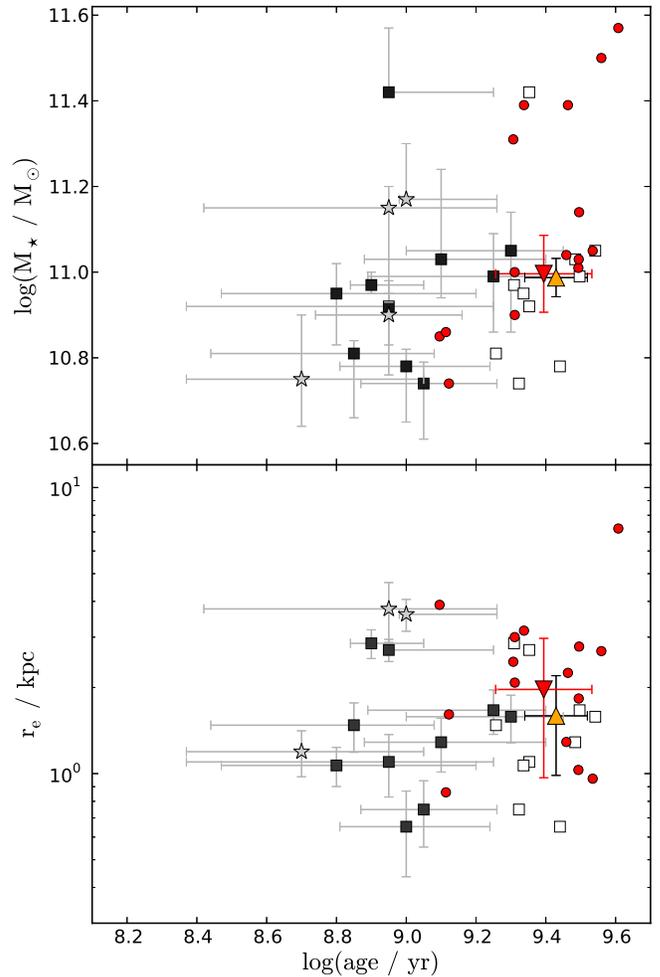


Figure 8. Stellar mass versus age (*top*) and circularized effective radii versus age (*bottom*) of our quiescent sample galaxies in black squares with error-bars. The grey stars indicate SFGs in the sample. The open squares indicate the quiescent sample passively evolved to a redshift of $z = 1.4$. The red points show the sample by Onodera et al. (2012) of galaxies with spectroscopic redshifts. For comparison, their sample has been passively evolved to $z = 1.4$ as well. The average of the samples at comparable masses ($10.8 < \log(M_*/M_\odot) < 11.2$) for this work and the Onodera sample are shown by the big, orange and red triangles, respectively.

shifts, consistent with simple passive evolution of their stellar populations. To illustrate this, the open squares in Fig. 8 are the $z \sim 2$ quiescent galaxies (black points), passively aged by the time passed between their observed redshifts and $z = 1.4$. For comparison the $\langle z \rangle = 1.6$ galaxies (red points) have also been passively evolved to the same redshift (the lowest redshift in the Onodera et al. (2012) sample).

In order to compare the two samples, we calculate the mean of the samples at similar masses ($10.8 < \log(M/M_\odot) < 11.2$) indicated by the big, orange triangle pointing up (passively evolved $z \sim 2$ galaxies) and red triangle pointing down (passively evolved $z \sim 1.6$ galaxies). As can be seen in the figure, the two samples are consistent when compared at similar masses, consistent with simple passive evolution between the two redshifts.

5.0.2. Size Evolution

We now take a closer look at the offset towards smaller sizes visible in Fig.7 between our sample at high redshift and the local sample. This offset has been studied in great detail (e.g., Daddi et al. 2005; Trujillo et al. 2006; Toft et al. 2007; Zirm et al. 2007; van Dokkum et al. 2008; Damjanov et al. 2011; Newman et al. 2012) and various explanations have been put forward to explain the required evolution in sizes, e.g., merging or feedback from quasars (Fan, Lapi, De Zotti, & Danese 2008). We here investigate a simple scenario in which the individual galaxies themselves do not need to increase significantly in size, but rather that the average of the population as a whole increases (e.g., Cassata et al. 2011; Trujillo et al. 2012; Carollo et al. 2013; Poggianti et al. 2013). We use our measurements of sizes and scatter at high redshift in combinations with those from Newman et al. (2012) to motivate the initial values for the size evolution.

Newman et al. (2012) study the size evolution of massive galaxies both star-forming and quiescent and find that the star-forming galaxies on average are a factor of 2 larger than the quiescent population at all times above redshift $z > 0.5$. This is in good agreement with the two star-forming galaxies in our sample (above our mass completeness limit) that are a factor of 2 larger than our quiescent sample (see Fig. 7). The evolution of the mean size of quiescent galaxies might then simply be driven by the addition of larger, newly quenched galaxies at lower redshifts to the already quenched population. Carollo et al. (2013) recently showed that the evolution of the sizes of passively evolving galaxies at $z < 1$ is driven by this "dilution" of the compact population. In order to test this picture and evaluate the effect on the scatter in sizes, we have taken the measured sizes normalized to a stellar mass of $10^{11} M_{\odot}$ from Newman et al. (2012) at redshift $2.0 < z < 2.5$ and generated an initial population of quiescent (QG) and star-forming (SFG) galaxies taking into account the observed number densities at that redshift for galaxies with comparable masses from Brammer et al. (2011). We have shifted the data from Newman et al. from a Salpeter IMF to the assumed Chabrier IMF in this work. The distribution of sizes for the populations are drawn from a log-normal distribution with an average size initially dictated by the observations for QGs while for SFGs we simply use the fact that star-forming galaxies on average are twice as big. Both distributions are assumed to have a scatter of 0.16 dex initially, motivated by the findings in this work.

We then simply assume that the SFGs at the given redshift will be quenched after a fixed time, t_{quench} , and add them to the already existing population of quiescent galaxies. For each time-step, we generate a new population of SFGs with a mean size that is twice as big as the mean size of the quiescent galaxies already in place, and after another t_{quench} these will be added to the quiescent population. The generated number in the SFG population varies according to the observed number density of SFGs. We have assumed that the scatter of the SFG population is constant with time and that no galaxy-galaxy interactions occur, i.e., no new massive galaxies form by merging of lower-mass galaxies. Furthermore, we assume that galaxies maintain their sizes after they have been quenched and that no further star formation

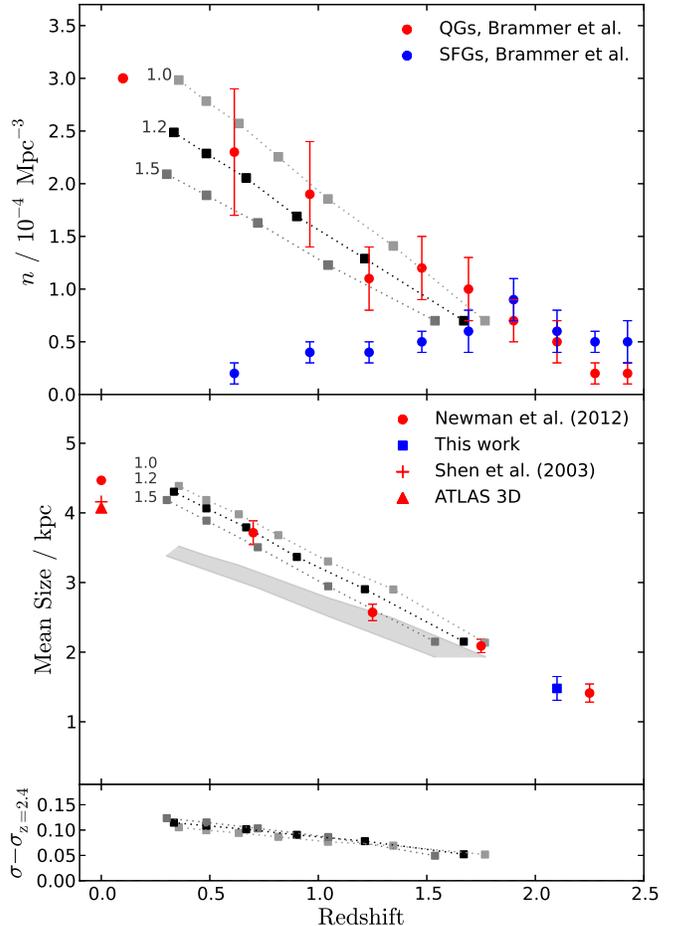


Figure 9. (*Top*) Number density evolution with redshift. The red and blue points show the observed number densities for quiescent and star-forming galaxies, respectively, with masses $\log(M/M_{\odot}) > 11$ from Brammer et al. (2011). The black and grey, connected points indicate our modeled evolution with varying quenching time indicated in Gyr by the small number at each line. (*Middle*) Average size of the quiescent galaxy population at fixed mass of $10^{11} M_{\odot}$ as function of redshift. The black and grey points are the same as in the top plot. The red circles show the observations from Newman et al. (2012), the cross and triangle are the local data from Shen et al. (2003) and the ATLAS^{3D} sample, and the blue square shows the size of our sample. The grey filled area indicates the evolution including fading of star-forming galaxies after quenching assuming the same quenching times (see text for details). (*Bottom*) Modelled scatter as a function of redshift relative to the initial scatter of 0.16 dex at redshift $z = 2.4$, the first redshift-bin from Brammer et al. (2011).

occurs once the galaxies have been quenched. We run this model three times for various quenching time-scales, t_{quench} : 1.0, 1.2, and 1.5 Gyr.

The results of this simple model are shown in Fig. 9. The top panel shows the evolution in number density. The red and blue points are data from Brammer et al. (2011) for quiescent and star-forming galaxies, respectively. The black and grey points show the modeled evolution in the number density assuming different quenching times indicated in Gyr by the number at each of the tracks. The middle panel shows the evolution in average size of the sample of quiescent galaxies. Data from Newman et al. (2012) is shown in red circles, our sample is indicated by the blue square, and the local size mea-

measurements from Shen et al. (2003) and ATLAS^{3D} are shown by the red plus and triangle, respectively. Again, the black and grey points indicate the modeled evolution at different quenching times. We performed a run where we included an estimated decrease in effective radius after the star-forming discs have faded. This is shown in Fig. 9 as the grey shaded area where the upper and lower boundaries correspond to, respectively, $t_{\text{quench}} = 1.0$ and 1.5 Gyr. We assumed that half of the star-forming population will be disc-dominated and that these will decrease their effective radii up to 30 per cent. The bottom panel shows the evolution of the scatter, σ , of the distribution of quiescent galaxies relative to the initial value at $z = 2.45$.

From these assumptions, we are able to reproduce the observed increase in number density and size of quiescent galaxies. However, the modeled scatter increases in contrast with the constant scatter observed in this work.

6. DISCUSSION

The evolution of galaxies in the mass–size plane is undoubtedly influenced by merging, star-formation and its cessation. As we increase the samples of well-studied, spectroscopically confirmed galaxies over a range of redshifts we can forge new diagnostic tools to address the weight with which each of these processes influences the evolution of galaxies.

In Sect. 5, we investigated the relation between stellar mass and half-light radius by parametrizing the relationship with a power-law. From the best fit to our quiescent grism sample we found the slope, $\beta = 0.51$, and the scatter, $\sigma_{\log r_e} = 0.12$ dex, consistent with their $z = 0$ values. From the ATLAS^{3D} data and from a large SDSS sample from the work of Shen et al. (2003) and Newman et al. (2012), a local slope and scatter of $\beta_0 = 0.56$ and $\sigma_0 = 0.12 - 0.16$ dex was inferred. One complication in comparing samples of galaxies at different redshifts, and from different samples, lies in the fact that the distinction between star-forming and quiescent galaxies becomes less clear at higher redshifts. Various studies use different criteria to define quiescence, e.g., a cut in sSFR or rest-frame colors, which makes any comparison between different datasets non-trivial. Even at low redshift, the classification of early type galaxies is performed in different ways. It is thus reassuring to see that we get very consistent results from the SDSS data and the ATLAS^{3D} team.

The importance of a clean separation and definition of star-forming and quiescent galaxies becomes clear when we look at the scatter as a tool to unravel the evolution in the mass–size relation, since the scatter is highly sensitive to outliers. Newman et al. (2012) find a scatter of $\sigma_{\log r_e} = 0.26$ dex for galaxies at redshifts $2.0 < z < 2.5$, much higher than what we find in our data. The large scatter observed in the Newman et al. sample may be due, at least partly, to the uncertainty in photometric redshifts and contamination from star-forming galaxies.

6.1. Mechanisms for size growth

In large photometric samples it has also been shown that the slope of the mass–size relation evolves very little from $z \sim 2$ to $z \sim 0.2 - 0.4$ despite there being strong redshift evolution of the galaxy distribution in the mass–size

plane (primarily a shift to larger sizes, see Newman et al. 2012, and McLure et al. (2013)). While the unchanging slope may be theoretically plausible as the slope may reflect initial formation rather than subsequent evolution (Ciotti, Lanzoni, & Volonteri 2007), the lack of evolution in the scatter observed in this work is puzzling. The scatter about the mean mass–size relation should evolve with redshift according to the underlying physical driver for the evolution in the mass–size plane, i.e., merging, quenching, or further star formation.

Merging will typically move galaxies to higher masses and larger radii, with the direction and amplitude of the change in the mass–size plane determined by the mass ratio, orbital parameters and gas content of the merger (Naab, Johansson, & Ostriker 2009). In gas-rich mergers, the remnant may become more compact due to the gas falling to the center, which leads to strong star formation activity. Star formation at later times (e.g., merger induced) will increase the mass, alter the size depending on the location of the star formation, and will decrease the mean age of the sample of *quiescent* galaxies at subsequent redshifts.

Quenching of star-forming galaxies will conserve mass while the individual galaxy sizes may even slightly decrease (as low-surface brightness star-forming regions fade) but is operating on a separate galaxy population that has intrinsically larger sizes than most of the quiescent galaxies already in place (Khochfar & Silk 2006). The addition of these quenched galaxies will then drive the evolution of the mean size of the whole population without changing the individual galaxies that have already been quenched. However, it is still not entirely clear what happens to star-forming galaxies after they stop forming stars in terms of morphology and size; star-forming galaxies show a variety of morphologies but the quiescent population is more dominated by spheroidal morphologies. Carollo et al. (2013) suggest that the star-forming galaxies shrink by ~ 30 per cent after they get quenched presumably by fading of the star-forming disc. We have assessed the effect of this fading of the star-forming population on the size evolution of the quiescent population in our toy model, see Fig. 9. We assume that around 50 per cent of the star-forming population will be dominated by a disc-like surface brightness profile and hence only half the population will be strongly affected by the disc-fading. Moreover, the disc-like, star-forming galaxies will on average have higher ellipticities than their quenched remnants, thus the *circularized* effective radii will be affected less by the fading of the disc component. However, assuming that half the population decreases 30 per cent in size when quenched we still are able to reproduce the size-increase from $z = 2.5$ to $z = 1$, see grey shaded area in Fig. 9. At later times the size-increase is not strong enough due to the declining number of star-forming galaxies. This is a quite conservative estimate as the details will depend on quenching mechanism, bulge-to-disc ratio and secular evolution and merging after quenching. We note, however, that this is qualitatively a possible way to reconcile the findings from Newman et al. (2012) who are able to reproduce the size increase by minor merging from $z = 1$ to $z = 0$, but fail to do so at earlier times.

6.2. Evolution of the scatter in sizes

Each of the above processes, in addition to directing the mass–size evolution, will affect the observed scatter of the mass–size relation and its evolution in different ways. Merging has been shown by Nipoti et al. (2012) to increase the scatter in the mass–size relation. The authors show that size evolution within a dissipation-less (“dry”) merger-only scenario leads to significantly higher scatter than is observed at $z = 0$ (Nipoti, Treu, Leauthaud, Bundy, Newman, & Auger 2012). Mergers are certainly on-going between $z = 2$ and $z = 0$ and Nipoti et al. conclude that there must be a finely tuned balance between the different processes in their merging model in order to reproduce the tight observed relation at $z = 0$. This type of fine tuning is not a general characteristic of the galaxy population(s) and is extremely unlikely to persist in real-world systems. However, the models by Nipoti et al. 2012 only consider dry mergers of spheroids, which given the diverse population of galaxies at high redshift is an unrealistic scenario.

In the case of “dilution” of the population via quenching, the scatter will increase due to the addition of a new population of larger, quiescent galaxies. By using our toy model (see Sect. 5.0.2) for the quenching case, we have shown that we are able to reproduce the observed increase in both number density and mean size of quenched galaxies as functions of redshift out to $z \sim 2$, see Fig. 9. A similar result has been found by others, e.g., Carollo et al. (2013) out to $z \sim 1$. However, our model shows that the scatter should increase by up to ~ 0.1 dex in the redshift range, $0.4 < z < 2.5$. This is inconsistent with the observations presented here, i.e., that the observed scatter in sizes is consistent with being constant from redshift $z \approx 2$ to $z = 0$. With the effect of disc-fading after quenching we still observe a significant increase in the scatter, though smaller (~ 0.05 dex).

So far, most studies have focused on the role of merging only, especially dry minor merging, as the driver of size evolution since this mechanism is very efficient in terms of increasing the size of a galaxy without adding too much mass to the system (McLure et al. 2013). However, Nipoti et al. (2009, 2012) find that dry merging in a Λ CDM cosmology is insufficient to explain the needed increase in size. As we show with our model for size evolution in Sect. 5.0.2, the addition of larger, quenched galaxies means that each individual galaxy needs to undergo less size-evolution. The combination of different galaxy-galaxy interactions, both gas-rich and gas-poor, may then regulate the size-evolution of individual systems such that the scatter remains constant through time. Also, individual systems must evolve at high redshift as such compact galaxies locally are very rare (Trujillo et al. 2009) and merging of galaxies is an obvious mechanism for this evolution (Damjanov et al. 2009; Taylor et al. 2010; van de Sande et al. 2011; Toft et al. 2012). A cascade of mergers is also the most likely way for galaxies to undergo morphological changes from clumpy and in some cases disc-like at high redshift to spheroidal at low redshift (Naab & Trujillo 2006; Ciotti et al. 2007; Wuyts et al. 2010). In order to study the evolution of galaxies and disentangle the various processes, high resolution, cosmological simulations are needed, which take into account both gas-rich and gas-poor galaxy in-

teractions on the entire population of star-forming and quenched galaxies. These should be able to reproduce the lack of evolution in the scatter of the sizes.

7. CONCLUSION

We have presented a spectroscopic sample of quiescent galaxies at redshift $z \approx 2$ with 12 targets (16 including star-forming galaxies). We have shown that our sample is nearly complete for masses above $\log(M_*) > 10.9$. We are therefore able to draw representative conclusions about the population of quiescent galaxies. We find that the galaxies are smaller than locally observed galaxies by a factor of 2.5 on average, consistent with previous results, and we show for the first time for a *spectroscopic* sample that the slope and scatter of the mass–size relation at $z = 2$ are consistent with their local values. We use the fact that the scatter remains constant from $z = 2$ to $z = 0$ as a tool to study the evolutionary mechanism that drives the size-increase of this population. We find that while the addition of larger galaxies quenched at later times can explain the increase of the average size of the population the scatter increases in contrast with the results presented here. Other processes, such as the combined influence from dry and wet mergers, must therefore be needed in order to keep the scatter constant and to make the number density of the most compact galaxies evolve in a way that is consistent with their rarity in the local Universe.

The Dark Cosmology Centre is funded by the DNRF. JK acknowledges support from an ESO studentship. ST and AZ acknowledge support from the Lundbeck foundation. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

REFERENCES

- Bashtannyk, D. M., & Hyndman, R. J. 2001, *Computational Statistics & Data Analysis*, 36, 279
- Bezanson, R., van Dokkum, P. G., Tal, T., Marchesini, D., Kriek, M., Franx, M., & Coppi, P. 2009, *ApJ*, 697, 1290
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Brammer, G. B., et al. 2012, *ApJS*, 200, 13
- . 2011, *ApJ*, 739, 24
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Bruzual A., G. 1983, *ApJ*, 273, 105
- Cappellari, M., et al. 2011, *MNRAS*, 413, 813
- . 2013, *MNRAS*, 432, 1709
- Carollo, C. M., et al. 2013, preprint (arXiv:1302.5115)
- Cassata, P., et al. 2011, *ApJ*, 743, 96
- Chabrier, G. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763
- Ciotti, L., Lanzoni, B., & Volonteri, M. 2007, *ApJ*, 658, 65
- Daddi, E., Cimatti, A., Renzini, A., Fontana, A., Mignoli, M., Pozzetti, L., Tozzi, P., & Zamorani, G. 2004, *ApJ*, 617, 746
- Daddi, E., et al. 2005, *ApJ*, 626, 680
- Damjanov, I., et al. 2011, *ApJ*, 739, L44
- . 2009, *ApJ*, 695, 101
- Fan, L., Lapi, A., De Zotti, G., & Danese, L. 2008, *ApJ*, 689, L101
- Franx, M., et al. 2003, *ApJ*, 587, L79
- Franx, M., van Dokkum, P. G., Schreiber, N. M. F., Wuyts, S., Labbé, I., & Toft, S. 2008, *ApJ*, 688, 770
- Gobat, R., et al. 2013, arXiv:1305.3576

- Grogin, N. A., et al. 2011, *The Astrophysical Journal Supplement Series*, 197, 35
- Guo, Y., et al. 2009, *MNRAS*, 398, 1129
- Hamilton, D. 1985, *ApJ*, 297, 371
- Hopkins, P. F., Somerville, R. S., Hernquist, L., Cox, T. J., Robertson, B., & Li, Y. 2006, *ApJ*, 652, 864
- Ilbert, O., et al. 2013, *A&A*, 556, A55
- Khochfar, S., & Silk, J. 2006, *ApJ*, 648, L21
- Koekemoer, A. M., et al. 2011, *The Astrophysical Journal Supplement Series*, 197, 36
- Kriek, M., van Dokkum, P. G., Franx, M., Illingworth, G. D., & Magee, D. K. 2009a, *ApJ*, 705, L71
- Kriek, M., et al. 2008, *ApJ*, 677, 219
- Kriek, M., van Dokkum, P. G., Labbé, I., Franx, M., Illingworth, G. D., Marchesini, D., & Quadri, R. F. 2009b, *ApJ*, 700, 221
- Krist, J. E., Hook, R. N., & Stoehr, F. 2011, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8127, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*
- Man, A. W. S., Toft, S., Zirm, A. W., Wuyts, S., & van der Wel, A. 2012, *ApJ*, 744, 85
- McLure, R. J., et al. 2013, *MNRAS*, 428, 1088
- Muzzin, A., et al. 2013, *ApJS*, 206, 8
- Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, *ApJ*, 699, L178
- Naab, T., & Trujillo, I. 2006, *MNRAS*, 369, 625
- Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, *ApJ*, 746, 162
- Nipoti, C., Treu, T., Auger, M. W., & Bolton, A. S. 2009, *ApJ*, 706, L86
- Nipoti, C., Treu, T., Leauthaud, A., Bundy, K., Newman, A. B., & Auger, M. W. 2012, *MNRAS*, 422, 1714
- Onodera, M., et al. 2010, *ApJ*, 715, L6
- . 2012, *ApJ*, 755, 26
- Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H. 2012, *ApJ*, 744, 63
- Patel, S. G., et al. 2013, *ApJ*, 766, 15
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, 124, 266
- Pirzkal, N., et al. 2004, *ApJS*, 154, 501
- Poggianti, B. M., et al. 2013, *ApJ*, 762, 77
- Shen, S., Mo, H. J., White, S. D. M., Blanton, M. R., Kauffmann, G., Voges, W., Brinkmann, J., & Csabai, I. 2003, *MNRAS*, 343, 978
- Szomoru, D., Franx, M., & van Dokkum, P. G. 2012, *ApJ*, 749, 121
- Szomoru, D., et al. 2010, *ApJ*, 714, L244
- Taylor, E. N., Franx, M., Glazebrook, K., Brinchmann, J., van der Wel, A., & van Dokkum, P. G. 2010, *ApJ*, 720, 723
- Toft, S., Franx, M., van Dokkum, P., Förster Schreiber, N. M., Labbe, I., Wuyts, S., & Marchesini, D. 2009, *ApJ*, 705, 255
- Toft, S., Gallazzi, A., Zirm, A., Wold, M., Zibetti, S., Grillo, C., & Man, A. 2012, *ApJ*, 754, 3
- Toft, S., et al. 2007, *ApJ*, 671, 285
- Trujillo, I., Carrasco, E. R., & Ferré-Mateu, A. 2012, *ApJ*, 751, 45
- Trujillo, I., Cenarro, A. J., de Lorenzo-Cáceres, A., Vazdekis, A., de la Rosa, I. G., & Cava, A. 2009, *ApJ*, 692, L118
- Trujillo, I., et al. 2006, *ApJ*, 650, 18
- van de Sande, J., et al. 2013, *ApJ*, 771, 85
- . 2011, *ApJ*, 736, L9
- van der Wel, A., et al. 2011, *ApJ*, 730, 38
- van Dokkum, P. G., et al. 2008, *ApJ*, 677, L5
- van Dokkum, P. G., Kriek, M., & Franx, M. 2009, *Nature*, 460, 717
- van Dokkum, P. G., et al. 2010, *ApJ*, 709, 1018
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., Toft, S., Kriek, M., & Labbé, I. 2010, *ApJ*, 713, 738
- Wuyts, S., Cox, T. J., Hayward, C. C., Franx, M., Hernquist, L., Hopkins, P. F., Jonsson, P., & van Dokkum, P. G. 2010, *ApJ*, 722, 1666
- Wuyts, S., et al. 2007, *ApJ*, 655, 51
- Zirm, A. W., et al. 2007, *ApJ*, 656, 66