Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **501,** 2659–2676 (2021) doi:10.1093/mnras/staa3766
Advance Access publication 2020 December 5

Compact, bulge-dominated structures of spectroscopically confirmed quiescent galaxies at $z \approx 3$

Peter Lustig [1]★ Veronica Strazzullo,[1,2,3,4] Chiara D'Eugenio,[5] Emanuele Daddi [5] Maurilio Pannella [1,2] Alvio Renzini [6] Andrea Cimatti,[7,8] Raphael Gobat,[9] Shuowen Jin [10,11] Joseph J. Mohr[1,12] and Masato Onodera[13,14]

[1]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[2]*University of Trieste, Piazzale Europa, 1, I-34127 Trieste, Italy*
[3]*INAF-Osservatorio Astronomico di Brera, Via Brera 28, I-20121 Milano, Italy*
[4]*INAF - Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
[5]*CEA, IRFU, DAp, AIM, Universit Paris-Saclay, Universit Paris Diderot, Sorbonne Paris Cit, CNRS, F-91191 Gif-sur-Yvette, France*
[6]*INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[7]*University of Bologna, Department of Physics and Astronomy (DIFA), Via Gobetti 93/2, I-40129 Bologna, Italy*
[8]*INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[9]*Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla, 4059 Valparaíso, Chile*
[10]*Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Spain*
[11]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Spain*
[12]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[13]*Subaru Telescope, National Astronomical Observatory of Japan, National Institutes of Natural Sciences, 650 North A'ohoku Place, Hilo, HI 96720, USA*
[14]*Department of Astronomical Science, The Graduate University for Advanced Studies, SOKENDAI, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

Accepted 2020 November 30. Received 2020 November 3; in original form 2020 August 11

ABSTRACT

We study structural properties of spectroscopically confirmed massive quiescent galaxies at $z \approx 3$ with one of the first sizeable samples of such sources, made of ten $10.8 < \log(M_\star/\mathrm{M}_\odot) < 11.3$ galaxies at $2.4 < z < 3.2$ in the COSMOS field whose redshifts and quiescence are confirmed by Hubble Space Telescope (*HST*) grism spectroscopy. Although affected by a weak bias toward younger stellar populations, this sample is deemed to be largely representative of the majority of the most massive and thus intrinsically rarest quiescent sources at this cosmic time. We rely on targeted *HST*/Wide-Field Camera 3 observations and fit Sérsic profiles to the galaxy surface brightness distributions at $\approx$ 4000- Å rest frame. We find typically high Sérsic indices and axis ratios (medians $\approx$ 4.5 and 0.73, respectively) suggesting that, at odds with some previous results, the first massive quiescent galaxies may largely be already bulge-dominated systems. We measure compact galaxy sizes with an average of ≈ 1.4 kpc at $\log(M_\star/\mathrm{M}_\odot) \approx 11.2$, in good agreement with the extrapolation at the highest masses of previous determinations of the stellar mass–size relation of quiescent galaxies, and of its redshift evolution, from photometrically selected samples at lower and similar redshifts. This work confirms the existence of a population of compact, bulge dominated, massive, quiescent sources at $z \approx 3$, providing one of the first statistical estimates of their structural properties, and further constraining the early formation and evolution of the first quiescent galaxies.

Key words: galaxies: evolution – galaxies: high-redshift – galaxies: structure.

1 INTRODUCTION

Structural properties of galaxies in the nearby universe correlate with their stellar population properties. Early-type galaxies are characterized by a higher central concentration and typically lower apparent ellipticity than late-type galaxies, and generally have a low specific star formation rate (sSFR). Up to a stellar mass of $\log(M_\star/\mathrm{M}_\odot) \approx 11$, early-type galaxies are more compact than late-type galaxies and show a steeper stellar mass versus size relation (e.g. Shen et al. 2003; Guo et al. 2009).

Up to $z \approx 1$, high axis ratios are largely ubiquitous in the most massive $\log(M_\star/\mathrm{M}_\odot) \gtrsim 11$ quiescent galaxies, although larger fractions of lower mass galaxies show lower axis ratios; this suggests that the mechanisms forming the most massive quiescent sources also result in the formation of bulge-dominated, spheroidal structures (van der Wel et al. 2009b; Holden et al. 2012). In fact, integral field spectroscopy showed that the vast majority of early-type galaxies in the nearby universe are fast rotators, with slow rotators dominating the early-type galaxy population only at the high-mass end ($M_\star \gtrsim 2 \times 10^{11}\,\mathrm{M}_\odot$; e.g. Emsellem et al. 2011; Cappellari 2016).

Structural properties of massive galaxies at higher redshift are more sparsely investigated and have produced more controversial results. Stockton, Canalizo & Maihara (2004) and Stockton et al. (2008) provided the first constraints on the structure of massive

★ E-mail: peter.lustig@physik.lmu.de

quiescent galaxies at $z \approx 2.5$ and revealed a higher fraction of quiescent galaxies with low Sérsic index profiles and smaller axis ratios with respect to low-redshift samples. Such scenario has been strengthened by following works with larger samples (McGrath et al. 2008; van Dokkum et al. 2008; Bundy et al. 2010; van der Wel et al. 2011; Chang et al. 2013; McLure et al. 2013; Hsu, Stockton & Shih 2014; Bezanson et al. 2018). Recently, Hill et al. (2019) investigated the axial ratio evolution of star-forming and quiescent galaxies over the redshift range $0.2 < z < 4.0$, finding that massive [$\log(M_\star/\mathrm{M}_\odot) > 11$] quiescent galaxies at $2.5 < z < 3.5$ are as flat as star-forming galaxies. Limited measurements of rotation curves indeed provide evidence for the existence of rotationally supported massive quiescent galaxies at high redshift (Newman, Belli & Ellis 2015; Toft et al. 2017; Newman et al. 2018). None the less, the coupling of structural and stellar population properties of galaxies at higher redshifts remains debated, as other studies find that the correlation between early-type structure and low sSFR holds at least up to $z \approx 3$, suggesting that morphological transformation towards bulge-dominated systems is tightly related to quenching of star formation already at high redshift (Bell et al. 2012; Lang et al. 2014; Tacchella et al. 2015; Mowla et al. 2019; Esdaile et al. 2020). It has in fact been shown that galaxies beyond a given stellar mass or central stellar mass density threshold are largely quiescent (e.g. Kauffmann et al. 2003; Brinchmann et al. 2004; Franx et al. 2008; Peng et al. 2010b; van Dokkum et al. 2015; Whitaker et al. 2017), and that – although it remains unclear whether mass or density is the actual driver (Lilly & Carollo 2016) – the most massive star-forming galaxies that at high-redshift approach such density threshold are very likely to rapidly quench, given the drop in their number density at lower redshifts (Mowla et al. 2019).

A possible mechanism to explain the correlation between structural and stellar population properties is the compaction of a star-forming disc in a first step, followed by quenching (possibly also as a consequence of the morphological transformation). The compaction of the disc can be a result of gas inflow from filaments or mergers (e.g. Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Dekel et al. 2009a), causing violent disc instabilities that drive dissipative gas inflow in the centre. This leads to a compact galaxy with a high star formation rate (SFR; e.g. Dekel, Sari & Ceverino 2009b; Burkert et al. 2010; Dekel et al. 2013; Dekel & Burkert 2014; Zolotov et al. 2015; Gómez-Guijarro et al. 2019; Wu et al. 2020). Multiple mechanisms can then quench star formation, as suggested by simulations: gas consumption by star formation, stellar, and active galaxy nucleus (AGN) feedback, as well as morphological quenching, can produce fast quenching at high redshift, while virial shock heating, gravitational infall, and AGN feedback can maintain quenching at lower redshift (e.g. Dekel & Silk 1986; Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Ciotti & Ostriker 2007; Dekel & Birnboim 2008; Khochfar & Ostriker 2008; Martig et al. 2009; Dekel et al. 2009a; Tacchella et al. 2016). Bulges embedded in star-forming discs can remain starved from accreted gas and maintain quenching if the infalling gas has a too high angular momentum to reach the bulge (Renzini et al. 2018).

Many studies have shown that the average size of distant quiescent galaxies at a given stellar mass is lower than for lower redshift counterparts (e.g. Daddi et al. 2005; Trujillo et al. 2006; Toft et al. 2007; Cimatti et al. 2008; Buitrago et al. 2008; Cimatti, Nipoti & Cassata 2012; Carollo et al. 2013; Cassata et al. 2013; Kubo et al. 2018; Mowla et al. 2019, among many others). Although an evolution in the average size at fixed mass is also observed for late-type galaxies, it is milder than for early types. With a large sample drawn from the CANDELS/3D-HST survey (Grogin et al. 2011; Koekemoer et al. 2011; Momcheva et al. 2016), van der Wel et al. (2014) studied morphologies of quiescent and star-forming galaxies with redshifts $0 < z < 3$. They find a redshift independent slope of the mass–size relation that is steeper for quiescent than for star-forming galaxies, and a size growth of massive quiescent galaxies of nearly an order of magnitude since $z \approx 3$, compared to a factor of $\approx$3 for star-forming sources. Using measurements from the COSMOS-DASH survey, Mowla et al. (2019) extended the van der Wel et al. (2014) sample to higher stellar masses (162 galaxies at $1.5 < z < 3.0$ with $\log(M_\star/\mathrm{M}_\odot) > 11.3$), which are poorly probed in the CANDELS/3D-HST survey due to the intrinsically very low number density of such sources, and find consistent results. However, even this survey only adds two quiescent galaxies to the van der Wel et al. (2014) sample at $z > 2.5$. As an alternative to overcome the problem of small sample sizes of the most distant, massive quiescent galaxies in deep fields, targeted imaging has been used to study these objects up to $z \approx 4$ (Straatman et al. 2015; Kubo et al. 2018), supporting the findings of strong average size growth of the quiescent galaxy population.

The observed redshift evolution of the mass–size relation of quiescent galaxies can be explained by a combination of different effects. Although gas-rich mergers, resulting in central starbursts, are not an efficient way to increase galaxy size (Lin et al. 2007, 2008; Perez, Michel-Dansac & Tissera 2011; Athanassoula et al. 2016), gas-poor minor mergers are often considered a viable and potentially significant channel for size growth of quiescent galaxies (Khochfar & Silk (Bell et al. 2006; Khochfar & Silk 2006; Naab, Khochfar & Burkert 2006; Lin et al. 2008; Bezanson et al. 2009; Naab, Johansson & Ostriker 2009; Oser et al. 2010, 2012; Trujillo, Ferreras & de La Rosa 2011; Bédorf & Portegies Zwart 2013). Progenitor bias is also often considered as an important contribution to the evolution of the mass–size relation, because of the significant drop of the quiescent galaxy population towards higher redshifts, implying a progenitor-descendant mismatch when comparing quiescent galaxy samples at different redshifts (van Dokkum & Franx 1996, 2001; Carollo et al. 2013; Cassata et al. 2013; Poggianti et al. 2013). To minimize the effect of progenitor bias, Belli, Newman & Ellis (2014) and Stockmann et al. (2020) investigated size evolution at constant velocity dispersion (which is found to remain approximately unchanged for quiescent systems, van der Wel et al. 2009a; Bezanson, van Dokkum & Franx 2012), finding that size growth of individual galaxies may, in fact, have a significant role in the observed mass–size evolution. The observed evolution is thus likely produced by a combination of both galaxy growth and progenitor bias. Additional complications come from the use of light as a tracer of stellar mass. Radial colour – and thus mass-to-light ratio (M/L) – gradients can lead to significant differences between half-light and half-mass radii. Suess et al. (2019a, b) find that colour gradients of quiescent galaxies are nearly flat at $z \gtrsim 2$, increase with decreasing redshift and are stronger in massive, larger and redder galaxies. Stellar mass versus half-mass–size relations of quiescent galaxies are shallower than stellar mass versus (rest-frame optical) half-light–size relations, and the growth of half-mass sizes towards lower redshifts is milder than for optical half-light sizes.

In most studies of the highest redshift quiescent sources, relying on purely photometric observations, the classification of star-forming versus quiescent galaxies is performed by exploiting the correlation between sSFR and galaxy colours in properly chosen passbands (Daddi et al. 2004; Labbé et al. 2005; Williams et al. 2009; Ilbert et al. 2010). Especially at high redshift, where the number density of massive quiescent galaxies and the quiescent galaxy fraction decrease significantly (Whitaker et al. (Marchesini et al. 2010; Whitaker et al. 2010; Brammer et al. 2011; Ilbert et al. 2013;

Table 1. Main properties and estimated morphological parameters of the 10 targets. The column ID_C lists for convenience the IDs from the COSMOS Laigle et al. (2016) catalogue. The z_{phot} column lists the photometric redshifts used for target selection, while z_{spec} lists the spectroscopic redshifts from D'Eugenio et al. (2020). The column $\log(M_\star/M_\odot)$ lists stellar mass estimates (see Section 2.2 for details, including a discussion of the uncertainties). The columns mag, r (sky), n, and q list the GALFIT best-fitting values for the Sérsic profile; radii are effective radii along the semimajor axis. Uncertainties are estimated as explained in Section 3.1. The physical radii in kpc, r_e (5000 Å), are inferred sizes at 5000 Å, derived as explained in Section 4.2.

ID	ID_C	RA (h:m:s)	Dec. (d:m:s)	z_{phot}	z_{spec}	$\log(M_\star/M_\odot)$	mag	r_e (sky) (arcsec)	r_e (5000 Å) (kpc)	n	q
1	135730	10:01:39.98	01:29:34.49	2.6	$2.841^{+0.021}_{-0.018}$	11.14	$21.99^{+0.10}_{-0.09}$	$0.40^{+0.07}_{-0.06}$	$3.07^{+0.57}_{-0.47}$	$4.7^{+0.8}_{-0.8}$	$0.50^{+0.03}_{-0.03}$
2	137182	10:00:57.35	01:29:39.46	2.7	$2.557^{+0.005}_{-0.005}$	11.27	$21.32^{+0.02}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$1.18^{+0.04}_{-0.04}$	$3.0^{+0.3}_{-0.2}$	$0.96^{+0.02}_{-0.02}$
3	252568	09:57:48.57	01:39:57.82	2.8	$3.124^{+0.003}_{-0.003}$	11.32	$21.88^{+0.10}_{-0.10}$	$0.33^{+0.08}_{-0.05}$	$2.37^{+0.58}_{-0.37}$	$6.2^{+1.2}_{-1.1}$	$0.78^{+0.04}_{-0.04}$
4	361413	10:02:00.97	01:50:24.28	3.2	$3.230^{+0.007}_{-0.006}$	10.75	$23.37^{+0.08}_{-0.11}$	$0.07^{+0.01}_{-0.01}$	$0.46^{+0.08}_{-0.07}$	$4.8^{+2.9}_{-1.3}$	$0.86^{+0.11}_{-0.11}$
5	447058	09:59:11.77	01:58:32.96	2.5	$2.665^{+0.003}_{-0.007}$	11.11	$22.20^{+0.02}_{-0.02}$	$0.21^{+0.01}_{-0.01}$	$1.63^{+0.05}_{-0.05}$	$1.2^{+0.1}_{-0.1}$	$0.79^{+0.03}_{-0.03}$
6	478302	09:59:01.31	02:01:34.15	2.6	$2.801^{+0.005}_{-0.002}$	11.13	$22.23^{+0.02}_{-0.02}$	$0.109^{+0.004}_{-0.004}$	$0.84^{+0.03}_{-0.03}$	$2.3^{+0.3}_{-0.3}$	$0.56^{+0.04}_{-0.04}$
7	503898	10:01:31.86	02:03:58.79	2.6	$2.674^{+0.005}_{-0.009}$	11.32	$21.60^{+0.13}_{-0.11}$	$0.57^{+0.13}_{-0.11}$	$4.45^{+1.04}_{-0.89}$	$6.3^{+1.0}_{-1.2}$	$0.60^{+0.03}_{-0.03}$
8	575436	10:00:43.76	02:10:28.71	2.8	$2.998^{+0.002}_{-0.003}$	11.17	$22.30^{+0.03}_{-0.04}$	$0.10^{+0.01}_{-0.01}$	$0.75^{+0.05}_{-0.04}$	$4.3^{+0.8}_{-0.7}$	$0.33^{+0.03}_{-0.03}$
9	707962	09:59:32.52	02:22:21.99	2.6	$2.667^{+0.015}_{-0.002}$	11.3	$21.66^{+0.15}_{-0.17}$	$0.29^{+0.16}_{-0.07}$	$2.27^{+1.24}_{-0.58}$	12[a]	$0.87^{+0.05}_{-0.05}$
10	977680	10:00:12.65	02:47:23.47	2.5	$2.393^{+0.011}_{-0.000}$	11.1	$22.39^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$1.19^{+0.06}_{-0.05}$	$2.6^{+0.4}_{-0.3}$	$0.67^{+0.04}_{-0.04}$

[a]If the Sérsic index is fixed to 4, the effective radius decreases to 0.14 arcsec (2.1σ), see Section 3.

Muzzin et al. 2013; Mowla et al. 2019) and the bimodality in colour sequences is less pronounced (Muzzin et al. 2013; Laigle et al. 2016), misclassification can lead to a significant contamination of quiescent galaxy samples from star-forming objects. Spectroscopic confirmation of quiescence can help securing higher purity samples of quiescent galaxies. However, spectroscopically confirming very distant quiescent sources is difficult and observationally expensive compared to star-forming galaxies at similar redshifts because of the lack of strong emission lines. Direct spectroscopic confirmation of quiescent sources currently reaches out to $z \approx 4$, and is based on the 4000 Å break, overall continuum shape, and/or weaker features as Fe and Mg absorption lines (Cimatti et al. 2004; Glazebrook et al. 2004, 2017; Kriek et al. 2006; Gobat et al. 2012; Onodera et al. 2012, 2015; Marsan et al. 2015; Newman et al. 2015; Hill et al. 2016; Gobat et al. 2017; Marsan et al. 2017; Newman et al. 2018; Schreiber et al. 2018; Tanaka et al. 2019; Esdaile et al. 2020; Valentino et al. 2020; Forrest et al. 2020a, b). The morphological properties of sizeable samples of spectroscopically confirmed quiescent galaxies have only been analysed up to $z < 2.3$ (Cimatti et al. 2008; van Dokkum et al. 2008; Belli, Newman & Ellis 2017; Stockmann et al. 2020). At higher redshifts, investigations are limited to a handful of galaxies at most (Gobat et al. 2012; Marsan et al. 2015; Hill et al. 2016; Tanaka et al. 2019; Esdaile et al. 2020). In this work, we investigate structural properties of a spectroscopically confirmed sample of 10 quiescent galaxies at $2.4 < z < 3.2$ with stellar masses of $\log(M_\star/M_\odot) \gtrsim 11$, relying on targeted *Hubble Space Telescope* (*HST*) WFC3/*F*160*W* imaging and G141 grism observations. This sample contains ≈1/4 of all spectroscopically confirmed quiescent galaxies at $z > 2.4$. Our sample is presented in Section 2. In Section 3, we explain our analysis and methods. In Section 4, we present and discuss our results. Section 5 summarizes our findings and conclusions.

We assume a Lambda cold dark matter cosmology with $H_0 = 71$, $\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. Magnitudes are given in the AB system.

2 THE QUIESCENT GALAXY SAMPLE

2.1 Sample selection

We selected high-redshift quiescent galaxy candidates for *HST* grism follow-up from the McCracken et al. (2010) photometric catalogue of the 2 deg^2 COSMOS field. Initially, we selected passive BzK (pBzK) galaxies (Daddi et al. 2004) that satisfy the conditions:

$$\mathrm{BzK} = (z_{AB} - K_{AB}) - (B_{AB} - z_{AB}) > -0.2, \quad (1)$$

$$\mathrm{zK} = (z_{AB} - K_{AB}) > 2.5. \quad (2)$$

These criteria select high-redshift (typically $z \gtrsim 1.4$) passive galaxies purely based on observed colours, without relying on photometric redshift estimation or spectral energy distribution (SED) analysis to identify quiescent sources. Given the low signal-to-noise ratio (S/N) of even massive quiescent galaxies at $z > 2$ in the available *B*- and *z*-BAND IMAGing, leading to large uncertainties in the formal passive versus star-forming BzK classification of such sources, we also retained galaxies with S/N < 5 in these bands, independent of their classification as quiescent or star-forming BzK galaxies. We then considered photometric redshifts (z_{phot}) for the selected galaxies estimated with the software EASY (Brammer, van Dokkum & Coppi 2008) and specifically calibrated to better estimate photometric redshifts of high-redshift quiescent galaxies (see details in Strazzullo et al. 2015). These photometric redshifts are listed in Table 1. We removed from the sample all galaxies with $z_{phot} < 2.5$, as well as galaxies classified as star-forming from their rest-frame UVJ colours (Williams et al. 2009) as estimated by EAZY assuming the galaxy photometric redshift. We performed SED fitting with FAST (Kriek et al. 2009) with different model libraries, including (1) a generic setup with delayed exponential star formation history (SFH) and dust attenuation up to $A_V = 5$ mag, assuming a Calzetti (2001) dust attenuation law, (2) constant SFR with A_V up to 5 mag, (3) only quiescent (including very young quiescent, given the redshift of our targets) models (age/$\tau > 4$, age > 0.5 Gyr). Based on this analysis, we discarded all candidates with an SED suggesting a possible star-forming solution. To further reduce potential contamination of the *HST* follow-up target sample from star-forming sources, we also deprioritised candidates with an S/N ≥ 4 at 24 μm in the Le Floc'h et al. (2009) catalogue, except if they had a well probed, convincingly quiescent SED with no plausible star-forming solution, suggesting that the 24 μm emission could be powered by an AGN. We then selected from these candidates suitable targets for *HST* grism follow-up observations. In order to observe a first, sizeable sample of $z \approx 3$ quiescent candidates, we focused on massive galaxies for which a

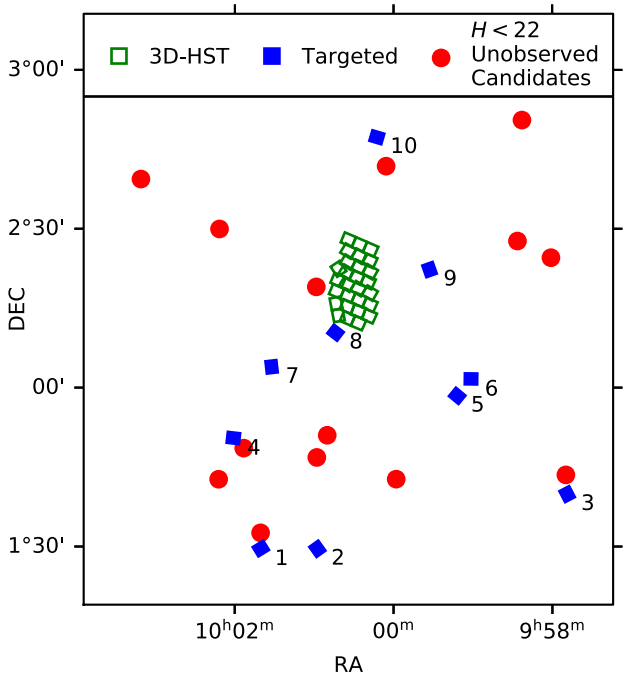

Figure 1. Selection of targets for *HST* follow-up observations in the 2 deg^2 COSMOS field. Blue rectangles show the WFC3 footprint of the observations acquired for this project. Red symbols show the remaining quiescent galaxies from our $H < 22$, $z > 2.5$ candidate sample that were not included in the target list. Green rectangles show the footprint of the 3D-HST survey field in COSMOS.

sufficiently high S/N spectrum to measure a reliable redshift could be obtained in 1−2 orbits. To this aim, we simulated for each candidate the grism spectrum that could be obtained within this observing time, assuming the source photometric redshift and best-fitting SED model, modelling the simulated spectrum to estimate the redshift. This observational constraint largely limited the viable targets to sources brighter than $H_{\rm AB} \approx 22$, leading to a sample of 23 sources that are shown in Fig. 1. Owing to the low number density of such massive, quiescent galaxies at $z \approx 3$, none of these objects is found in the CANDELS/3D-HST COSMOS field (Grogin et al. 2011; Koekemoer et al. 2011; Momcheva et al. 2016). Although such bright ($H_{\rm AB} < 22$) targets were favoured because of the observational reasons discussed above, as well as of higher S/N photometry resulting in a more robust characterization of the galaxy SED, we also explored fainter candidates that potentially allow us to probe higher redshift galaxies. We thus included in the final target sample a fainter ($H_{\rm AB} \approx 23$) source at $z_{\rm phot} = 3.2$ for which – in contrast to most similarly faint candidates – the SED modelling discussed above was able to reject star-forming solutions at high confidence. The final target sample of 10 sources with photometric redshifts between 2.5 and 3.2 is listed in Table 1 and shown in Fig. 1.

The selected targets have been observed with the G141 grism and direct imaging in the *F*160*W* band with the Wide-Field Camera 3 (WFC3) on board of *HST* (program ID 15229, PI: E. Daddi). D'Eugenio et al. (in preparation, see also D'Eugenio et al. 2020) have estimated spectroscopic redshifts from the grism spectra, which are shown in Table 1. They combined the grism spectra with photometric measurements from the Laigle et al. (2016) catalogue and performed a stellar population analysis. By comparing the goodness-of-fit of constant SFR versus passive templates with exponentially declining SFHs, star-forming solutions could be rejected for all galaxies in the sample (see full details and discussion in D'Eugenio et al., in preparation).

From the analysis of the stacked spectrum of all galaxies in the sample (except ID 7), D'Eugenio et al. (2020) derived an sSFR of $(4.35 \pm 2.47) \times 10^{-11}$ yr^{-1}, which is 60 times below the main sequence of star-forming galaxies at the median redshift of $z = 2.8$ (Schreiber et al. 2015). The lookback time where 50 per cent of the stellar mass of the stacked sample was formed is $t_{50} = 300^{+200}_{-50}$ Myr.

We note that the median and NMAD (normalized mean absolute difference) scatter of $(z_{\rm spec} - z_{\rm phot})/(1 + z_{\rm spec})$ for the sample studied here, using the grism redshifts from D'Eugenio et al. (2020) and the Strazzullo et al. (2015) photometric redshifts used for the sample selection, are 0.03 and 0.06; we thus assume that no significant biases are introduced in the sample studied here by uncertainties in the photometric redshifts used for the sample selection.

2.2 SED-modelling and stellar mass estimates

We perform SED fitting to estimate stellar masses of the targets from multiband photometry from the COSMOS2015 catalogue (Laigle et al. 2016) adopting the spectroscopic redshifts measured in D'Eugenio et al. (2020). We use FAST++ [1] to fit Bruzual & Charlot (2003) population synthesis models to 29 photometric bands[2] from 0.42 to 8 μm (including narrow bands). We assume a Chabrier (2003) initial mass function, a Calzetti (2001) dust attenuation law and a delayed exponentially declining SFH with $7 \leq \log(\tau/{\rm yr}) \leq 10$. To allow for a more direct comparison with van der Wel et al. (2014, see Section 4), we also estimate stellar masses assuming an exponentially declining SFH, finding no systematics and individual stellar mass estimates differing by at most 0.05 dex for this specific sample, having no impact on our analysis. The metallicity is fixed to solar; leaving it free affects the mass estimates by at most 0.07 dex. The best-fitting SED models are shown in Fig. 2. The formal uncertainties on the estimated stellar masses with the given SED-fitting setup are $\lesssim 0.06$ dex; we stress that these uncertainties do not include known sources of statistical and systematic errors (e.g. Maraston et al. 2006; Longhetti & Saracco 2009; Muzzin et al. 2009; Conroy 2013; Pacifici et al. 2015), and that more realistic absolute uncertainties on the individual mass estimates are likely around a factor of ≈2.

IDs 2, 4, 7, and 10 have close neighbours in our *F*160*W* imaging that are undetected in the Laigle et al. (2016) catalogue (see Section 3). For these targets, we scale the estimated stellar masses by the fraction of the target flux to the total flux including the undetected neighbours within the 3-arcsec aperture used in Laigle et al. (2016), assuming the *F*160*W* fluxes measured in Section 3. This correction decreases the masses of IDs 2, 4, 7, and 10 by 0.01, 0.11, 0.02, and 0.01 dex, respectively. The resulting stellar masses are listed in Table 1.[3]

The median estimated stellar mass of our sample is $\log(M_\star/{\rm M}_\odot) = 11.16$ with individual masses in the range $10.8 < \log(M_\star/{\rm M}_\odot) < 11.3$. To ensure that no systematics affect our comparisons with van

[1] https://github.com/cschreib/fastpp

[2] For sources that are observed in the *H* and *Ks* band by both UltraVISTA and WIRCam, we have checked that there is no impact on the stellar mass estimates if the shallower WIRCAM data is removed.

[3] We note that these masses reflect the total fluxes reported in the Laigle et al. (2016) catalogue. We have verified that the total flux estimated by GALFIT on the *F*160*W* imaging is fully consistent with the total flux in the *H* band from the Laigle et al. (2016) catalogue (the average flux ratio for these targets is 0.96 with a dispersion of ≈0.15, accounting for the small colour term between the two filters).

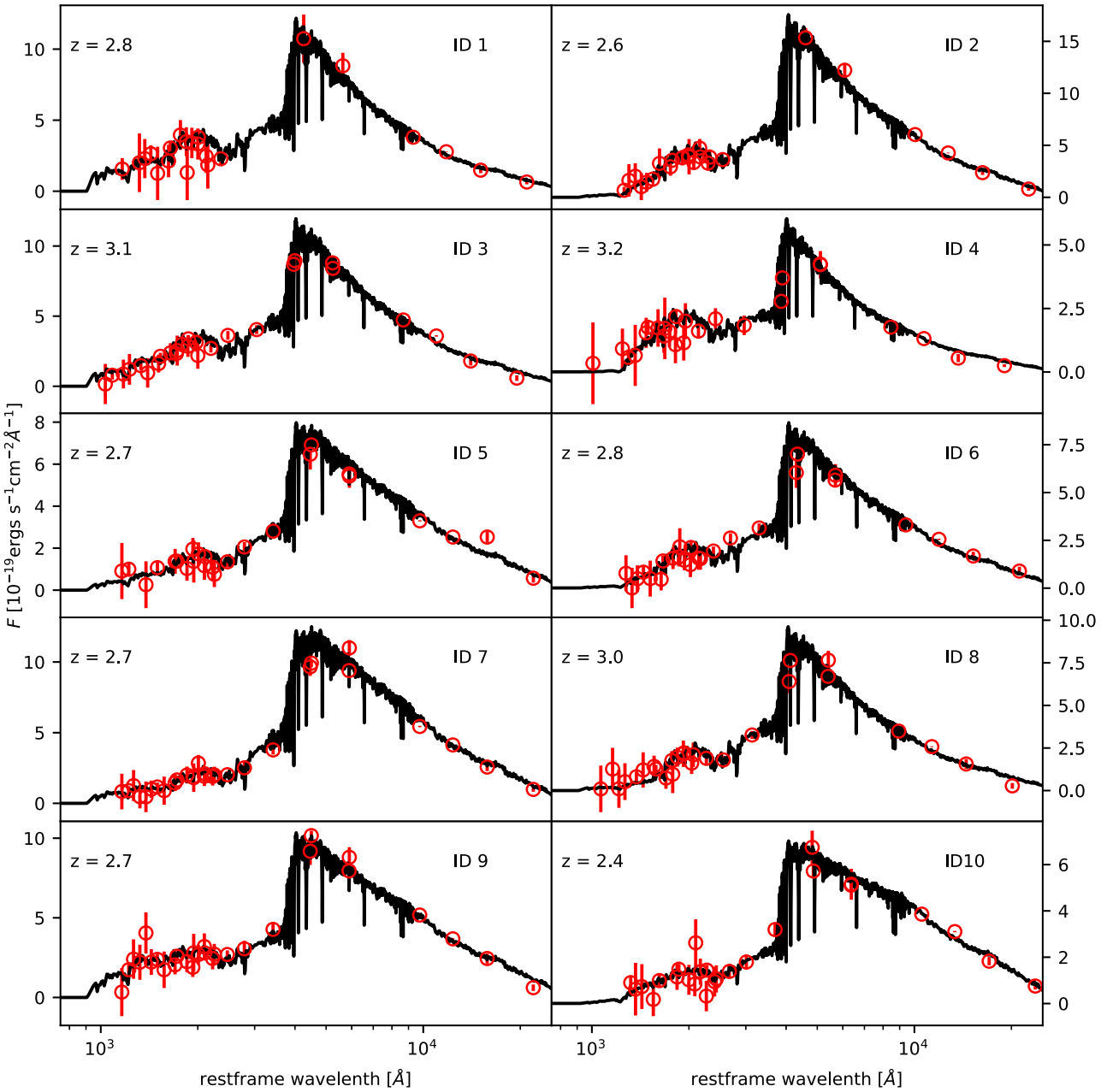

Figure 2. Observed SEDs from 0.42 to 8 μm (Laigle et al. 2016) and best-fitting stellar population models (see Section 2.2).

der Wel et al. (2014, see section 4), who use stellar mass estimates from Skelton et al. (2014), we estimate stellar masses with the same setup for sources from the Skelton et al. (2014) catalogue using Laigle et al. (2016) photometry and redshifts from Skelton et al. (2014). By comparison of the two estimates, we find a statistical scatter on the estimated stellar masses of 0.1 dex and no systematics.[4]

[4]We do not use stellar mass estimates from Laigle et al. (2016) because: (1) we re-estimate stellar masses adopting the grism redshift (see also related discussion in Section 2.3 and Appendix A), and (2) as also reported in Mowla et al. (2019) the stellar mass estimates from Laigle et al. (2016) are systematically higher than those from Skelton et al. (2014) by $\gtrsim 0.1$ dex for sources with $\log(M_\star/\mathrm{M}_\odot) > 10.75$.

2.3 Sample characterization and representativeness

For sources at $z \approx 3$, the observed *H*-band probes the galaxy SED at $\approx$ 4000- Å rest frame, where the M/L ratio is sensitive to the age of the stellar population. Selecting *H*-band bright sources, as discussed in Section 2.1, may therefore bias the sample towards younger and/or less dust-attenuated stellar populations. Depending on quenching mechanisms, and at least at lower redshifts on progenitor bias effects, sizes of younger versus older quiescent sources at fixed stellar mass may differ on average (e.g. Saracco, Longhetti & Andreon 2009; Belli, Newman & Ellis 2015; Yano et al. 2016; Almaini et al. 2017; Williams et al. 2017; Zahid & Geller 2017; Wu et al. 2018). Our $H < 22$ selection could thus potentially result in a bias on the average quiescent galaxy size at a given mass inferred from this sample. In this section, we thus discuss the representativeness of this *H*-selected

sample with respect to the parent (mass-selected) sample of massive quiescent galaxies at this redshift.

To address the relevance of the potential bias in the quiescent sample caused by the H-band selection, we compare the UVJ rest-frame colours of the $H < 22$ quiescent population with those of the full massive galaxy population at $2.5 < z < 3$. For the full parent sample, we match the Laigle et al. (2016) and Muzzin et al. (2013) catalogues in order to fit the Laigle et al. (2016) photometry assuming the Muzzin et al. (2013) photometric redshifts and spectroscopic redshifts from D'Eugenio et al. (2020) for our targets. We choose this approach in order to make use of the deeper photometry in the Laigle et al. (2016) catalogue (that we use throughout in the analysis of our target sample in Section 2.2) and at the same time of the more accurate Muzzin et al. (2013) photometric redshifts for massive quiescent sources at this redshift, as inferred by comparison with spectroscopic samples as shown in Appendix A. We use the same FAST+ + setup as in Section 2.2 to estimate stellar masses and EASY to estimate rest-frame UVJ colours. We consider galaxies more massive than the mass-completeness limit of $\log(M_\star/\mathrm{M}_\odot) = 11.1$ at $z = 3$ from Muzzin et al. (2013), providing a sample of 43 UVJ quiescent galaxies at $2.5 < z < 3.0$.

The location of quiescent galaxies in the UVJ plane correlates with the age of their stellar populations (e.g. Belli, Newman & Ellis 2019). We thus investigate in Fig. 3 the distribution of our targets in the UVJ plane, and more generally of sources brighter than $H = 22$, with respect to the parent population, to constrain possible biases in our sample. All of our targets are well within the UVJ-quiescent region (Williams et al. 2009) except ID 10, which is anyway consistent with being UVJ-quiescent.

Belli et al. (2019) parametrize the relation between stellar population age and UVJ colours by adopting t_{50} as an age estimate. We use this parametrization to investigate the impact of the H-band selection on the fraction of 'post-starburst' ($t_{50} < 800$ Myr) to old passive galaxies in the full sample of $\log(M_\star/\mathrm{M}_\odot) > 11.1$ quiescent galaxies. As it may be expected given the high redshift, a large fraction of the quiescent galaxy sample is made of relatively young sources which at lower redshift are typically classified as post-starburst based on their colours (see also e.g. Whitaker et al. 2012; Marchesini et al. 2014; Maltby et al. 2018; Merlin et al. 2018). To account for the uncertainties on the photometric measurements and redshift estimates, we perturb the source photometry and photometric redshift within the uncertainties and estimate rest-frame UVJ colours accordingly for 10 000 realizations. The inferred median distribution of UVJ colour combination which translates to t_{50} in the Belli et al. (2019) parametrization is shown in the bottom panel of Fig. 3. With this approach, the estimated fraction of post-starburst galaxies in the full massive quiescent sample is (50 ± 9) per cent. Considering only galaxies with $H < 22$, this fraction increases to (77 ± 9) per cent, consistent with our sample in which 9 of 10 galaxies have $t_{50} < 800$ Myr, according to the relation from Belli et al. (2019). Indeed, D'Eugenio et al. (2020) found $t_{50} \leq 800$ Myr for all galaxies in the sample. Therefore, at face value the average stellar age of galaxies in the $\log(M_\star/\mathrm{M}_\odot) > 11.1$, $H < 22$ sample is indeed younger than in the whole $\log(M_\star/\mathrm{M}_\odot) > 11.1$ sample, suggesting that our sample may be more representative of younger, post-starburst quiescent systems (see Fig. 3), and likely biased against the oldest quiescent galaxies at this redshift. If significant morphological transformations happen on longer time-scales than the typical age of this sample we would not be able to see it in our analysis because our sample does not contain these older sources. On the other hand, we stress that the uncertainties on the estimated rest-frame UVJ colours are significantly higher – as expected given the

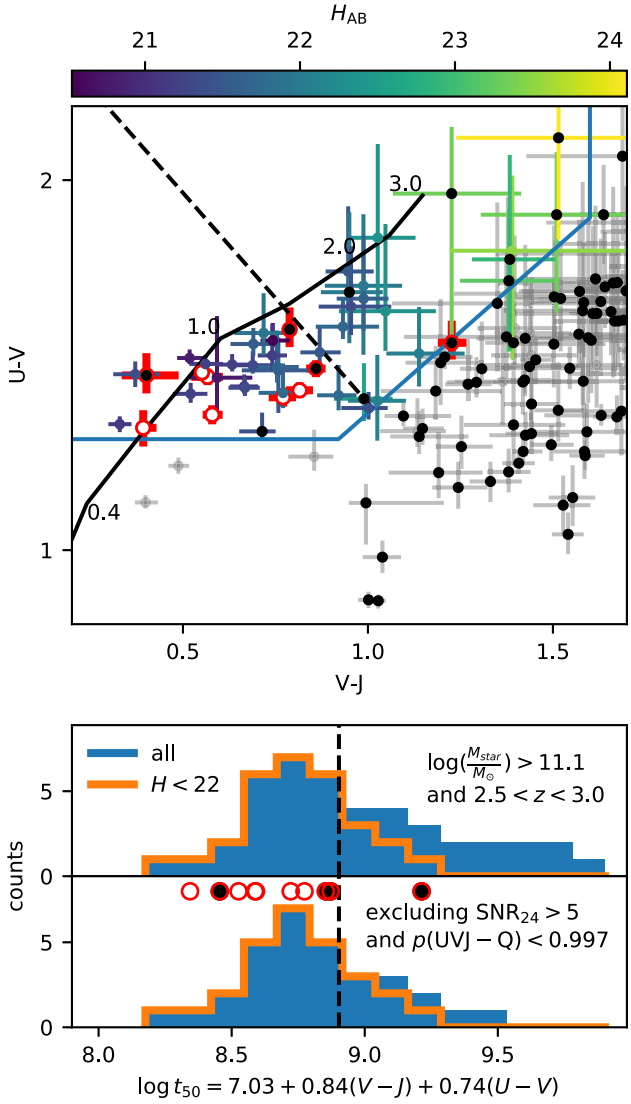

Figure 3. Top panel: rest-frame UVJ colours of massive ($\log(M_\star/\mathrm{M}_\odot) > 11.1$) galaxies in the UltraVISTA COSMOS field (see Section 2.3) in the redshift range $2.5 < z < 3$. For UVJ-quiescent sources, the colours of the symbols scale with the H-band magnitude, as indicated by the colour bar. The sources studied in this work are overplotted as red circles, with colours computed assuming the spectroscopic redshift. The 1σ colour uncertainties account for photometric and redshift uncertainties. 24-μm detections from Jin et al. (2018) are marked with a black dot. The blue line shows the adopted separation between star-forming and quiescent galaxies (Williams et al. 2009). The black solid line shows for reference the evolution of a simple stellar population (numbers along the line show ages in Gyr). The dashed line shows the location in the diagram corresponding to an average stellar age of $t_{50} = 800$-Myr based on the empirical relation between rest-frame UVJ colours and t_{50} from Belli et al. (2019). Bottom panel: histograms of the UVJ colour combination translating into t_{50} with the Belli et al. (2019) relation (see Section 2.3), for the full $\log(M_\star/\mathrm{M}_\odot) > 11.1$ UVJ-quiescent sample (top panel) and for the UVJ-quiescent sample excluding 24-μm detections with a probability of being quiescent <0.997 (see the text). In both cases, orange histograms refer to the $H < 22$ subsample. Red symbols show the 10 targets of this work.

quality of the available photometry – for older quiescent galaxies, possibly resulting in a more significant contamination from dusty star-forming sources. To investigate this further, we also highlight in Fig. 3 sources that are detected at 24 μm (≈ 6-μm rest frame) with an S/N > 5 in the Jin et al. (2018) catalogue. The fraction of 24 μm detected sources is higher for old quiescent galaxies than for 'post-starburst' systems. Out of the six oldest UVJ quiescent sources in the sample shown in Fig. 3, five are detected at 24 μm. However, the 24-μm emission could also originate from nuclear activity (see D'Eugenio et al., in preparation for a discussion of 24-μm emissions of our targets), considering the large photometric uncertainties for the oldest sources, this suggests that a significant fraction of the full massive sample considered in Fig. 3 might be star-forming contaminants. We thus re-estimate the fraction of post-starburst galaxies excluding all galaxies that are both 24 μm detected and UVJ quiescent with a probability lower than $p(\mathrm{UVJ\text{-}Q}) = 0.997$ (3σ) and find that (65 ± 10) per cent of the full $\log(M_\star/\mathrm{M}_\odot) > 11.1$ sample have $t_{50} < 800$ Myr compared to (77 ± 10) per cent of the galaxies of the full sample with $H < 22$. Although some of the 24-μm detections could be due to nuclear activity, the distribution of 24-μm detections across the UVJ plane and the estimated uncertainties in UVJ colours strongly suggest that a possibly significant fraction of the oldest quiescent galaxies are actually contaminants, and that the impact of the $H < 22$ selection on the age distribution of our target sample is smaller than would be suggested by face-value comparison of UVJ colours alone. Indeed, the independent estimate of the selection bias for this sample presented in D'Eugenio et al. (2020) consistently concludes that our target sample is representative of $\gtrsim 70$ per cent of the overall quiescent population in the probed mass and redshift range. A specific – and currently very expensive – follow-up of a sample of the highest M/L ratio candidates would be necessary to conclusively address the picture of the potentially oldest massive quiescent galaxies at this redshift.

3 MORPHOLOGICAL ANALYSIS

We investigate the morphological properties of our targets by means of parametric modelling of the surface brightness distribution in the *F*160*W*-band images. For each target, we have three to five dithered observations with total exposure times ranging from 980 to 1130 s at an observed wavelength of ≈ 16 000 Å. We reduce the preprocessed flat-fielded single exposures retrieved from the STScI archive in two different ways to investigate the robustness and sensitivity of the fit results to the reduction procedure. For the first reduction, we use DRIZZLEPAC release 2.2.6[5] to subtract the background, remove cosmic rays and for each source use a square kernel to drizzle the exposures to a pixel scale of 0.06 arcsec before median stacking them to the final image. Each image covers an area of ≈ 4.8 arcmin^2. For reference, the estimated 90 per cent point-source completeness of the images is ≈ 26.7 mag.

We use SEXTRACTOR (Bertin & Arnouts 1996) to detect sources in the *F*160*W*-band images. We select point-like sources in each image by means of a magnitude (MAG_AUTO) versus half-light radius (FLUX_RADIUS 50 per cent) diagram. The comparison of point-like sources across the images of the 10 different fields and at different positions on the detector suggests that the point spread function (PSF) is relatively stable with no significant variations for the purposes relevant to this work. This allows us to create a single PSF by stacking high S/N ($H \lesssim 21$) point-like sources from all 10 fields with SWARP (Bertin et al. 2002), improving the S/N of the model. To estimate the effect of possible systematics of the PSF modelling on our results (see discussion in Section 3.1), we vary the point-like source selection criteria to create a set of PSFs from our observations. We also compare these PSFs with a synthetic model, created with TIINYTIM, and with the hybrid model from van der Wel et al. (2014). Both are more peaked than our models and, if fitted to point sources in our images, subtract systematically too much flux in the centre, while our PSF models do not cause systematic features in the residuals, confirming that they are appropriate descriptions of the PSF of our images. A possible reason why our PSF is less sharp is the low number (3–5) of dithered exposures per target.

For the second reduction, we use the GRIZLI pipeline[6] to produce science ready images from the single exposures, detect sources, and create a PSF for each image. The science images produced with the GRIZLI pipeline are slightly sharper and have less residual cosmic rays compared to the images from the former procedure. None the less, the results of our analysis are largely independent of the reduction method as discussed in detail later in this section. Cutouts of all targets are shown in Fig. 4.

We use GALFIT (Peng et al. 2002, 2010a) to fit PSF convolved Sérsic (1963, 1968) profiles to the *F*160*W*-band images of the sources from both reductions. We create uncertainty maps by quadratically adding Poisson source noise to the background root mean square (RMS), estimated in 9×9 arcsec^2 boxes across the images. We fit sources in cutouts with a sidelength of 9 arcsec, allowing GALFIT to fit a constant background simultaneously with the Sérsic profiles. Estimating the local background and subtracting it from the image, rather than fitting it, has no significant impact on the estimated parameters for the targets considered here. For cutouts containing multiple galaxies, we simultaneously fit Sérsic profiles for all sources. We do not set prior constraints on any of the fit parameters (position, magnitude, effective radius r_e, Sérsic index n, axial ratio q, position angle). Starting values for the fitted parameters are estimated based on the SEXTRACTOR output except for the Sérsic index for which we use a starting value of 1. We verified that varying the initial parameters in a reasonable range has no impact on the results for our targets. Estimated effective radii and axis ratios are stable against the use of the different reductions and corresponding PSFs, being entirely consistent within the estimated uncertainties with no systematic biases. Sérsic indices are systematically lower by 20 per cent in the GRIZLI reduction. In the following, we always refer to the measurements obtained on the GRIZLI reductions unless otherwise stated. We stress again that all conclusions would be unchanged if referring to the other reduction, and that all results would be fully consistent except Sérsic indices which would be on average larger, thus resulting in even stronger conclusions on the typically high Sérsic indices of these sources as discussed in Section 4. The target images, best-fitting models and corresponding residuals, together with *HST* ACS *F*814*W* imaging (Koekemoer et al. 2007; Massey et al. 2010, only available for IDs 4-10), are shown in Fig. 4. The resulting profile parameters are presented in Table 1. Radii in this paper refer to effective radii along the semimajor axis.

Some of our targets have faint close neighbouring sources with unknown redshifts (see also Stockmann et al. 2020), in particular, ID 7 has a very close neighbour ($d \approx 0.8$ arcsec, corresponding to 6 kpc if at the target redshift). With the available data, we do not see evidence of interaction between these sources. Furthermore, after

[5] https://github.com/spacetelescope/drizzlepac/blob/master/doc/source/index.rst

[6] https://github.com/gbrammer/grizli/

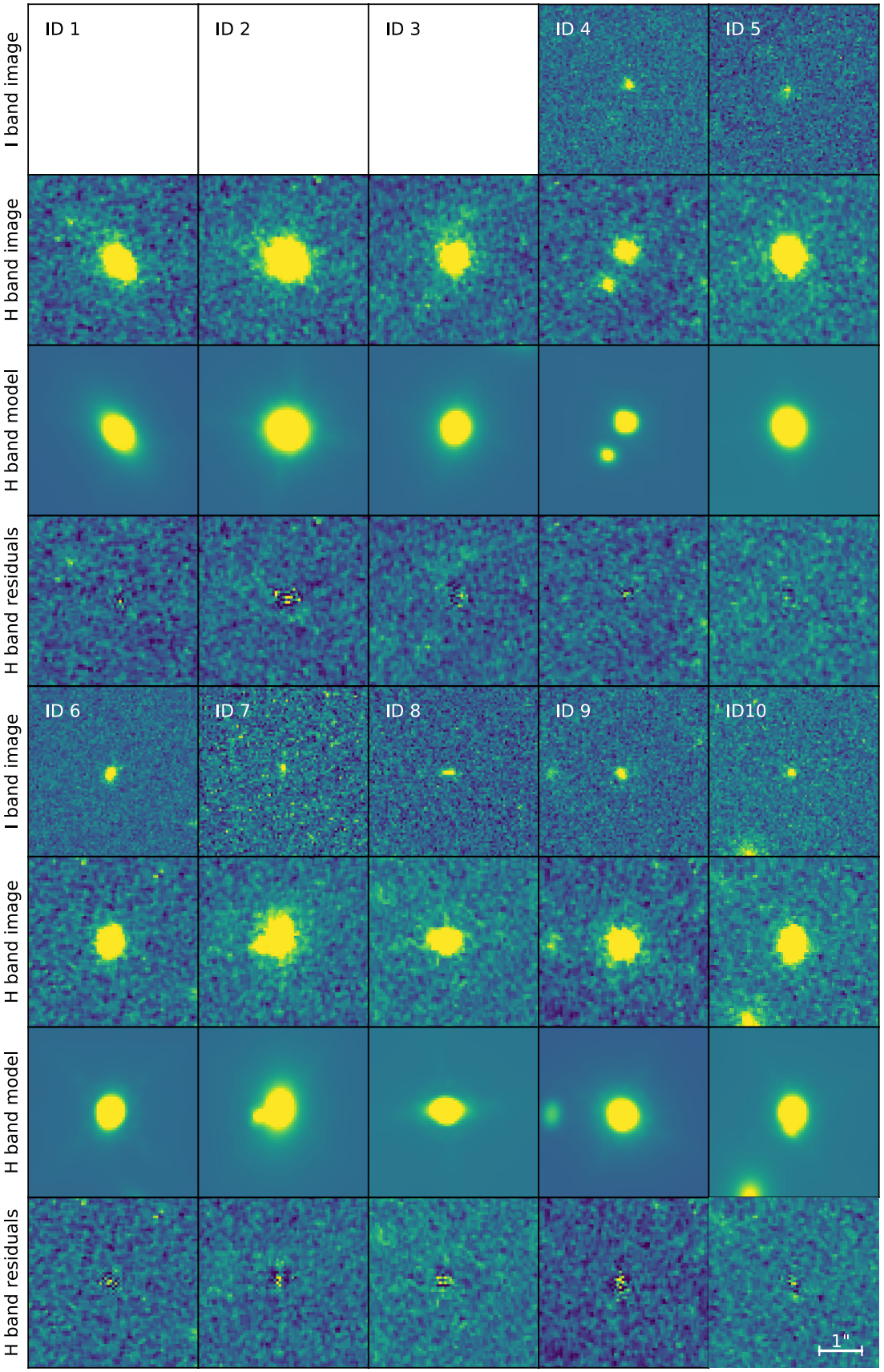

Figure 4. *HST F*160*W* (*H*-band) images of the 10 observed targets, the best-fitting models (see Section 3) and the corresponding residuals. For IDs 4–10, *F*814*W* (*I*-band) imaging is available and also shown. The cutouts have a size of 4 arcsec (≈ 32 kpc) by side.

subtracting the best-fitting model the residuals of a single Sérsic profile fit for IDs 2 and 10 also show faint neighbours close to the targets (0.56 and 0.1 arcsec, respectively, corresponding to 4.6 and 1.0 kpc if they are at the targets redshift). For these sources, an additional Sérsic component is thus used in the following to simultaneously model the faint neighbours. The models and residuals in Fig. 4 show the modelling accounting for these sources. Estimated parameter values (r_e, n, q) change by $\lesssim 10$ per cent, except for the estimated effective radius of ID 10 which decreases by 19 per cent and its axial ratio which increases by 25 per cent.

For several sources, namely IDs 6, 7, 8, 9, and 10 the residual images show a central residual. For ID 8, about 4 per cent of the pixels associated with the source have a significance of more than 3σ in the residuals. For all other sources, this fraction is $\lesssim 2$ per cent. This is also shown in the Appendix in Fig. B1. We verified that this is not a PSF effect by fitting the targets with different PSF models, including the more peaked h TIINYTIM and van der Wel et al. (2014) models and examining a larger number of residuals from fits of other sources in the images. We also verified that adding an additional point-like component to the fit does not produce an appreciable improvement on the residuals for any of these targets. We note that three of these sources (IDs 6, 7, and 10) are detected in both the Jin et al. (2018) 24-μm catalogue and the Marchesi et al. (2016) catalogue of X-ray sources. Therefore, some level of star formation (in the galaxy centre) and/or nuclear activity might possibly cause the central excess. By comparing the 24-μm and X-ray luminosities, D'Eugenio et al. (in preparation) suggest that in at least two out of these three sources the IR- and X-ray emission is likely AGN dominated. The flux of the central residual for these sources is smaller than 10 per cent of the image flux in any pixel.

3.1 Uncertainties

The precision and accuracy of our measurements are limited by noise and by uncertainties in the PSF model, which are analysed in the following.

To estimate the impact of PSF uncertainties on the parameter estimates, we create on empty images without noise artificial sources with an effective radius ranging from 0.03 to 0.72 arcsec (0.2 kpc $\leq r_e \leq$ 5.8 kpc at $z = 2.75$) and with Sérsic indices from 0.5 to 8, spanning a reasonably wide range of Sérsic parameters for quiescent galaxies in the mass and redshift range of our targets. We then fit these artificial sources with the same procedure used for our targets, using different PSF models (see Section 3) for the convolution in the creation and in the fitting process. The deviations of the retrieved parameters (effective radius, Sérsic index, axial ratio) from the input ones are generally smaller than 5 per cent; for very small radii ($r_e \lesssim 0.05$ arcsec or $r_e \lesssim 0.4$ kpc in the probed redshift range) they can exceed 10 per cent. All of our targets except one (ID 4, $r_e \approx 0.07$ arcsec) are significantly larger than this size. The uncertainties on the PSF model are therefore expected to have a subdominant impact on our results.

To investigate the uncertainties due to noise, we create Sérsic models with Poisson noise in different empty areas of the observed images. We first create sources with the approximate magnitude of our targets, $H \approx 22$, and effective radius and Sérsic index in the same range as discussed above for the evaluation of systematic uncertainties. In this case, we use the same PSF for the creation of the artificial sources and for their modelling.

In Fig. 5, we show the deviation of the retrieved best-fitting values for the Sérsic index, effective radius and axial ratio with respect to the input values. The estimated scatter in r_e and n is of the order of 10 per cent but depends on the actual values of r_e and n. The scatter of the axial ratio is of the order of 5 per cent except for very small ratios of ≈0.2 where it reaches ≈ 10 per cent. Models with larger Sérsic indices have generally larger uncertainties on the retrieved parameters and the Sérsic index tends to be underestimated, because of the extended tails and low S/N in the outskirts (see also e.g. Marleau & Simard 1998; Pignatelli, Fasano & Cassata 2006; Sargent et al. 2007; Pannella et al. 2009). This underestimation, $\Delta n \approx 5$ per cent for $n \approx 4$ and $\Delta n \approx 10$ per cent for very high Sérsic indices of $n \approx 8$, is about four times smaller than the statistical uncertainty. Sources with very small radii, close to the resolution limit, are affected by larger uncertainties, as well as those with very large radii because of the lower S/N per pixel at fixed magnitude. Sources with $n \approx 1$ and with $r_e \gtrsim 0.15$ arcsec have uncertainties $\sigma_{r_e} < 5$ per cent and $\sigma_n < 10$ per cent. The effective radius of models with large n and r_e is also affected by a systematic underestimation of r_e. For large sources with $r_e = 0.3$ arcsec and $n \approx 4$, r_e is underestimated by ≈ 5 per cent, for $n = 8$ by ≈ 10 per cent. These systematics are small compared to the statistical uncertainties for the same models.

To properly estimate statistical uncertainties on the measured Sérsic parameters for each target, we then create artificial sources with parameters in a 10 per cent range around the best-fitting models, motivated by the previous results, adding as usual Poisson noise. We add them to different empty areas of the observed images and fit them again with the same PSF as used for the creation. Since the estimated systematics are typically small compared to the statistical uncertainties we do not apply any correction for the described systematics in the following.

4 RESULTS AND DISCUSSION

4.1 Broad structural properties

Sérsic indices of our sources range from 1.2 to 6.3 with median statistical uncertainties of 16 per cent except for source ID 9, which formally has a best-fitting Sérsic index of 12. The median Sérsic index of all targets is $4.5^{+0.3}_{-1.4}$. The high Sérsic index of ID 9 has no strong influence on the median Sérsic index, excluding it leads to a median of $n = 4.3^{+0.5}_{-1.2}$. The low Sérsic index $n = 1.2^{+0.1}_{-0.1}$ of ID 5 is also reinforced by the diffuse appearance in the $F814W$ image compared to the other sources (see Fig. 4).

In Fig. 6, we show the median Sérsic index of massive quiescent galaxy samples as a function of redshift. For comparison to lower redshift quiescent and star-forming galaxies, we show median Sérsic indices of galaxies with stellar masses of $11 < \log(M_\star/M_\odot) < 11.5$ from the morphological analysis of van der Wel et al. (2014). At all redshifts, median Sérsic indices of quiescent galaxies are significantly larger than those of star-forming galaxies. The median Sérsic indices of quiescent galaxies from the van der Wel et al. (2014) sample decrease from $n = 4.5^{+0.2}_{-0.3}$ at $z = 0.4$ to $n = 3.3^{+1.0}_{-0.4}$ at $z = 2.7$. Our results are consistent with no significant evolution in the median Sérsic index of quiescent galaxies up to $z \approx 3$, in agreement with other studies from Patel et al. (2017), Mowla et al. (2019), Marsan et al. (2019), Stockmann et al. (2020), and Esdaile et al. (2020), although some investigations have reported lower Sérsic indices at $z \gtrsim 1.5$ (e.g. van Dokkum et al. 2008; van der Wel et al. 2011).

Axis ratios of our targets range between 0.33 and 0.96 with a median of $0.73^{+0.06}_{-0.12}$. The only source with $q < 0.5$ is ID 8 having $q = 0.33^{+0.03}_{-0.03}$ in spite of a high Sérsic index of $4.3^{+0.8}_{-0.7}$. The source also appears rather flat in the $F814W$ image, which could be explained

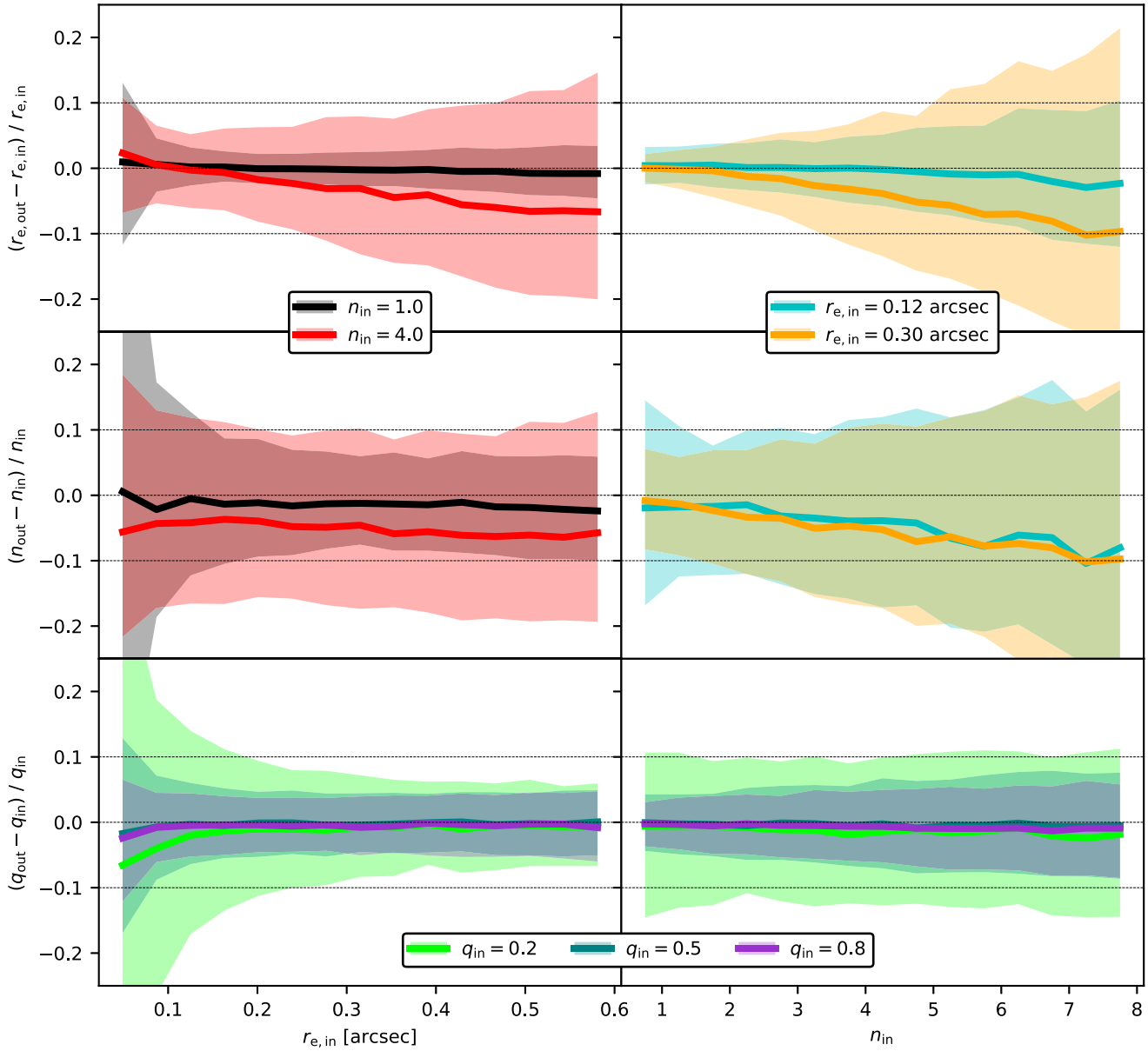

Figure 5. Relative uncertainties on the estimated effective radius, Sérsic index and axial ratio from the simulations (Section 3.1), defined as $(p_{out} - p_{in})/p_{in}$ where p_{in} is the input parameter value of the artificial source and p_{out} the retrieved value from the fit. Left- and right-hand panels show the deviation of the retrieved versus input parameters as a function of effective radius and Sérsic index of the source, respectively. The solid lines represent the median deviation from the input value, the shaded areas show the scatter estimated from the $16 - 84$percentile range. We show as an example in the upper two panels results for a Sérsic index of 1 (black) and 4 (red) in the left-hand panels and for $r_e = 0.06$ arcsec (cyan) and $r_e = 0.12$ arcsec (orange) in the right panels, both with axis ratios $0.5 \leq q_{in} \leq 1$. In the lower panels, we show uncertainties on the retrieved axial ratio for input axis ratios of 0.2, 0.5, and 0.8, as indicated.

by a combination of an older bulge with redder colours and a younger and bluer disc that is seen edge-on.

Hill et al. (2019) find a redshift and stellar mass-independent linear relation between Sérsic index and apparent axial ratio with a slope of $dq/dn = 0.062$, yielding an axial ratio of 0.71 at our median Sérsic index of 4.5, in perfect agreement with our measurement and reinforcing our conclusions on the generally high Sérsic indices of these sources.

In Fig. 6, we also show the median axial ratio of massive quiescent galaxies as a function of redshift from Hill et al. (2019), comparing with measurements from our work and other studies at $z \gtrsim 1.5$. While at $z < 2$ median axis ratios of quiescent galaxy samples are larger than those of star-forming galaxies, at $z \gtrsim 2$ no clear difference can be seen, although uncertainties become large and quiescent galaxy sample contamination from star-forming sources is likely more significant. Our measurement of the average axial ratio is in agreement with typical axis ratios of quiescent galaxies at low redshift (e.g. Holden et al. 2012; Hill et al. 2019). Consistent with our measurements, Patel et al. (2017) and Marsan et al. (2019) also find high axis ratios and Sérsic indices for massive ($\log(M_\star/M_\odot) > 11.26$) quiescent galaxies at $z \approx 2.6$ as well as Esdaile et al. (2020) at $z \approx 3.3$, suggesting that already at $z \approx 3$ a large fraction of quiescent galaxies are bulge dominated. On the other hand, van Dokkum et al. (2008) investigated morphologies of nine spectroscopically confirmed massive ($\log(M_\star/M_\odot) > 11.1$) quiescent galaxies at $z \approx 2.3$. The median Sérsic index of their sample is $2.3^{+0.5}_{-0.0}$ and the

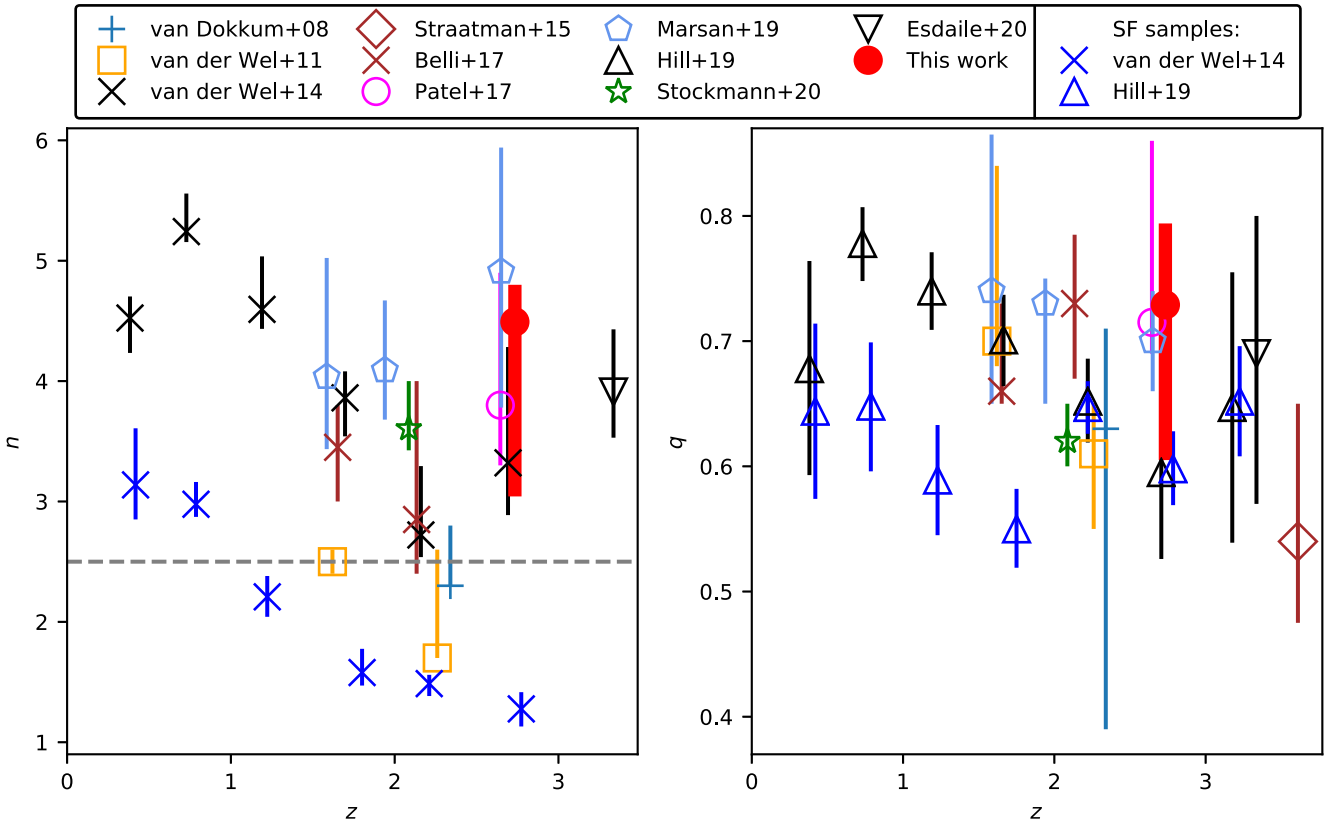

Figure 6. Median Sérsic indices (left-hand panel) and axis ratios (right-hand panel) of massive quiescent galaxy samples as a function of redshift. For comparison, we also show the evolution of the Sérsic index for star-forming galaxies from van der Wel et al. (2014, left-hand panel) and of the axis ratios from the same sample analysed by Hill et al. (2019, right). Sources from van der Wel et al. (2014) have stellar masses of $11 < \log(M_\star/\mathrm{M}_\odot) < 11.5$ and sources from Hill et al. (2019) have $\log(M_\star/\mathrm{M}_\odot) > 11$. Results from other works are reported as published, with no further mass selection applied. If the same mass selection is applied to the other samples, measurements change within the uncertainties with no systematics. Uncertainties for the samples of van der Wel et al. (2014) and Hill et al. (2019) are taken from the papers while we use bootstrapping for the calculation for the other samples.

median apparent axial ratio $0.63^{+0.08}_{-0.24}$. In agreement with these results, van der Wel et al. (2011) analysed a colour selected sample of 14 massive ($\log(M_\star/\mathrm{M}_\odot) > 10.8$) quiescent galaxies at $1.5 < z < 2.5$ finding a median Sérsic index of $2.45^{+0.15}_{-0.40}$ and a median axial ratio of $0.67^{+0.10}_{-0.06}$. Belli et al. (2017) find a median Sérsic index of $3.25^{+0.45}_{-0.30}$ and a median axial ratio of $0.69^{+0.05}_{-0.04}$ in the same redshift range. Hill et al. (2019) investigated the median flattening of galaxies in the redshift range $0.2 < z < 4.0$, based on the structural analysis from van der Wel et al. (2014) and also find that, for quiescent galaxies with $\log(M_\star/\mathrm{M}_\odot) > 11.0$, the apparent axial ratio decreases to $q = 0.60 \pm 0.07$ at $z = 2.7$. In contrast to results from this work and other previous investigations as discussed above, these studies suggest that massive quiescent galaxies at high redshift are flatter than low-redshift counterparts, with a large fraction of disc-dominated systems. Considering at face value our results on both Sérsic indices and axis ratios, our measurements do not lend support to this picture. None the less, concerning axis ratios, we note that given the large statistical uncertainties our median axial ratio is still consistent with results from van Dokkum et al. (2008), van der Wel et al. (2011), and Hill et al. (2019). Furthermore, one of our targets has $q < 0.5$ and three have $n < 3$, suggesting that some sources in our sample might indeed be disc dominated or have a significant disc component. We note that differences in results and conclusions from the studies discussed above may partly derive from different sample selection criteria [in particular, van der Wel et al. (2011) and Hill et al. (2019) rely on different flavours of photometrically selected samples].

4.2 The mass–size relation

The estimated effective radii of the galaxies in our sample are between 0.07 and 0.57 arcsec, corresponding to physical sizes of 0.5 to 4.5 kpc at rest-frame wavelengths from 3800 to 4700 Å. Fitting ID 9 with the Sérsic index fixed to a typical value for bulge-dominated systems of $n = 4$ (close to the sample median of $n = 4.5^{+0.3}_{-1.4}$) leads to a decrease of the estimated effective radius by 50 per cent.

Because of – mostly negative – colour gradients of galaxies (Szomoru et al. 2011; Wuyts et al. 2012; van der Wel et al. 2014; Suess et al. 2019a, b), galaxy sizes inferred from light profiles depend on the probed wavelength with sizes being larger at shorter wavelengths. For a proper comparison with previous works, we convert all measured sizes to the same rest-frame wavelength of 5000 Å, adopting the correction appropriate for quiescent galaxies from van der Wel et al. (2014):

$$r_\mathrm{e}(5000\,\text{Å}) = r_\mathrm{e}(\lambda_\mathrm{obs}) \left(\frac{1+z}{\lambda_\mathrm{obs}/5000\,\text{Å}} \right)^{\frac{\Delta \log r_\mathrm{e}}{\Delta \log \lambda}} \tag{3}$$

with

$$\frac{\Delta \log r_\mathrm{e}}{\Delta \log \lambda} = -0.35 + 0.12z - 0.25 \log \left(\frac{M_\star}{10^{10}\,\mathrm{M}_\odot} \right), \tag{4}$$

where λ_obs is the observed wavelength of 16000 Å. This results in a very small correction decreasing the measured sizes of our targets by about 5 per cent; the final sizes adopted in the following are between

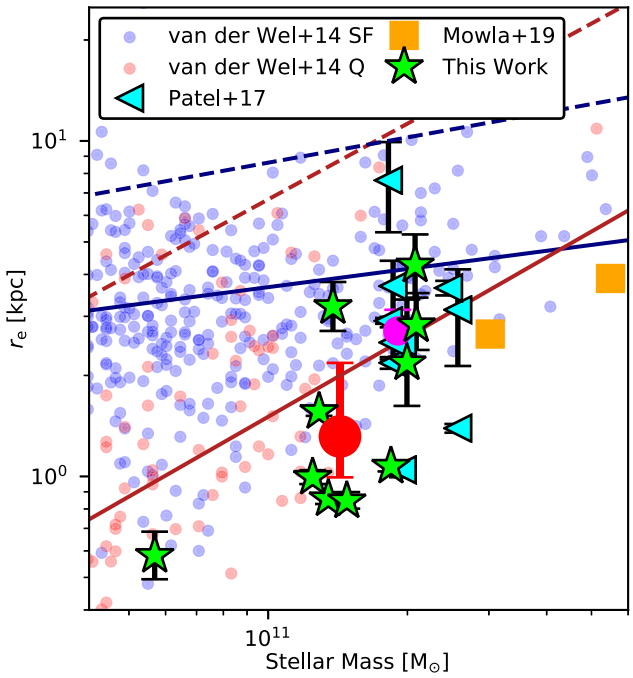

Figure 7. Estimated effective radii as a function of stellar mass. Sizes of individual galaxies from this work (stars), Patel et al. (2017, triangles) and Mowla et al. (2019, squares) are scaled to a pivot redshift of 2.75 (see Section 4.2). The red and purple dots show the median size and median mass from our sample and from Patel et al. (2017), respectively. Red (blue) dots are quiescent (star-forming) galaxies with $2.5 < z < 3.0$ from van der Wel et al. (2014), with red and blue solid lines showing the corresponding best-fitting mass–size relations. For comparison, the best-fitting relations for $z < 0.5$ from the same work are shown with dashed lines.

0.5 and 4.4 kpc with a median size of $1.4^{+0.9}_{-0.2}$ kpc. Uncertainties on the median are obtained by bootstrapping. These sizes are reported in Table 1.

In Fig. 7, we compare our results for the stellar mass versus size relation of quiescent galaxies at $z \approx 3$ with previous measurements of photometrically (UVJ) selected quiescent and star-forming galaxies from van der Wel et al. (2014), as well as quiescent galaxies from Patel et al. (2017) and Mowla et al. (2019), in the same redshift range. To ensure that no systematics on stellar masses affect our comparison with size estimates from van der Wel et al. (2014), we fit Sérsic profiles to all galaxies in our fields and estimate their stellar masses as explained in Section 2.2. Their mass–size relations are consistent with results from van der Wel et al. (2014) at the corresponding redshifts, indicating no significant systematics between the mass and size measurements in the two studies. For a more proper comparison of the mass–size relation within the probed $2.4 < z < 3.2$ range, and given the small sample size of the plotted samples from this work, Patel et al. (2017) and Mowla et al. (2019), we scale individual sizes for galaxies from these samples to a pivot redshift of 2.75 using the size evolution dependence on the Hubble parameter from van der Wel et al. (2014). This scaling leads to a maximum decrease of sizes by 17 per cent at the lowest redshift $z = 2.4$ and a maximum increase by 25 per cent at the highest redshift $z = 3.2$.[7] The median sizes of our, Patel et al. (2017) and Mowla et al. (2019) samples decrease by ≈7, 8, and 10 per cent, respectively.

Our sources specifically probe the mass–size relation at the highest stellar masses, for the first time with a statistical, homogeneously analysed, spectroscopically confirmed quiescent galaxy sample at this redshift. Our measurements thus extend towards the highest masses the determination of the mass–size relation of quiescent sources at $z \approx 3$, which in deep fields is typically dominated by lower mass galaxies because of the intrinsically low number density of very massive quiescent sources. Our measurement of the median quiescent galaxy size at the tip of the mass-size relation ($\log(M_\star/M_\odot) \gtrsim 11$) at $z \approx 3$ is none the less consistent with the relation measured in van der Wel et al. (2014).

4.2.1 *Central stellar mass densities*

Although with the available data we can only probe the projected surface brightness distribution of our targets in the *F160W* band, we attempt a conversion of the observed light profile to a stellar mass density profile, to estimate central densities of these galaxies for the purpose of comparing with other similar studies. This conversion relies, in particular, on the assumption that the observed *F160W* light traces stellar mass across the galaxy: we stress that, also given the rest-frame wavelength probed by the *F160W* imaging at the redshift of these sources, this assumption has, in fact, significant limitations, which are neglected in the following calculations.

We deproject the observed surface brightness distributions of our targets and calculate their central densities within 1 kpc following the procedure in Whitaker et al. (2017). Briefly, we calculate a circularized density profile from the best-fitting structural parameters derived in Section 3 by performing an Abel transform as described in Bezanson et al. (2009). In the assumption that light traces mass the central stellar mass density within 1 kpc is then given by

$$\rho_1 = \frac{\int_0^{1\,\mathrm{kpc}} \rho(r) r^2 \mathrm{d}r}{\int_0^{\infty} \rho(r) r^2 \mathrm{d}r} \frac{M_\star}{\frac{4}{3}\pi (1\,\mathrm{kpc})^3}, \quad (5)$$

where $\rho(r)$ is the spherical density profile as a function of radius. We estimate uncertainties coming from the measurement of structural parameters by perturbing r_e, n, and q within their estimated uncertainties and recalculating the central densities 1000 times. These uncertainties are at most 0.09 dex; the uncertainties on the central densities are therefore dominated by the uncertainties on the stellar mass estimates (see Section 2.2) as well as by the limitations of the adopted assumptions to convert the observed surface brightness distribution to a stellar mass density profile. The central densities of the targets are $9.8 \lesssim \log(\rho_1\,\mathrm{kpc}^3/M_\odot) \lesssim 10.4$ with a median of $\log(\rho_1\,\mathrm{kpc}^3/M_\odot) = 10.1 \pm 0.1$. Such central densities translate in circular velocities at $r = 1$ kpc of $330\,\mathrm{km\,s^{-1}} \lesssim v_1 \lesssim 640\,\mathrm{km\,s^{-1}}$ (median 480^{+50}_{-60} km s^{-1}), as obtained by $v_1 = \sqrt{\frac{4\pi}{3}(1\,\mathrm{kpc})^2 \rho_1 G}$, where G is the gravitational constant, by balancing gravitational and centrifugal forces (see e.g. Whitaker et al. 2017).

These high inferred central densities – and implied circular velocities – are in line with previous determinations for high-redshift massive, quiescent sources (e.g. van Dokkum et al. 2014; Whitaker et al. 2017; Mowla et al. 2019).

4.3 Size evolution

To constrain the redshift evolution of massive quiescent galaxies at early times, we compare sizes from our work with measurements from van der Wel et al. (2014), Straatman et al. (2015), Patel et al.

[7] If rather than using the size evolution dependence on the Hubble parameter we use the dependence on $1 + z$, always from van der Wel et al. (2014), the maximum increase (decrease) is 14 per cent (20 per cent) which does not impact the results of this analysis.

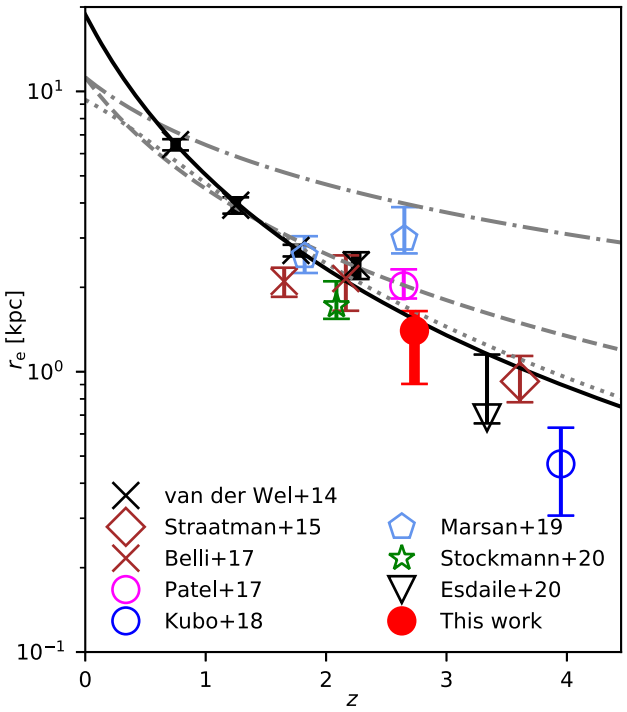

Figure 8. Median effective radii of quiescent galaxy samples with $10.6 < \log(M_\star/\mathrm{M}_\odot) < 11.8$ as function of redshift. Sizes are scaled to a pivot mass of $\log(M_\star/\mathrm{M}_\odot) = 11.1$ as explained in Section 4.2. The Marsan et al. (2019) (Belli et al. 2017) sample is split into two subsamples with $1.5 < z < 2.1$ ($1.5 < z < 1.9$) and $2.6 < z < 3$ ($2.1 < z < 2.4$). The dotted line shows the best-fitting model from van der Wel et al. (2014) assuming a $r_e \propto h(r)^\beta$ relation, while the dashed and solid lines show best-fitting models assuming $r_e \propto (1+z)^\beta$ from van der Wel et al. (2014) and Kubo et al. (2018), respectively. The dashed–dotted line shows for reference the median evolution for star-forming galaxies from van der Wel et al. (2014).

(2017), Kubo et al. (2017, 2018), Belli et al. (2017), Marsan et al. (2019), Stockmann et al. (2020), and Esdaile et al. (2020) in Fig. 8. Sources from Belli et al. (2017), Stockmann et al. (2020), Esdaile et al. (2020), and from this work are spectroscopically confirmed quiescent galaxies, while the other studies we compare with rely on photometrically selected quiescent sources.

Sources from van der Wel et al. (2014) in this figure have masses between $11.0 < \log(M_\star/\mathrm{M}_\odot) < 11.5$ with a median mass of $\log(M_\star/\mathrm{M}_\odot) \approx 11.1$ in each redshift bin. We use the redshift independent slope of the mass–size relation from van der Wel et al. (2014) ($\mathrm{dlog}(r_e)/\mathrm{dlog}(M_\star) = 0.7$) to scale sizes of individual galaxies of all other samples (with masses in the range $10.6 < \log(M_\star/\mathrm{M}_\odot) < 11.8$) to $\log(M_\star/\mathrm{M}_\odot) = 11.1$. We then calculate median sizes and uncertainties for all samples by bootstrapping. The aforementioned scaling to a common mass of $\log(M_\star/\mathrm{M}_\odot) = 11.1$ has a very limited impact on our measurement of the median size of our sample being $1.4^{+0.2}_{-0.5}$ kpc at the pivot mass, basically affecting only the uncertainties. For the least massive sample (Straatman et al. 2015), this correction increases the median size by ≈ 0.1 dex, while for the most massive sample (Marsan et al. 2019) the median size decreases by ≈ 0.25 dex. For the study of Kubo et al. (2018), the size and uncertainty of the stack is shown. All points are plotted at the median redshift of the respective sample.[8]

[8]We do not use any size versus redshift relation to scale individual galaxy sizes to the median redshift before calculating the median size. However, if any of the relations shown in Fig. 8 is adopted to do so, the impact on the median size is smaller than 0.03 dex for any sample and does not affect our discussion.

Our measurements are in line with previous determinations indicating that sizes of quiescent galaxies at fixed stellar mass have increased by nearly one order of magnitude since $z = 3$. The different models from van der Wel et al. (2014) and Kubo et al. (2018), that parametrize the redshift evolution of the mass–size relation either as a function of the Hubble parameter, that is related to halo properties, or as a function of the scale factor, differ by a maximum of 0.1 dex at the median target redshift of 2.73. The median of our size measurements is consistent with the van der Wel et al. (2014) evolution as a function of $h(z)$ and the Kubo et al. (2018) evolution as a function of $1+z$. Our measurement is 2σ smaller than expected from the van der Wel et al. (2014) evolution as a function of $1+z$, which suggests, together with the higher redshift measurements by Straatman et al. (2015), Kubo et al. (2018), and Esdaile et al. (2020) that size evolution is steeper than in this relation.

At redshifts closest to our measurements, the median sizes from Patel et al. (2017) and especially from the highest redshift sources in Marsan et al. (2019) tend to be larger than our estimate as well as than extrapolations from most of the other high-redshift measurements discussed above. Both measurements are largely based on the same sample of galaxies with a median mass of $\log(M_\star/\mathrm{M}_\odot) \approx 11.3$. Based on these measurements, Patel et al. (2017) and Marsan et al. (2019) suggest that very massive galaxies with $\log(M_\star/\mathrm{M}_\odot) > 11.25$ are systematically larger than expected from the mass–size relation determined at lower masses, and thus that the size evolution factor may be different at the highest masses. On the other hand, results from the very massive samples with $\log(M_\star/\mathrm{M}_\odot) \approx 11.5$ from Marsan et al. (2019) at $z \approx 1.8$ and from Stockmann et al. (2020) at $z \approx 2$ – the latter likely affected by minimal contamination from star-forming sources, due to spectroscopic confirmation – do not seem to support such a scenario. From the four most massive galaxies in our sample with $\log(M_\star/\mathrm{M}_\odot) > 11.25$, we see excellent agreement with the extrapolation of the van der Wel et al. (2014) mass–size relation, although the statistics are very limited due to the small sample size.

5 SUMMARY AND CONCLUSIONS

We have analysed structural properties of a first sizeable sample of spectroscopically confirmed, massive, quiescent galaxies at $z \approx 3$ (D'Eugenio et al. 2020). Due to the rarity of these objects, we relied on targeted *HST*/WFC3 imaging of 10 robust candidates.

We estimate structural properties by fitting Sérsic profiles to the *F*160*W* images and obtain half-light radii of about $r_e \approx 1$ kpc at stellar masses of $\log(M_\star/\mathrm{M}_\odot) \approx 11.2$, in agreement with photometrically selected samples at this redshift. The comparison with sizes of massive quiescent galaxies at different redshifts shows substantial agreement with the expected evolution of the mass–size relation of quiescent galaxies as determined in previous work, pointing towards a size evolution factor at the fixed stellar mass of almost a factor of 10 from $z \approx 3$ to today.

Although our observations are consistent with a fraction of our sample being made of disc-dominated galaxies, and a larger sample would be needed to better quantify the prevalence of such sources, our measurements of both axis ratios and Sérsic indices suggest that massive, quiescent galaxies are already largely bulge dominated at $z \approx 3$. Based on a sample of massive galaxies in the redshift

range $0.5 < z < 3$, Barro et al. (2017) find a redshift and mass-independent relation between the offset of a galaxy's SFR from the main sequence and the central mass density, that strongly correlates with Sérsic index. This implies that star-forming galaxies first grow inside out while increasing the radius and the central mass density, followed by a phase of enhanced bulge growth that increases the Sérsic index. Star formation is then suppressed and galaxies become quiescent. This picture is also in line with other studies by e.g. Lang et al. (2014), van Dokkum et al. (2014), van Dokkum et al. (2015), Gobat et al. (2017), Whitaker et al. (2017), and Gómez-Guijarro et al. (2019) and is consistent with the large fraction of bulge-dominated systems in our sample. The presence of bulge-dominated, quiescent galaxies already at $z \approx 3$ constraints the time-scales of quenching and of morphological transformations at early times. Although merging is believed to be a critical process to explain the size and structural evolution of quiescent high-redshift progenitors into local massive ellipticals, the combination of young ages and dense, compact structures of the most distant quiescent galaxies such as those studied here suggest different mechanisms for the fast formation of the stellar core. Matching number densities and high SFRs at high redshift have suggested an evolutionary path linking the intense bursts of star formation in high-redshift sub-mm galaxies to the high stellar densities and old stellar populations of massive elliptical galaxies at lower redshifts down to the nearby universe (e.g. Lilly et al. 1999; Genzel et al. 2003; Cimatti et al. 2008; Tacconi et al. 2008; Simpson et al. 2014, 2017), including, in particular, the most distant massive, dusty star-forming galaxies being likely progenitors of (at least some of) the first massive, compact quiescent galaxies at $z > 2$ (e.g. Toft et al. 2014; Valentino et al. 2020; Forrest et al. 2020a). High-resolution ALMA imaging of z 4-6 dusty, massive, highly star-forming sources confirms the existence of possible star-forming progenitors with already compact morphologies (Oteo et al. 2017; Jin et al. 2019). The majority of bright sub-mm galaxies in simulations (Hopkins et al. 2008; Dekel et al. 2009b; Wellons et al. 2015; Zolotov et al. 2015; Lagos et al. 2020) are experiencing central starbursts driven by two main channels, gas-rich major mergers and disc instabilities, that increase the central density forming a compact remnant. Such remnants may still have discs and disc-dominated kinematics (e.g. Belli et al. 2017; Toft et al. 2017; Newman et al. 2018, and references therein), suggesting that the morphological transformations creating dispersion-supported ellipticals are not necessarily coincident with quenching. The mechanism by which star formation would stop in the compact star-forming progenitors is still unclear, with proposed processes including dynamical heating ('morphological quenching', Martig et al. 2009), stellar and AGN feedback (Hopkins et al. 2006), shock heating (Dekel & Birnboim 2006), cosmological starvation (Feldmann & Mayer 2015), starvation by the circumgalactic medium having too high angular momentum to be accreted by the central galaxy (Peng & Renzini 2020) (see also e.g. Man & Belli 2018, and references therein). Some observations (Gilli et al. 2014; Nelson et al. 2014) have identified possible compact star-forming progenitors suggestive of dense stellar cores in their formation phase (see also Patel et al. 2013; Barro et al. 2013, 2014a, b; Stefanon et al. 2013; Williams et al. 2015) or quenching progenitors suggestive of the transition stage to compact quiescent remnants (Marsan et al. 2015). Recent and upcoming efforts to secure samples of very distant quiescent galaxies and of their immediate progenitors, their observation with state-of-the-art and new instruments to probe their stellar population, gas content, and structural and kinematical properties, and the detailed comparison with state-of-the-art simulations, will soon provide new constraints on the early formation of massive quiescent galaxies.

ACKNOWLEDGEMENTS

We thank the anonymous referee for a constructive report that improved the presentation of this work. PL and VS acknowledge support from the Deutsches Zentrum für Luft- und Raumfahrt Verbundforschung grant 50OR1805. VS and MP acknowledge support from the ERC-StG ClustersXCosmo grant agreement 716762. AC acknowledges support from grant PRIN MIUR 2017 - 20173ML3WW_001. SJ acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities (MICIU) under AYA2017-84061-P, co-financed by FEDER (European Regional Development Funds). This research is based on observations made with the NASA/ESA *Hubble Space Telescope* obtained from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program ID 15229. Partly based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This study is partly based on a K_s-selected catalogue of the COSMOS/UltraVISTA field from Muzzin et al. (2013). The catalogue contains PSF-matched photometry in 30 photometric bands covering the wavelength range 0.15μm → 24μm and includes the available *GALEX* (Martin et al. 2005), CFHT/Subaru (Capak et al. 2007), UltraVISTA (McCracken et al. 2012), S-COSMOS (Sanders et al. 2007), and zCOSMOS (Lilly et al. 2009) data sets.

DATA AVAILABILITY

The data of the *HST* program underlying this paper are available in the *HST* archive.[9] References for additional data are given in the text.

REFERENCES

Almaini O. et al., 2017, MNRAS, 472, 1401
Athanassoula E., Rodionov S. A., Peschken N., Lambert J. C., 2016, ApJ, 821, 90
Barro G. et al., 2013, ApJ, 765, 104
Barro G. et al., 2014a, ApJ, 791, 52
Barro G. et al., 2014b, ApJ, 795, 145
Barro G. et al., 2017, ApJ, 840, 47
Bell E. F. et al., 2006, ApJ, 640, 241
Bell E. F. et al., 2012, ApJ, 753, 167
Belli S., Newman A. B., Ellis R. S., 2014, ApJ, 783, 117
Belli S., Newman A. B., Ellis R. S., 2015, ApJ, 799, 206
Belli S., Newman A. B., Ellis R. S., 2017, ApJ, 834, 18
Belli S., Newman A. B., Ellis R. S., 2019, ApJ, 874, 17
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Bohlender D. A., Durand D., Handley T. H., eds, ASP Conf. Ser. Vol. 281, Astronomical Data Analysis Software and Systems XI, Astron. Soc. Pac., San Francisco, p. 228
Bezanson R., van Dokkum P., Franx M., 2012, ApJ, 760, 62
Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, ApJ, 697, 1290
Bezanson R. et al., 2018, ApJ, 858, 60
Bédorf J., Portegies Zwart S., 2013, MNRAS, 431, 767
Birnboim Y., Dekel A., 2003, MNRAS, 345, 349
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brammer G. B. et al., 2011, ApJ, 739, 24

[9] https://archive.stsci.edu/hst/

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buitrago F.et al., 2008, ApJ, 687, L61
Bundy K. et al., 2010, ApJ, 719, 1969
Burkert A. et al., 2010, ApJ, 725, 2324
Calzetti D., 2001, PASP, 113, 1449
Capak P. et al., 2007, ApJS, 172, 99
Cappellari M., 2016, ARA&A, 54, 597
Carollo C. M. et al., 2013, ApJ, 773, 112
Cassata P. et al., 2013, ApJ, 775, 106
Chabrier G., 2003, PASP, 115, 763
Chang Y.-Y. et al., 2013, ApJ, 773, 149
Cimatti A., Nipoti C., Cassata P., 2012, MNRAS, 422, L62
Cimatti A. et al., 2004, Nature, 430, 184
Cimatti A. et al., 2008, A&A, 482, 21
Ciotti L., Ostriker J. P., 2007, ApJ, 665, 1038
Conroy C., 2013, ARA&A, 51, 393
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
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Birnboim Y., 2008, MNRAS, 383, 119
Dekel A., Burkert A., 2014, MNRAS, 438, 1870
Dekel A., Sari R., Ceverino D., 2009b, ApJ, 703, 785
Dekel A., Silk J., 1986, ApJ, 303, 39
Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, MNRAS, 435, 999
Dekel A. et al., 2009a, Nature, 457, 451
D'Eugenio C. et al., 2020, ApJ, 892, L2
Emsellem E. et al., 2011, MNRAS, 414, 888
Esdaile J. et al., 2020, preprint (arXiv:2010.09738)
Feldmann R., Mayer L., 2015, MNRAS, 446, 1939
Forrest B. et al., 2020a, ApJ, 890, L1
Forrest B. et al., 2020b, ApJ, 903, 47
Franx M., van Dokkum P. G., Förster Schreiber N. M., Wuyts S., Labbé I., Toft S., 2008, ApJ, 688, 770
Genzel R., Baker A. J., Tacconi L. J., Lutz D., Cox P., Guilloteau S., Omont A., 2003, ApJ, 584, 633
Gilli R. et al., 2014, A&A, 562, A67
Glazebrook K. et al., 2004, Nature, 430, 181
Glazebrook K. et al., 2017, Nature, 544, 71
Gobat R. et al., 2012, ApJ, 759, L44
Gobat R. et al., 2017, A&A, 599, A95
Grogin N. A. et al., 2011, ApJS, 197, 35
Guo Y. et al., 2009, MNRAS, 398, 1129
Gómez-Guijarro C. et al., 2019, ApJ, 886, 88
Hill A. R., Muzzin A., Franx M., van de Sande J., 2016, ApJ, 819, 74
Hill A. R., van der Wel A., Franx M., Muzzin A., Skelton R. E., Momcheva I., van Dokkum P., Whitaker K. E., 2019, ApJ, 871, 76
Holden B. P., van der Wel A., Rix H.-W., Franx M., 2012, ApJ, 749, 96
Hopkins P. F., Cox T. J., Kereš D., Hernquist L., 2008, ApJS, 175, 390
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1
Hsu L.-Y., Stockton A., Shih H.-Y., 2014, ApJ, 796, 92
Ilbert O. et al., 2010, ApJ, 709, 644
Ilbert O. et al., 2013, A&A, 556, A55
Jin S. et al., 2018, ApJ, 864, 56
Jin S. et al., 2019, ApJ, 887, 144
Kauffmann G. et al., 2003, MNRAS, 341, 54
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Khochfar S., Ostriker J. P., 2008, ApJ, 680, 54
Khochfar S., Silk J., 2006, ApJ, 648, L21
Koekemoer A. M. et al., 2007, ApJS, 172, 196
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221
Kriek M. et al., 2006, ApJ, 649, L71
Krogager J. K., Zirm A. W., Toft S., Man A., Brammer G., 2014, ApJ, 797, 17
Kubo M., Tanaka M., Yabe K., Toft S., Stockmann M., Gómez-Guijarro C., 2018, ApJ, 867, 1
Kubo M., Yamada T., Ichikawa T., Kajisawa M., Matsuda Y., Tanaka I., Umehata H., 2017, MNRAS, 469, 2235
Labbé I. et al., 2005, ApJ, 624, L81
Lagos C. d. P., da Cunha E., Robotham A. S. G., Obreschkow D., Valentino F., Fujimoto S., Magdis G. E., Tobar R., 2020, MNRAS, 499, 1948
Laigle C. et al., 2016, ApJS, 224, 24
Lang P. et al., 2014, ApJ, 788, 11
Le Floc'h E. et al., 2009, ApJ, 703, 222
Lilly S. J., Carollo C. M., 2016, ApJ, 833, 1
Lilly S. J., Eales S. A., Gear W. K. P., Hammer F., Le Fèvre O., Crampton D., Bond J. R., Dunne L., 1999, ApJ, 518, 641
Lilly S. J. et al., 2009, ApJS, 184, 218
Lin L. et al., 2007, ApJ, 660, L51
Lin L. et al., 2008, ApJ, 681, 232
Longhetti M., Saracco P., 2009, MNRAS, 394, 774
Maltby D. T., Almaini O., Wild V., Hatch N. A., Hartley W. G., Simpson C., Rowlands K., Socolovsky M., 2018, MNRAS, 480, 381
Man A., Belli S., 2018, Nat. Astron., 2, 695
Maraston C., Daddi E., Renzini A., Cimatti A., Dickinson M., Papovich C., Pasquali A., Pirzkal N., 2006, ApJ, 652, 85
Marchesini D. et al., 2010, ApJ, 725, 1277
Marchesini D. et al., 2014, ApJ, 794, 65
Marchesi S. et al., 2016, ApJ, 817, 34
Marleau F. R., Simard L., 1998, ApJ, 507, 585
Marsan Z. C., Marchesini D., Brammer G. B., Geier S., Kado-Fong E., Labbé I., Muzzin A., Stefanon M., 2017, ApJ, 842, 21
Marsan Z. C. et al., 2015, ApJ, 801, 133
Marsan Z. C. et al., 2019, ApJ, 871, 201
Martig M., Bournaud F., Teyssier R., Dekel A., 2009, ApJ, 707, 250
Martin D. C. et al., 2005, ApJ, 619, L1
Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulian E., 2010, MNRAS, 401, 371
McCracken H. J. et al., 2010, ApJ, 708, 202
McCracken H. J. et al., 2012, A&A, 544, A156
McGrath E. J., Stockton A., Canalizo G., Iye M., Maihara T., 2008, ApJ, 682, 303
McLure R. J. et al., 2013, MNRAS, 428, 1088
Merlin E. et al., 2018, MNRAS, 473, 2098
Momcheva I. G. et al., 2016, ApJS, 225, 27
Mowla L. A. et al., 2019, ApJ, 880, 57
Muzzin A., van Dokkum P., Franx M., Marchesini D., Kriek M., Labbé I., 2009, ApJ, 706, L188
Muzzin A. et al., 2013, ApJ, 777, 18
Muzzin A. et al., 2013, ApJS, 206, 8
Naab T., Johansson P. H., Ostriker J. P., 2009, ApJ, 699, L178
Naab T., Khochfar S., Burkert A., 2006, ApJ, 636, L81
Nelson E. et al., 2014, Nature, 513, 394
Newman A. B., Belli S., Ellis R. S., 2015, ApJ, 813, L7
Newman A. B., Belli S., Ellis R. S., Patel S. G., 2018, ApJ, 862, 126
Onodera M. et al., 2012, ApJ, 755, 26
Onodera M. et al., 2015, ApJ, 808, 161
Oser L., Naab T., Ostriker J. P., Johansson P. H., 2012, ApJ, 744, 63
Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, ApJ, 725, 2312
Oteo I. et al., 2017, preprint (arXiv:1709.04191)
Pacifici C. et al., 2015, MNRAS, 447, 786
Pannella M. et al., 2009, ApJ, 701, 787
Patel S. G., Hong Y. X., Quadri R. F., Holden B. P., Williams R. J., 2017, ApJ, 839, 127
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
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010a, AJ, 139, 2097
Peng Y.-j., Renzini A., 2020, MNRAS, 491, L51
Peng Y.-j. et al., 2010b, ApJ, 721, 193

Perez J., Michel-Dansac L., Tissera P. B., 2011, MNRAS, 417, 580
Pignatelli E., Fasano G., Cassata P., 2006, A&A, 446, 373
Poggianti B. M. et al., 2013, ApJ, 762, 77
Renzini A. et al., 2018, ApJ, 863, 16
Sanders D. B. et al., 2007, ApJS, 172, 86
Saracco P., Longhetti M., Andreon S., 2009, MNRAS, 392, 718
Sargent M. T. et al., 2007, ApJS, 172, 434
Schreiber C. et al., 2015, A&A, 575, A74
Schreiber C. et al., 2018, A&A, 618, A85
Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Sérsic J. L., 1968, Atlas de Galaxias Australes
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Simpson J. M. et al., 2014, ApJ, 788, 125
Simpson J. M. et al., 2017, ApJ, 839, 58
Skelton R. E. et al., 2014, ApJS, 214, 24
Stefanon M., Marchesini D., Rudnick G. H., Brammer G. B., Whitaker K. E., 2013, ApJ, 768, 92
Stockmann M. et al., 2020, ApJ, 888, 4
Stockton A., Canalizo G., Maihara T., 2004, ApJ, 605, 37
Stockton A., McGrath E., Canalizo G., Iye M., Maihara T., 2008, ApJ, 672, 146
Straatman C. M. S. et al., 2015, ApJ, 808, L29
Strazzullo V. et al., 2015, A&A, 576, L6
Suess K. A., Kriek M., Price S. H., Barro G., 2019a, ApJ, 877, 103
Suess K. A., Kriek M., Price S. H., Barro G., 2019b, ApJ, 885, L22
Szomoru D., Franx M., Bouwens R. J., van Dokkum P. G., Labbé I., Illingworth G. D., Trenti M., 2011, ApJ, 735, L22
Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mand elker N., Primack J. R., 2016, MNRAS, 458, 242
Tacchella S. et al., 2015, Science, 348, 314
Tacconi L. J. et al., 2008, ApJ, 680, 246
Tanaka M. et al., 2019, ApJ, 885, L34
Toft S. et al., 2007, ApJ, 671, 285
Toft S. et al., 2014, ApJ, 782, 68
Toft S. et al., 2017, Nature, 546, 510
Trujillo I., Ferreras I., de La Rosa I. G., 2011, MNRAS, 415, 3903
Trujillo I. et al., 2006, MNRAS, 373, L36
Valentino F. et al., 2020, ApJ, 889, 93
van der Wel A., Bell E. F., van den Bosch F. C., Gallazzi A., Rix H.-W., 2009a, ApJ, 698, 1232
van der Wel A., Rix H.-W., Holden B. P., Bell E. F., Robaina A. R., 2009b, ApJ, 706, L120
van der Wel A. et al., 2011, ApJ, 730, 38
van der Wel A. et al., 2014, ApJ, 788, 28
van Dokkum P. G., Franx M., 1996, MNRAS, 281, 985
van Dokkum P. G., Franx M., 2001, ApJ, 553, 90
van Dokkum P. G. et al., 2008, ApJ, 677, L5
van Dokkum P. G. et al., 2014, ApJ, 791, 45
van Dokkum P. G. et al., 2015, ApJ, 813, 23
Wellons S. et al., 2015, MNRAS, 449, 361
Whitaker K. E., Kriek M., van Dokkum P. G., Bezanson R., Brammer G., Franx M., Labbé I., 2012, ApJ, 745, 179
Whitaker K. E. et al., 2010, ApJ, 719, 1715
Whitaker K. E. et al., 2017, ApJ, 838, 19
Williams C. C. et al., 2015, ApJ, 800, 21
Williams C. C. et al., 2017, ApJ, 838, 94
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Wu P.-F. et al., 2018, ApJ, 868, 37
Wu P.-F. et al., 2020, ApJ, 888, 77
Wuyts S. et al., 2012, ApJ, 753, 114
Yano M., Kriek M., van der Wel A., Whitaker K. E., 2016, ApJ, 817, L21
Zahid H. J., Geller M. J., 2017, ApJ, 841, 32
Zolotov A. et al., 2015, MNRAS, 450, 2327

APPENDIX A: COMPARISON OF SPECTROSCOPIC AND PHOTOMETRIC REDSHIFTS

In Section 2.3, we compare rest-frame UVJ colours of our sample with the full massive parent sample at $2.5 < z < 3.0$. To derive UVJ colours and stellar masses for the parent sample, we make use of photometric redshifts. To investigate how reliable the photometric redshifts are we compare in Fig. A1 spectroscopic redshifts of quiescent galaxies at $z_{\rm spec} \gtrsim 1.2$ with photometric redshifts from Muzzin et al. (2013) and Laigle et al. (2016). Redshifts from Krogager et al. (2014) and D'Eugenio et al. (2020) rely on *HST* grism data while redshifts from Onodera et al. (2015), Marsan et al. (2015), Gobat et al. (2017), Belli et al. (2017), Glazebrook et al. (2017), Schreiber et al. (2018), Valentino et al. (2020), and Stockmann et al. (2020) rely on spectroscopic observations from ground-based telescopes. Photometric redshifts from Laigle et al. (2016) have a larger scatter and are systematically underestimated in this redshift range, especially at $z_{\rm spec} \gtrsim 2.5$. We therefore use redshifts from Muzzin et al. (2013) together with the deeper photometry from Laigle et al. (2016) for our analysis in Section 2.3.

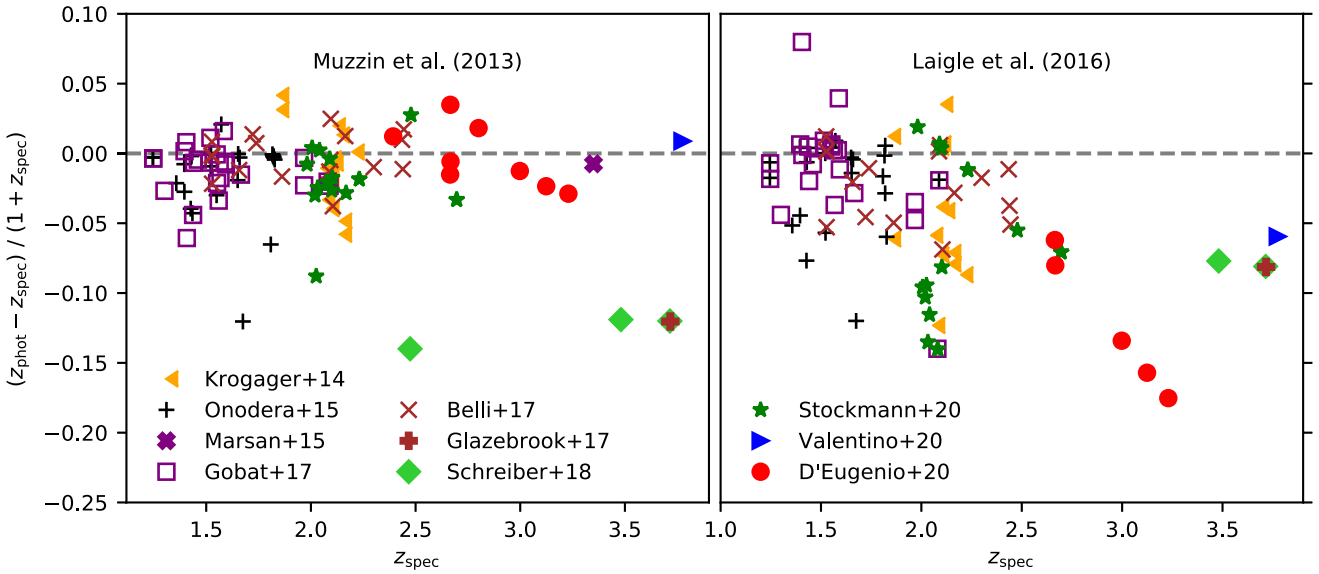

Figure A1. Comparison of spectroscopic redshifts of quiescent galaxies with photometric redshifts from Muzzin et al. (2013, left-hand panel) and Laigle et al. (2016, right-hand panel). Sources from Schreiber et al. (2018) at $z \approx 2.5$ and (Marsan et al. 2015) at $z \approx 3.4$ have a photometric redshift estimates in Laigle et al. (2016) of 4.9 and $z = 0.3$, respectively, and are not shown in the right-hand panel. The sources from Glazebrook et al. (2017) and Schreiber et al. (2018) at $z = 3.7$ are the same.

APPENDIX B: SIGNIFICANCE OF THE CENTRAL RESIDUALS

In Fig. B1, we show the $F160W$ images and residuals after subtracting the best-fitting Sérsic profiles (see Fig. 4 and Section 3) together with a plot of the significance of the residuals that we define as the absolute value of the residuals divided by the noise in each pixel. Considering only pixels of the $F160W$ images with a flux higher than 3 times the RMS of the background we find that the fraction of pixels in the residual images with a significance higher than 3 is $\lesssim 2$ per cent for all sources except for ID 8, where we find 4 per cent.

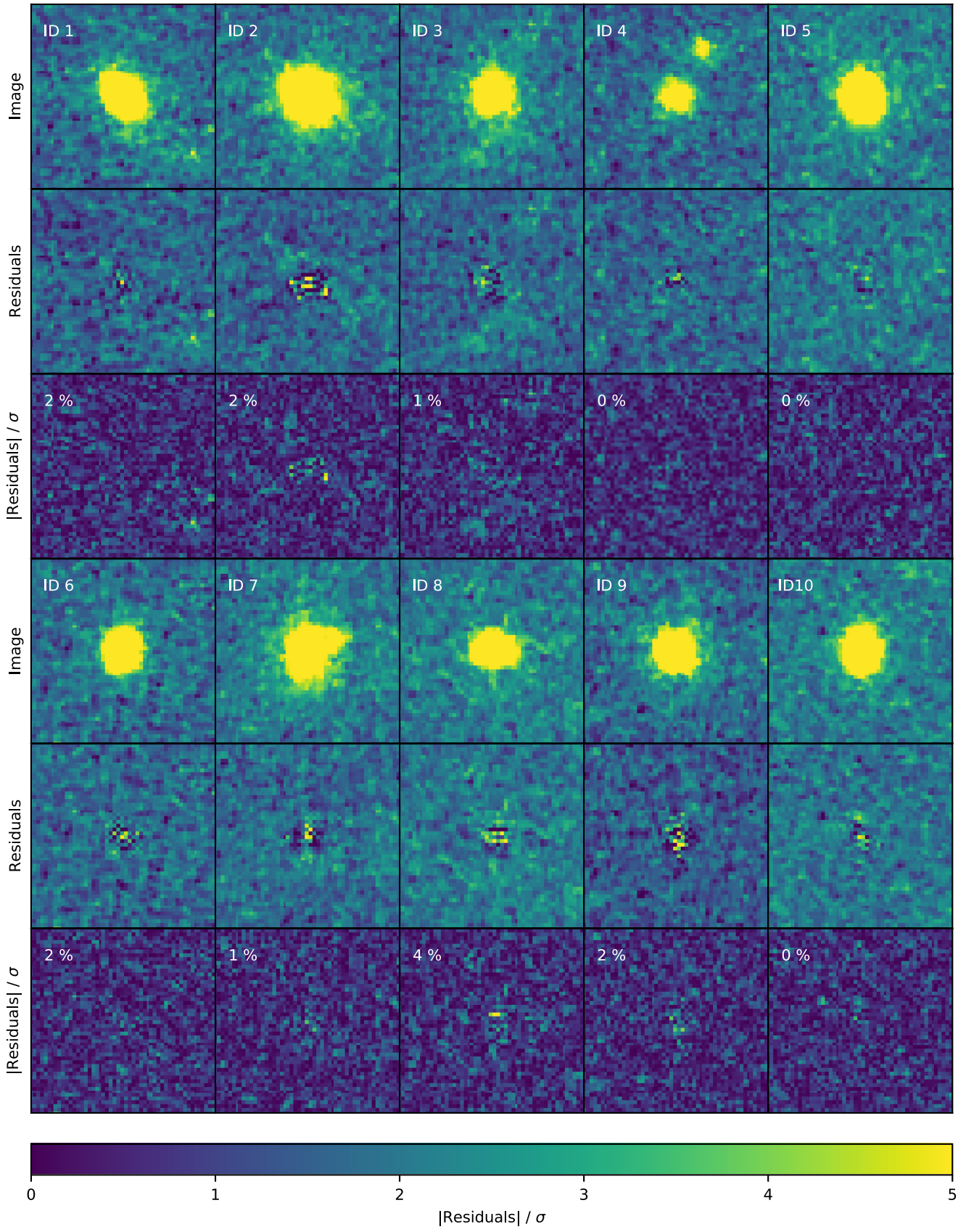

Figure B1. *HST F160W* images of the targets, residuals, and their significance, defined as |residuals|/σ. For each source, the fraction of pixels associated with the sources that have |residuals|/$\sigma > 3$ is indicated.

This paper has been typeset from a TEX/LATEX file prepared by the author.