MINOR MERGERS OR PROGENITOR BIAS? THE STELLAR AGES OF SMALL AND LARGE QUENCHED GALAXIES

Martina Fagioli1, C. Marcella Carollo2, Alvio Renzin2, Simon J. Lilly3, Masato Onodera1,3, and Sandro Tacchella1

1 Institute for Astronomy, ETH Zurich, 8093 Zurich, Switzerland; martina.fagioli@phys.ethz.ch
2 INAF-Osservatorio Astronomico di Padova, Vicolo dell’Osservatorio 5, I-35122, Padova, Italy
3 National Astronomical Observatory of Japan, National Institutes of Natural Sciences

Received 2016 March 24; revised 2016 July 13; accepted 2016 August 3; published 2016 November 4

ABSTRACT

We investigate the origin of the evolution of the population-averaged size of quenched galaxies (QGs) through a spectroscopic analysis of their stellar ages. This evolution has been claimed to arise from either the size growth of individual galaxies through a sequence of dry minor mergers, or the addition of larger, newly quenched galaxies to the pre-existing population (i.e., a progenitor bias effect). We use the 20k zCOSMOS-bright spectroscopic survey to select bona fide QGs at $0.2 < z < 0.8$. We stack their spectra in bins of redshift, stellar mass, and size to compute stellar population parameters through fits to the rest-frame optical spectra and Lick indices. The size–age relation differs below and above $\sim 10^{11}M_\odot$: at $10.5 < \log M_*/M_\odot < 11$, at all redshifts the stellar populations of the largest galaxies are younger than those of the smaller counterparts, indicating progenitor bias as the main driver of the average size evolution. In contrast, at higher masses, there is no clear size–age trend, supporting a substantial role of dry mergers in increasing with cosmic time the sizes of these most massive QGs. The [$\alpha$/Fe] abundance ratios of QGs are (i) above-solar over the entire redshift range of our analysis, hinting at universally short timescales for the build-up of the stellar populations of QGs, and (ii) similar at all masses and sizes, suggesting similar (short) timescales for the whole QG population—and strengthening the role of mergers in the build-up of the most massive QGs in the universe.

Key words: galaxies: evolution – galaxies: formation – galaxies: fundamental parameters

1. INTRODUCTION

The observed evolution with cosmic time in the population-averaged size of quenched galaxies (QGs, here often also referred to as "passive" or "quiescent" galaxies, as opposed to "star-forming" galaxies) at fixed stellar mass has received a lot of attention in the past decade (e.g., Daddi et al. 2005; Trujillo et al. 2007; Cimatti et al. 2008; van Dokkum et al. 2008; Cassata et al. 2011; Carollo et al. 2013, hereafter C13; Poggianti et al. 2013; van der Wel et al. 2014). The median half-light radius of QGs is about a factor $\sim 3$–5 larger in the local universe than at redshift $z \sim 2$ (Newman et al. 2012). The size growth scales as roughly $(1+z)^{-3}$, and it is similar to the rate of growth of the sizes of dark matter halos, but is somewhat steeper than the latter. This has sparked an intense debate concerning the physical mechanism behind this size evolution.

There are two main scenarios to which the evolution of the size–mass relation has been ascribed: the growth of individual QGs through a series of dry minor merger events, or the continuous addition of larger, recently quenched, galaxies at later epochs. This effect is an example of so-called "progenitor bias" in the sense that the population changes because of a change in membership rather than through changes in individual members (e.g., Franx et al. 2008; Newman et al. 2012; C13; Poggianti et al. 2013; Belli et al. 2015). In the individual growth scenario, the compact cores of QGs would remain constant in mass within a few kiloparsecs, but would accrete extended stellar envelopes around them (Cimatti et al. 2008; Hopkins et al. 2009; Naab et al. 2009; Trujillo et al. 2011; Cappellari 2013). Contrary to major mergers, minor gas-poor (dubbed "dry") mergers could have a key role: for a given amount of added mass, mergers with higher mass-ratios (i.e., minor mergers) result in a larger size increase (Villumsen 1983; Hilz et al. 2012; see also Feldmann et al. 2010; Taylor et al. 2010; Szomoru et al. 2011; McLure et al. 2013; van der Wel et al. 2014). This scheme would require $\sim 10$ dry mergers with $\sim 1:10$ mass ratio to account for the observed growth in size (Naab et al. 2009; van de Sande et al. 2013). The mergers are required to be dry since "wet" mergers, involving gas-rich galaxies, are expected to lead to central star formation and therefore to a reduction of the half-light radius of the primary galaxy. At a mass ratio of 1:10, the companion of a $10^{11}M_\odot$ galaxy is a $10^9M_\odot$ galaxy. Galaxies of this mass are generally gas-rich systems (e.g., Santini et al. 2014; Genzel et al. 2015). Therefore, the sequence and number of the required dry mergers, without a substantial contribution of wet mergers, is quite problematical, an aspect which has been substantially ignored so far.

Regardless of whether the merger scenario can explain the observed effect, the possible effects of progenitor bias must in any case be taken into consideration. An implicit assumption of the individual size growth view is that galaxies which are being quenched at different epochs have similar properties. If not, however, this could lead to a progenitor bias effect. In the context of the evolution of the average size–mass relation, the addition of newly quenched galaxies to a pre-existing population of QGs could lead to an observed growth of the average size of the population even if individual quiescent galaxies do not grow at all. This is particularly important especially in the light of the observed increase by about one order of magnitude of the comoving number density of massive (i.e., $\sim 10^{11}M_\odot$) QGs from $z = 2$ to the present epoch (e.g., Ilbert et al. 2010; Cassata et al. 2013; Muzzin et al. 2013).
Tracing the evolution of the number density of QGs of different sizes offers clues toward discriminating between the two scenarios. Different studies well agree on the evolution of the number densities of the smallest and densest QGs at stellar masses above $\sim 10^{11} M_\odot$, where a steady decrease is observed with cosmic time. At lower masses, however, different authors report different results. For example, C13 did not find any change in the number density of their “compact” galaxies at masses $10.5 < \log M_*/M_\odot < 11$; they reported instead a substantial increase in the number density of large QGs. The constancy of the compact population and the increase in the large population led those authors to advance the progenitor bias interpretation. At similar $10.5 < \log M_*/M_\odot < 11$ masses, however, van der Wel et al. (2014) report a strong decrease in the number density for compact QGs since $z = 1.5$, and therefore interpret their observed disappearance of these objects at the lower redshifts as indication of a growth in size of individual QGs.

In comparing results from different studies, it is important however to note that the adopted definition of stellar mass is an important factor when discussing the evolution of the size–mass relation. In C13, and also in this paper, we will define the stellar masses to be the integral of the star formation rate (SFR). These are about 0.2 dex larger than the commonly used definition which subtracts the mass returned to the interstellar medium, i.e., the mass of surviving stars plus compact stellar remnants. The former has the feature of remaining constant after the galaxy ceases star formation, whereas the latter continually decreases. Thus, when comparing the properties of QGs in a given mass bin across cosmic time, one should clearly use the former. This effect explains part of the discrepancies found in the different number density analyses: effectively high redshift galaxies are given a spuriously high mass, which leads to them appearing to be too small and to have a higher number density at high redshift. Another factor that leads to different estimates for the evolution of the number densities of small QGs is the definition of the bins in which the densities are computed, in particular whether a single size threshold is used to compare number densities at different redshifts, or whether the bins are defined along the size–mass relation at each given redshift (which, due to its evolution, implies a comparison between populations of different sizes).

Number densities alone however are not conclusive. C13 and Damjanov et al. (2015) agree that, at masses below the $\sim 10^{11} M_\odot$ scale, the number densities of compact QGs remain constant since at least $z \sim 1$; these authors reached, however, different conclusions on the origin of this constancy. Damjanov et al. (2015) proposed that the compact QG population is continuously replenished with younger members, so as to compensate for the shift toward larger sizes of individual galaxies due to mergers. In contrast, C13 argued that the compact population remains stable since $z \sim 1$ and the newly accreted members of the population have increasingly larger sizes at steadily lower redshifts.

These two interpretations can be easily tested through the average ages of the populations involved. If the increase of the median size is due to the addition of newly quenched galaxies that are progressively larger toward lower redshifts then, at any epoch, the stellar populations of larger QGs should be younger than those of smaller QGs of similar mass. On the other hand, if individual QGs grow their sizes through mergers and the number density of compact QGs remains more or less constant due to the continuous production of compact QGs then, at any epoch, smaller QGs should be younger on average than their later relatives of similar mass. Therefore, the stellar ages of the galaxies offer a powerful discriminant between these two scenarios (see, e.g., Onodera et al. 2012; Belli et al. 2014; Keating et al. 2015; Yano et al. 2016).

C13 did a study of the colors of compact and large $< 10^{11} M_\odot$ QGs at different redshifts, and found that, at any epoch, larger QGs appear to be bluer than their smaller counterparts; it is this result that led those authors to conclude that, at these masses, the stellar populations of larger QGs are younger than those of smaller QGs and thus that the evolution in size of the whole populations is to a large extent ascribable to the addition of recently quenched, larger QGs. Galaxies quenched at later epochs are indeed expected to have larger sizes than those quenched earlier as (progenitor) star-forming galaxies also experience an evolution in their average size with cosmic time (e.g., Newman et al. 2012).

Stellar ages determined on the basis of a single rest-frame optical color (as done in C13), however, suffer heavily from the well-known degeneracy between age, metallicity, and also, possibly most problematically, dust effects (Worthey et al. 1994). We therefore push here the analysis of the stellar ages of small and large QG below and above the evidently important mass threshold of $10^{11} M_\odot$ using more robust spectroscopic measurements of stellar population properties. Our primary goal is to test whether and to what extent progenitor bias is driving the increase of the average size of passive galaxies as a function of stellar mass; specifically, we use two mass bins with boundaries $10.5 < \log M_*/M_\odot < 11$ and $11 < \log M_*/M_\odot < 11.5$. We also use the spectroscopic diagnostics to study the ratio of different elements in the attempt to constrain the timescales of buildup of the stellar populations of QGs of different masses and sizes.

Even spectroscopically, however, residual degeneracies between effects of age and metallicity continue to afflict galaxy ages, which are not straightforward to obtain. In the last few years, a number of “full spectral fitting” codes (e.g., Ocvirk et al. 2006a, 2006b; Cappellari & EMSL3mass 2004; Koleva et al. 2009, STECKMAP, pPXF and ULySS, respectively) have been developed to address this issue. In the full spectral technique, a set of templates is used to fit the overall shape of the spectrum. The most recent full spectral fitting codes do not fit the overall shape of the continuum, thereby avoiding common problems such as flux calibration and extinction. Instead, a polynomial function is used to fit the shape of the continuum. Full spectrum fitting codes are good at handling the impact of the age–metallicity degeneracy as they maximize the information used from the whole observed spectrum (Koleva et al. 2008; Sánchez-Blázquez et al. 2011; Beasley et al. 2015; Ruiz-Lara et al. 2015). We therefore adopt this methodology to derive our fiducial stellar population ages in this paper.

Besides the full spectral fitting analysis, we have also used the Lick line-strength indices to get independent estimates of ages and metallicities. The Lick system of spectral line indices is a commonly used method to determine ages and metallicities of stellar populations (e.g., Burstein et al. 1984; González 1993; Carollo & Danziger 1994; Worthey et al. 1994; Worthey & Ottaviani 1997; Trager et al. 1998, 2000, 2005; Poggianti et al. 2001; Thomas & Maraston 2003; Thomas et al. 2003, 2011; Korn et al. 2005; Schiavon 2007; Onodera et al. 2012, 2015). The system consists of a set of 25 optical absorption line
indices, spanning a wavelength range from ~4080 to ~6400 Å. The absorption features are particularly useful because they are largely insensitive to dust attenuation (MacArthur 2005). Even so, age-dating based on the Lick indices is not free from degeneracy effects. Its main pitfall is that most indices are sensitive to all the basic population parameters, namely age, metallicity, and the ratio of $\alpha$-elements to iron. Circulation of the errors can generate spurious correlations or anticorrelations (Kuntschner et al. 2001; Thomas et al. 2005; Renzini 2006). For example, an underestimate in the strength of a Balmer line (mainly sensitive to age), due, e.g., to partial filling in by an emission line, may lead to an overestimate in the age, but having an overestimated metallicity, the procedure is forced to underestimate metallicity in order to match the strength of the metal lines. In this way, a spurious age–metallicity anticorrelation can be generated. A strength of our analysis is thus to attempt to mitigate the intrinsic degeneracies by using and comparing for cross-validation both methodologies, i.e., the full spectral fitting approach and the Lick indices approach.

We include in our study a brief introspection of the $\alpha$-elements to Fe-element abundance ratios (i.e., $[\alpha$/Fe]) in QGs of different masses and sizes. The $[\alpha$/Fe] ratio is a well-known diagnostic to constrain formation timescales (Matteucci & Greggio 1986; Pagel & Tautvaisiene 1995), since $\alpha$-elements such as O, Ne, Mg, Si, S, Ar, Ca, and Ti (i.e., nuclei that are built up with $\alpha$-particles) are delivered mainly by core collapse (CC) supernova explosions of massive stars and thus on much shorter timescales than elements such as Fe and Cr, which come predominantly from the delayed explosion of SNe Ia (e.g., Nomoto et al. 1984; Woosley & Weaver 1995; Thielemann et al. 1996). Enhanced values of $[\alpha$/Fe] ratio indicate a short formation timescale.

The paper is organized as follows. In Section 2 we describe the data set, the zCOSMOS-bright 20k catalog, and its features. Section 3 summarizes the basic measurements and presents the spectroscopic sample selection in detail. Section 4 presents the steps we took in the course of our analysis. Section 4.1 describes the binning in mass, size and redshift, and Section 4.2 describes the stacking procedure that was used to obtain average spectra as a function of redshift, mass and size. The fitting used to correct for the emission lines contribution is described in Section 4.3. Section 4.4 describes how we derived ages with full spectral fitting using pPXF, and Section 4.5 describes how we measured the Lick strengths and derived the stellar population parameters from them. In Section 5 we present our results, followed by a discussion in Section 6. In Section 7 we summarize our paper and present the conclusions.

Through this paper we adopt a $\Lambda$-dominated cold dark matter ($\Lambda$CDM) cosmology, with $\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$ and $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$. All magnitudes are given in AB system. We use “dex” to refer to the antilogarithm, so that 0.3 dex represents a factor of 2.

2. DATA SET

We use the zCOSMOS-bright 20k sample, to which, however, we apply several cuts in order to limit the redshift range to the $0.2 \leq z \leq 0.8$ interval, the mass range to the $10.5 < \log M_\text{b}/M_\odot < 11.5$ interval, and to ensure that the selected galaxies are bona fide passive systems. In Figure 1 we show a schematic view of the selection criteria that we have applied to the original sample, and in the next sections we describe in detail the steps which lead to the final selection.

2.1. The 20k zCOSMOS-bright Catalog

The spectra we employed come from the full zCOSMOS-bright 20k catalog (Lilly et al. 2007, 2009). Here we briefly summarize the data. zCOSMOS-bright consists of about 20,000 galaxy spectra selected to have $I_{AB} < 22.5$ across the full 1.7 deg$^2$ in the COSMOS field (Scoville et al. 2007). The zCOSMOS project (Lilly et al. 2007) is a large redshift survey of galaxies undertaken on the ESO VLT. The bright part uses the VIMOS MR grating with a resolution of $R \sim 600$ and a pixel size of $\approx 2.553$ Å.

The VIMOS wavelength coverage spans from 5500 to 9700 Å. The greatest advantage of the zCOSMOS spectroscopic survey is to combine high-quality spectra with a compilation of multi-wavelength imaging of the COSMOS survey data set, including Hubble Space Telescope/ACS data (Koekemoer et al. 2007). Therefore we have available high-quality spectroscopic redshifts, photometrically derived quantities, and high-resolution images.

The typical redshift uncertainty in zCOSMOS-bright is $\pm 110$ km s$^{-1}$ (Lilly et al. 2007). A confidence class parameter is introduced to estimate the reliability of the redshift assignment. Also, objects flagged with confidence class 4 usually show high-quality spectra. The spectroscopic redshifts are compared with photometric redshifts derived from the COSMOS multi-band photometric data and a decimal number is used to flag the agreement or otherwise between the photometric and spectroscopic redshifts. For a complete description of confidence classes the reader is referred to Lilly et al. (2007, 2009).

2.2. The S-COSMOS MIPS 24 μm Catalog

The S-COSMOS survey (Sanders et al. 2007) is a deep infrared imaging survey, which comprises IRAC 3.6, 4.5, 5.8, and 8.0 μm and MIPS 24, 70 and 160 μm observations including the entire 2 deg$^2$ of the COSMOS-ACS field. It was carried out with the Spitzer Space Telescope as part of the
Spitzer Cycle 2 and 3 Legacy Programs. In Cycle 2, the COSMOS field was mapped at 24 μm. The observations performed in Cycle 3 mapped the entire COSMOS area reaching deeper flux limits, down to a flux density limit S_{24μm} ≈ 0.08 mJy.

3. MEASUREMENTS AND SAMPLE SELECTION

3.1. Redshifts

We selected objects within a restricted redshift range of 0.2 < z < 0.8 so as to be complete in mass down to log M_*/M_⊙ = 10.5 at z = 0.8 (Pozzetti et al. 2010). In order to achieve a high signal-to-noise ratio, adequate for a Lick-based stellar population analysis, we further restricted the sample to confidence Classes 3 and 4 (including secondary objects in Classes 23 and 24, but excluding objects with broad emission lines, Classes 13 and 14). Class 3 and 4 spectra have a very secure redshift assignment (reliability > 99.8%). Almost all the galaxies in our sample (98.5%) have been flagged with the 0.5 decimal number, indicating an agreement between spectroscopic and photometric redshift to within 0.08(1+z) (a subsequent visual inspection of all of the final set of objects confirmed the correctness of the assigned spectroscopic redshifts). After this first selection, we end up with a sample of 9208 objects.

3.2. Stellar Masses

Our sample is matched with that from C13 to get structural parameters, whose derivation was fully described in that paper. The stellar masses are derived with the ZEBRA+ (Oesch et al. 2010) code from synthetic SED fitting to 11 photometric broad bands from 3832 Å (u’, CFHT) to 4.5 μm (Spitzer/IRAC channel 2). A set of star formation history models with exponentially declining SFRs, ranging metallicities from 0.05 to 2 Z⊙, with decay timescales from τ ≈ 0.05 to 9 Gyr, and ages from 0.01 to 12 Gyr, constitutes the SED library. This was constructed with the Bruzual & Charlot (2003) (BC03) stellar population synthesis code with a Chabrier initial mass function (Chabrier 2003). We use dust reddening from Calzetti et al. (2000) with E(B−V) as a free parameter. As discussed in the Introduction, we stress that our stellar masses are defined as the time-integral of the SFRs, which are on average around 0.2 dex higher than stellar masses which exclude the mass that is subsequently returned to the interstellar medium.

3.3. Sizes

We adopt the half-light radius (r_{1/2}) as an estimation of the size of our galaxies. The procedure is fully described in C13. The sizes of our galaxies have been measured with the software ZEST+ (Zurich Estimator of Structural Types Plus), an extended version of ZEST (Scarlata et al. 2007). The main advantage is that it measures the half-light radius within elliptical apertures, instead of the circular apertures of SExtractor (Bertin & Arnouts 1996). ZEST+ requires as input the total apparent flux of each galaxy, which is taken from the 2.5 R_{Kron} (Kron 1980) value from SExtractor. Using a file with the positions of close objects, ZEST+ first replaces the segmentation maps of companion galaxies with random sky values, and then estimates the local sky background of the galaxy. Finally, the code outputs the semimajor axis of the corresponding elliptical aperture that encloses half of the output total flux. In Figure 3 we show the stellar mass and size distributions before and after the mass cut.

3.4. Spectroscopic Sample Selection

To proceed toward our goal of studying the properties of QGs, we give special attention to the separation between star-forming and quiescent galaxies. A number of studies have used photometric information to achieve this separation; a UVJ color–color diagram has been found to be particularly effective in this direction (Wuyts et al. 2007; Patel et al. 2009; Williams et al. 2009; Whitaker et al. 2012; Moresco et al. 2013; Muzzin et al. 2013). A cleaner separation can clearly be achieved however using spectroscopic diagnostics (see e.g., Onodera et al. 2015). Our selection of quiescent galaxies was done by choosing objects which show no, or only very weak, emission lines, with the following procedure.

1. Identify the expected wavelength of Hα, Hβ and [O II] 3727 Å.
2. Define a continuum region on either side of this location to be between one and five times the FWHM of the instrumental resolution (R = λ/Δλ = 600) and compute the mean (f_{cont}) and the standard deviation (σ) of the continuum per pixel.

We then consider a line to be detected in emission if the peak of the line exceeds three times the noise per pixel,

\[ \text{max}(f_{\text{line}} - f_{\text{cont}}) > 3\sigma. \]  

This is straightforwardly applied for Hα and [O II] and the object is excluded if either line is detected. The case of Hβ is complicated by stellar absorption lines. If neither of [O II] and Hα lines is in the observed wavelength range, we use an empirical calibration of the ratio of the peak of the [O II] and Hβ, [O II] ≈ 2.8Hβ − 2, which yields an equivalent threshold of ≥1.8σ, using the standard deviation of the continuum from the blue side of the feature only, due to the proximity of [O III] 5007 Å. Our wavelength range includes Hα for 1042 galaxies and [O II] 3727 Å for 3826 objects. For the rest of the galaxies in the sample, we use Hβ as the criterion to exclude emission line objects. As a result of this selection, we find a total of 2094 galaxies showing no detected emission lines.

To further check against the presence of star-forming objects, we cross-match these 2094 galaxies with the S-COSMOS MIPS 24 μm Photometry Catalog (2008 October) and with C-COSMOS (Chandra COSMOS) (Elvis et al. 2009), identifying sources within 2″ (see also Caputi et al. 2008). We find 257 galaxies having a possible detection in the MIPS catalog and 28 in the X-ray. We discarded 38 objects which had a MIPS detection and for which subsequent visual inspection of the spectra revealed emission lines that had not been found by the automatic algorithm described above. The emission lines were hidden at the edges of the wavelength range, where the signal-to-noise ratio is lower and fringing may alter the shape of the continuum. None of the objects with X-ray detection shows emission features in the zCOSMOS spectra and therefore we keep those galaxies in the quiescent sample.
As the selection of a purely quiescent sample is crucial for our analysis, because the star-forming objects are expected to be larger than quiescent ones, we visually inspected the spectra of all the galaxies in the sample. We further discard 51 galaxies for which visible emission lines had not been detected from the automatic code. At the end of this selection, 2005 objects are defined as purely quiescent galaxies. We stress that this selection has been made purely on the basis of the spectra (and MIPS and X-ray catalogs) with no reference to the images, sizes, or morphologies of the galaxies.

As a final check that we have succeeded in excluding all galaxies with significant emission lines, we stack the spectra of a subsample consisting of a randomly chosen 5% of the final set of 2005 galaxies and measure the equivalent width of [O II] (EW([O II])), with continuum and line passbands defined as in Balogh et al. (1999). We repeat this procedure 1000 times. Figure 2 shows the distribution of the EW([O II]) of these stacked spectra. The EW([O II]) of the stacked spectra was in every case far below the 5 Å limit which is commonly used to separate star-forming from quiescent galaxies (e.g., Mignoli et al. 2009; Moresco et al. 2010, see also Balogh et al. 1999 for the typical values of EW([O II]) they found for quiescent galaxies). In fact, the maximum EW for the [O II] emission line in these 1000 trials was <2 Å, which corresponds roughly to a log (sSFR/Gyr^{-1}) < −2 (see also Figure 5 in Moresco et al. 2013), and the mean is below zero.

In Figure 1, where we show the steps of our sample selection, we also give the number of galaxies remaining in the sample at each of the steps. The final step is to apply a cut in stellar mass at 10.5 < log M/M_⊙ < 11.5, which should be complete in the redshift range 0.2 < z < 0.8, applying the procedure described in Pozzetti et al. (2010). Figure 3 shows the distribution of the stellar masses of the passive sample, before the cut, and of the redshifts and sizes before and after the mass cut is applied.

4. ANALYSIS

4.1. Sample Binning

In order to study the average properties of galaxies in our sample, we split them into bins of redshift, stellar mass and size. Our galaxies have three equal bins in redshift within the redshift interval 0.2 < z < 0.8, with Δz = 0.2. We cut the redshift interval at z = 0.8 in order to be complete in mass down to log M*/M_⊙ = 10.5. We then divide the mass range into two bins with size of 0.5 dex, as in C13, and further divide the sample into three size bins, following two different procedures, as follows.

Quiescent galaxies are known to follow a tight relation between their size (r_1/2) and stellar mass (M_⋆), which evolves with redshift (Daddi et al. 2005; Williams et al. 2010; Newman et al. 2012; Mosleh et al. 2013; Patel et al. 2013; van der Wel et al. 2014; Tacchella et al. 2015b). Our final goal is to compare the average stellar population parameters for a sample of “small” and “large” QGs. Therefore, we fit a size--mass relation for the different redshift bins, finding minimal or no variation in the slope, as expected (e.g., van der Wel et al. 2014). We therefore fix the slope at the value obtained in the central redshift bin, 0.4 < z < 0.6, as shown in the central panel in the left-hand plot of Figure 4. The resulting values of the intercepts at different redshifts are given in Table 1. As expected, we find a decrease in the size (at given mass) with redshift. Our first approach to binning in size constructs bins relative to this evolving mean relation. This is shown in the left-hand plot in Figure 4. The two dashed lines, with the same slope as our fitted size--mass relation, splits the sample into 35:30:35% of the galaxies (across the whole redshift range). We define the galaxies lying in these areas as “small,” “intermediate-size” and “large” galaxies. The bold numbers in Figure 4 show how many passive galaxies we stacked as a function of redshift, size, and mass. We define this binning as the “size--mass cut.”

This cut is useful to compare the average stellar ages among different sizes in the same redshift bin. However, the comparison of stellar ages between different redshift intervals becomes difficult, as different populations of galaxies may enter into the definition of small and large as the mean relation evolves. To make such a comparison, we also apply a different binning in size, which we define as the “horizontal cut” (as shown in the right-hand panels in Figure 4). In this case, we define for each mass bin, three bins in size that are the same at all three redshifts. Specifically, for the mass bin 10.5 < log M*/M_⊙ < 11, we define as small the galaxies having r_1/2 < 2 kpc, intermediate, 2 < r_1/2 < 5 kpc, and large, r_1/2 > 5 kpc. For more massive galaxies (11 < log M*/M_⊙ < 11.5), small galaxies have r_1/2 < 4.5 kpc and large >7.5 kpc. This allows us to track the evolution of the stellar ages of galaxies with same size and same stellar masses (i.e., integrated SFRs, see Section 3.2) through different cosmic epochs.

4.2. Stacking

We computed the average stacked galaxy and noise spectra for the galaxies in each bin of redshift, stellar mass, and size. First, we de-redden the individual spectra for Galactic extinction following the extinction curve for diffuse gas from O’Donnell (1994) with R_V = 3.1, and using the Galactic E(B−V) values from the maps of Schlegel et al. (1998). We find a typical correction factor of f_{obs} ∼ 0.96 f_{dered}. We do not
correct for any internal dust extinction in the galaxies, since they are expected to be passive systems with negligible dust.

We then de-redshift the spectra to the rest frame and normalize them by the mean flux at rest frame $4100 < \lambda < 4700 \, \text{Å}$. The spectra are then interpolated onto a 1 Å linearly spaced wavelength grid. The associated noise spectra, which we use as weights during the stacking procedure, are normalized by the same factor as that for the object spectra and interpolated to the rest-frame in quadrature. The spectra are a straight average of the individual spectra, weighted by the signal to noise in each spectrum at each wavelength.

In Figure 5 we show the stacked spectra of all bins of our analysis, i.e., all redshifts, mass bins, and size bins. In each panel we show two spectra, respectively for the size–mass binning and for the horizontal-cut binning. To highlight the quality of the stacked spectra, we replot in Figure 6 the spectra of only the small galaxies, this time however with an expanded horizontal axis. The main absorption lines that we have used for our measurements of ages are marked with vertical gray bands; also shown, with vertical dashed lines, are the spectral features which were excluded from the computation of the stellar ages, as discussed in Section 4.5. Due to the variety of the redshifts in the sample, at all redshifts the stacked spectra overlap approximately in the wavelength range $4150 < \lambda < 4800$ (shaded yellow area in Figure 6).

We fit our stacked spectra with pPXF (Cappellari & Emsellem 2004), adopting stellar templates from version 9.1 of the MILES library, which consists of 985 stars, whose spectra cover a range of $3525$–$7500$ Å at $2.51$ Å (FWHM) spectral resolution (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011). During the pPXF fit, we use a fourth-order additive and fourth-order multiplicative polynomial correction for the spectral slope of our template. Additive polynomials are introduced to correct low-frequency differences in shape between the galaxy and the templates (Cappellari & Emsellem 2004). The multiplicative polynomials are included to ensure that the results are insensitive to the normalization or flux calibration of galaxy and stellar template spectra (Kelson et al. 2000). The polynomial degree is chosen to maximize the quality of the fit, which is also confirmed by a visual inspection.

We find values for the velocity dispersions of between $125$ and $215 \, \text{km s}^{-1}$, after subtracting the instrumental velocity dispersion in quadrature. We adopt the same procedure as Cappellari et al. (2009) and compute $\sigma_{\text{stack}}$ using

$$\sigma_{\text{stack}} \approx \frac{c \Delta z}{(1 + z)}$$

where $c$ is the speed of light and $\Delta z = 110 \, \text{km s}^{-1}$ is the zCOSMOS-bright redshift error. We find values for $\sigma_{\text{stack}}$ of $~60$–$90 \, \text{km s}^{-1}$, depending on the redshift. The best-fit templates are shown as dashed red lines in Figures 5 and 6.

For each of the stacked spectra of Figure 5, we show greatly expanded in Figure 7 the residuals obtained after subtraction of the best-fit stellar-superposition templates. The regions around the Balmer lines, $\text{[O III]}$ and $\text{[O II]}$ have been masked with the goodpixels function provided in pPXF. For the small galaxies, the residual spectra are also shown along the bottom of each panel in Figure 6.

In each panel of Figure 5, the upper spectrum refers to the cut along the size–mass relation, the lower one to the horizontal cut. The horizontal cut spectra have been shifted by subtracting a constant value from the normalized spectra. This is also done for the residuals in Figure 7. The different background colors in both figures indicate different redshift bins.

It is clear that all of the stacked spectra are dominated by the typical features of an old stellar population, such as a strong $4000$ Å break, the G-band at $4300$ Å, the $5180$ Å Mg and the Balmer absorption features.

Nevertheless, it is clear that the residuals from essentially all of the stacks show a weak narrow emission line contamination. For masses $10.5 < \log M/M_* < 11$ (first column of residuals in Figure 7), the emission line fill-in appears to be stronger for larger galaxies, especially in the H$\beta$, H$\gamma$ and [O II] region. The same does not seem to apply to more massive galaxies (second column of residuals in Figure 7). We fit the residual emission lines and compute the equivalent width of [O II] and H$\beta$. From the relation between sSFR, EW(H$\alpha$) and EW([O II]) we obtain

Figure 3. Histograms showing the distribution of the sample with respect to redshift, mass, and size, respectively. The left panel shows the redshift distribution of our objects. The central panel shows the distribution of stellar masses (i.e., integrated SFRs), measured with the ZEBRA+ (Oesch et al. 2010) code. The different colors represent the three redshift bins our sample is divided into. The vertical dashed lines in the central panel show our mass cuts at $10^{10.5} M_\odot$ and $10^{11} M_\odot$. The right panel shows the size distribution in kpc. The size measurements and their correction were described in C13.
log (sSFR/Gyr\(^{-1}\)) < -2 (see for example Moresco et al. 2013 for the relation between these quantities in zCOSMOS galaxies). We assume in deriving this the Balmer line ratios for the case B recombination with a temperature of 10^4 K and a typical electron density \(\leq 10^4\) cm\(^{-3}\) without reddening (Osterbrock & Ferland 2006) for the conversion between H\(\alpha\) and H\(\beta\) fluxes. This is a very low sSFR, \(< 10^{-11}\) yr\(^{-1}\); it corresponds closely to the inverse age of the universe at \(z \sim 0.3\).

4.3. Emission Line Correction

We correct for the emission line contribution that is visible in Figure 7 by fitting a Gaussian to the lines in the residual spectrum, while fitting at the same time the residual continuum level, which is very close to zero. We find velocity dispersions for the residual emission lines of H\(\beta\), H\(\gamma\) and H\(\delta\) lines that are consistent with the instrumental resolution. We then subtract the fitted emission lines from the observed spectra. In some bins, especially in the highest-redshift one, the correction is not applied, as the residuals corresponding to the Balmer lines are barely distinguishable from the noise. In order to check whether we are performing an over-subtraction of the lines, we also fit Gaussians fixing the continuum level at 1\(\sigma\). We find minimal or no variation between the two methods.

In the rest of our analysis, to derive stellar ages for the different stacked spectra, we use the spectra subtracted of their emission line component.

4.4. Full Spectral Fitting

We first estimate ages using pPXF for a full spectral fitting analysis, using a set of SSP templates with solar metallicity and Salpeter IMF from BC03 and exclude templates with age older than the age of the universe at each redshift and younger than 1 Gyr. The metallicity of quiescent galaxies at these masses is expected to be solar with a scatter of 0.1 dex (see also Gallazzi

![Table 1](image-url)

| Redshift (\(0.2 < z < 0.8\)) | Slope (\(\alpha\)) | Intercept (\(\beta\)) |
|-----------------------------|-----------------|---------------------|
| \(0.2 < z < 0.4\)          | 0.63            | 5.76                |
| \(0.4 < z < 0.6\)          | 6.06            |                     |
| \(0.6 < z < 0.8\)          | 6.43            |                     |

![Figure 4](image-url)

Figure 4. Size–mass relation for our sample (red circles) of QGs. Dashed lines indicate the adopted bins in size and stellar mass in each redshift interval. The three panels in each plot show the three different redshift bins. Numbers in the boxes are the number of objects to be stacked in each redshift, mass, and size bin. On the left, the two black dashed lines which follow the slope of the size–mass relation in each panel enclose the middle 30% of the distribution at each redshift. These bins allow the comparison of galaxies of the same size relative to the evolving mean relation. In contrast, the horizontal bins in the right-hand panel are chosen to allow the comparison of galaxies with the same absolute size at different redshifts.
Figure 5. Stacked spectra for all our redshift, mass and size bins. From top to bottom, the different shades of purple as background color separate the spectra in redshift. On the left side, we show galaxies with masses between 10.5 < \log M_*/M_\odot < 11, and on the right side, the spectra of more massive galaxies (11 < \log M_*/M_\odot < 11.5). The three size bins are as described in Section 4.1 and shown in Figure 4. In each panel, the upper spectrum shows the spectrum that is obtained by stacking spectra binned along the size–mass relation while the lower spectrum shows the stack obtained with the horizontal cut, shifted by subtracting a constant. In each case, the black line shows the observed stacked spectra, while the red dashed line shows the best-fit template that is obtained with pPXF.
et al. 2005, 2014; Thomas et al. 2010; Conroy & van Dokkum 2012). During the pPXF fit we use a tenth-order multiplicative polynomial correction and no additive polynomials. We use the emission line subtracted spectra as described in Section 4.3, without masking the regions of Balmer lines but masking out the regions corresponding to [O III] and [O II]. We compute the mass-weighted age among the best-fit templates found from pPXF.

To derive the error (error bars in Figures 8 and 10) on our age estimation, we use the jackknife technique:

$$\sigma_{\text{full}}^2 = \frac{N}{N-1} \sum_{i=1}^{N} (\text{Age}_{\text{full}} - \text{Age}_{\text{full}(i)})^2,$$

where $N$ is the number of objects stacked in each bin, $\text{Age}_{\text{full}}$ is the age measured on all $N$ spectra, and $\text{Age}_{\text{full}(i)}$ is the age
Figure 7. Residuals between the observed stacked spectra and the best-fit templates from the previous figure. The upper residuals come from the size–mass cut and the lower ones from the horizontal cut, having been shifted by a constant. For less massive galaxies, the residuals in the regions of Hβ, Hγ, Hδ and [O II] appear to show higher emission line contamination with increasing sizes. The same is not however seen for more massive galaxies. This effect is visible for galaxies binned with both a size–mass and an horizontal cut. However, from fitting the residuals we obtain very low sSFRs ($\log (sSFR/\text{Gyr}^{-1}) < -2$).
Figures 8 and 10 and listed in Table 3.

measured on a stacked spectrum made of $N - 1$ spectra by removing the $i$th spectrum. The ages obtained with this method are shown in Figures 8 and 10 and listed in Table 3.

4.5. Lick Indices Analysis

To derive the strengths of the spectral absorption features we use the bandpasses and pseudo-continua for each index from Trager et al. (1998). The resolution of the 20 indices that are covered in our different redshift ranges varies from 10.9 Å ($\sim 340$ km s$^{-1}$) for H$\delta_A$ to 8.4 Å ($\sim 200$ km s$^{-1}$) for Fe5406 (see Table 2 for the indices we use at each redshift bin). The zCOSMOS-bright spectra have a resolution that is better than the defining Lick spectra for some wavelengths and worse for others. To correct the measured values from our spectra to the Lick resolution, we carry out the following procedure:

$$I_{\text{Lick}} = I_{\text{obs}} \frac{r_{\text{obs}}}{r_{\text{best}}}$$

where $I_{\text{obs}}$ is the index derived from the pPXF best-fit at the observed velocity dispersion and $I_{\text{Lick}}$ on pPXF original template, convolved to the Lick resolution. The best-fit templates we use are those derived in Section 4.2 with MILES stellar library. For all indices the complete set of Lick’s resolutions is from Schiavon (2007). We derive the stellar population parameters from the corrected indices $I_{\text{corr}}$. In order to get the correct Lick resolution for $I_{\text{Lick}}$, the template FWHM = 2.51 Å has been taken into account.

In order to check whether the index values depend on the choice of the templates, we derived, for those cases where the zCOSMOS-bright resolution was higher than the original Lick resolution, stellar population parameters from the spectra directly convolved down to the Lick resolution, without using the best-fit templates. For these test cases, there are no significant variations from the values reported here. To derive errors on the indices, we used the jackknife technique as explained in Section 4.4.

We derive stellar population parameters as in Onodera et al. (2015), comparing our measured indices with those computed on SSP models with variable $\alpha$-abundances (Thomas et al. 2011). The code makes a 3D grid with a uniform interval of 0.02 dex of the Thomas et al. (2011) model values for age, [Z/H], and $[\alpha/Fe]$. The model spans the parameter space $0.1 < \text{age/Gyr} < 15$, $-2.25 < [Z/H] < 0.67$, and $-0.3 < [\alpha/Fe] < 0.5$ and adopts the Salpeter IMF. The best-fit values of these three stellar population parameters are derived by comparing our corrected indices with the Thomas et al. (2011) model and finding a set of parameters which gives minimum $\chi^2$:

$$\chi^2 = \sum \frac{(I_{\text{obs}} - I_{\text{measured}})^2}{\sigma^2_{\text{Lick}}}$$

where $\sigma_{\text{Lick}}$ is the error bar on our Lick indices computed with the jackknife procedure. This allows us to make use of all available indices at the same time and returns best-fit values for all three free parameters. We use all the indices available in each redshift bin, with few exceptions. The indices Ca4455, H$\delta_A$, Fe5015 and H$\beta$ have been excluded following the recommendation of Thomas et al. (2011). The iron line Fe5015 is a problem due to its proximity to the strongest of the emission lines of the doublet [O iii] at 5007 Å. The Balmer line H$\beta$ is also known to be problematic in deriving ages, as it is affected by fill-in from the emission line more than higher-order Balmer lines (see also Poole et al. 2010; Onodera et al. 2015). Ca4455 is found to be mostly sensitive to Ca abundance (Korn et al. 2005; Thomas et al. 2011) and its use should be therefore taken with caution, as is also demonstrated from comparison with globular clusters (Korn et al. 2005). We also excluded other indices sensitive to the C abundance, namely CN$\gamma$, CN$\delta$ and C2$\gamma$668, as the available models do not use the Mg$\gamma$ lines, the sensitivity of this index is therefore higher than Fe5015 for all redshifts, as these indices lie in the common wavelength range highlighted in yellow in Figure 6. The Balmer lines are the most sensitive indices to age variations. The index Fe4383 is particularly sensitive to the iron abundance and it is also influenced by total metallicity [Z/H] and by the magnesium abundance. Fe4531 and Ca4227 are also sensitive to the total metallicity. Ca4227 is also dominated by variations in Ca and C abundances and show very weak dependence to $\alpha$ variations. The same can be said about the G4300 band, whose main contributors are C, O and Fe abundances. We included H$\delta_A$, Mg$\gamma$, Mg$\delta$, Mg$\beta$, Fe5270, Fe5335 and Fe5406, when available. H$\delta_A$ is sensitive to age variations, despite being more affected by metallicity than H$\gamma$, and is also important for $[\alpha/Fe]$ ratio estimation, as the higher-order Balmer indices like H$\gamma_A$ are sensitive to $\alpha$ enhancement (Thomas et al. 2004). Mg$\gamma$, Mg$\delta$ and Mg$\beta$ are the major indicators of the $\alpha$-elements abundance, although being affected also by [Z/H], C and Fe. The Mg$\delta$ band-pass covers also the Mg$\beta$ lines, the sensitivity of this index is therefore higher than
Table 2

Indices Available in Each Redshift Bin, Bold Those Used to Determine the Stellar Population Parameters

| Redshift | Ca4227 | G4300 | HγA | Hβ | Fe4383 | Ca4455 | Fe4531 | C44668 | Hβ | Fe5015 | MgR | MgB | Fe5670 | Fe5335 | Fe5406 |
|----------|--------|-------|-----|-----|--------|--------|--------|--------|-----|--------|-----|-----|--------|--------|--------|
| 0.2 < z < 0.4 |        |       |     |     |        |        |        |        |     |        |     |     |        |        |        |
| 0.4 < z < 0.6 |        |       |     |     |        |        |        |        |     |        |     |     |        |        |        |
| 0.6 < z < 0.8 |        |       |     |     |        |        |        |        |     |        |     |     |        |        |        |

Table 3

The Ages Derived from Full Spectral Fitting (First Column) and Lick Indices, with Fixed [Z/H] (Second Column) and Free [Z/H] (Third Column)

| Size–Mass cut | 10.5 < Log m < 11 | 11 < Log m < 11.5 |
|---------------|-------------------|-------------------|
| Redshift      | Spectral Fitting  | Lick (fixed [Z/H]) | Spectral Fitting  | Lick (fixed [Z/H]) |
|               | (Log (Age/Gyr))   | (Log (Age/Gyr))   | (Log (Age/Gyr))   | (Log (Age/Gyr))   |
| 0.2 < z < 0.4 | Small             | 0.84 (0.05)       | 0.64 (0.08)       | 0.70 (0.04)       | 0.72 (0.03)       |
|               | Large             | 0.50 (0.10)       | 0.28 (0.03)       | 0.02 (0.04)       | 0.72 (0.04)       |
| 0.4 < z < 0.6 | Small             | 0.49 (0.04)       | 0.42 (0.04)       | 0.40 (0.06)       | 0.56 (0.06)       |
|               | Large             | 0.35 (0.02)       | 0.30 (0.02)       | 0.30 (0.02)       | 0.41 (0.05)       |
| 0.6 < z < 0.8 | Small             | 0.56 (0.03)       | 0.36 (0.02)       | 0.44 (0.04)       | 0.59 (0.06)       |
|               | Large             | 0.42 (0.06)       | 0.24 (0.02)       | 0.18 (0.08)       | 0.65 (0.07)       |

Note: The ages from the Lick indices have been in both cases derived from emission line free spectra. The numbers in parenthesis show the results obtained by fitting the residual emission lines having as base level 1σ above the continuum.

5. RESULTS

Before examining the age difference between large and small galaxies, we look first at the ages of passive galaxies, at each redshift, as a function of their stellar mass, but independent of their size. Thomas et al. (2010) have shown that, at least at redshift zero, there is a clear mass–age relation for QGs, in the sense that more massive QGs are found to be on average older than their less massive counterparts.

In Figure 8 we show the stellar population ages derived from the stacked spectra of our QGs, in bins of stellar mass and redshift, using in each mass-redshift bin all the galaxies, regardless of size. More massive galaxies in our sample (filled black circles) indeed have older ages than less massive ones (empty circles) at any given redshift, in agreement with the previous studies at z ~ 0, but now extending this result out to z = 0.8. The shaded areas represent the difference in ages obtained when applying the correction for emission line fill-in from the continuum level or from 1σ above the continuum. The decrease in age seen in the redshift bin 0.4 < z < 0.6 might conceivably be related to the fact that these galaxies lie in a less dense region of the universe with respect to the known over-densities of COSMOS at 0.3 (Knobel et al. 2009) and 0.7 (Guzzo et al. 2007; Scoville et al. 2007). The over-dense regions are also visible in Figure 3.

Overall this test, reproducing the well-known mass–age relation for QGs and extending this at z > 0 redshifts, gives us confidence that our measurements of ages, at least in a
comparative sense, are fairly robust. We therefore turn to the main question of our paper, which is to establish the age comparison between QGs of different sizes at constant redshift and stellar mass.

5.1. The Stellar Ages of Small and Large QGs

5.1.1. Visual Inspection of Spectral Ratios

Before proceeding to a more quantitative analysis, significant differences between the spectra of large and small galaxies are readily visible “by eye.” In Figure 9 we show the ratio of the spectra of large/small galaxies. At the lower masses, the overall redness of the small spectrum versus the large one is clearly visible, as well as the much deeper calcium H and K lines and the suppression of flux shortward of 4000 Å in the small spectrum. These features are both expected for older ages. The situation with the Balmer lines is less clear as stronger absorption would be expected in the younger object. This effect could however be greatly weakened because of the effects of residual line emission, as seen in Figure 7. Interestingly, the same cannot be said about the higher-mass spectra, with no hint of deeper H and K lines and a stronger 4000 Å break. Similar behavior can be identified for galaxies stacked with both the horizontal cut and the cut along the size–mass relation.

5.1.2. Ages of Small and Large Quenched Galaxies: Quantitative Estimates from the Full Spectral Fitting and Lick Indices Approaches

In Figure 10 we show our best-fit stellar population ages as a function of redshift for the large and small galaxies, obtained with the full spectral fitting technique using pPXF. The spectra have been corrected for emission line contribution visible in the residuals in Figure 7 and fitted with a combination of SSP with solar metallicity. The results show a clear trend in ages for less massive galaxies.

At masses $10.5 < \log M_*/M_\odot < 11$ (left panel) red symbols (small galaxies) are the oldest in each redshift bin, while blue points (large galaxies) are the youngest objects at each redshift, with the age difference between small and large galaxies being especially evident in the lowest redshift bin. The same trends are visible for both the size–mass and the horizontal cuts. The Lick-based analysis that was described in Section 4.5 also shows the same trends, though overall younger ages are returned for both the small and large galaxies, as expected since the model from Thomas et al. (2011) returns luminosity-weighted ages. However, the Lick-derived ages reported in Table 3, which were computed with metallicity as a free parameter, should be treated with some caution because of the strong age–metallicity degeneracy.

For masses $11 < \log M_*/M_\odot < 11.5$, the behavior is different and a clear difference between the ages of galaxies with different sizes is not visible. This lack of difference is present both in the full spectral fitting or in the Lick analysis.

We stress that the results shown here are derived using as the definition of stellar masses the integral of the SFR. However, repeating the analysis using bins of stellar masses computed
including the loss of mass due to the return of material to the ISM due to stellar evolution (which we have argued is not strictly correct), also gives the same qualitative results, as in the lower-mass bin larger galaxies are younger than smaller ones, and no clear trend with ages is visible in the higher-mass bin.

In Figures 11 and 12 we show an example of a Lick index–index diagram, (Fe) (mean obtained with indices Fe5270 and Fe5335) versus Mg2 (see for example Burstein et al. 1984; Carollo & Danziger 1994; Fisher et al. 1996), for the size–mass and the horizontal cut, respectively. We show results only for the two lowest-redshift bins, as at the highest redshift we do not have the wavelength coverage for Mg2, Fe5270 and Fe5335. The model lines are from Thomas et al. (2011). Dotted lines show loci of constant age (2, 4, 6 and 9 Gyr) and constant [α/Fe], all computed at fixed solar metallicity. Small galaxies (red squared points) and large galaxies (blue triangles), while clearly showing differences in ages, do not show differences in [α/Fe] for different sizes. The black points show the index values for galaxies with the same mass and redshift binning as the others in the same plot, but with no split for sizes. All the galaxies appear to have super-solar α to iron ratios ([α/Fe] > 0.3), as expected for massive quiescent galaxies (e.g., Lee & Worthey 2005; Thomas et al. 2005; Choi et al. 2014).

6. DISCUSSION

The analysis presented in the previous sections highlights an important change in behavior of QGs with masses respectively above and below ~10^{11} M_☉: Below this mass threshold, and precisely in our 10.5 < log M/M_☉ < 11 mass bin, the smaller galaxies have systematically older stellar populations than their larger counterparts at each cosmic epoch. This is true for both our definitions of smallness and largeness, i.e., considering both the size–mass cut (relative sizes below and above the evolving size–mass relation), and the horizontal cut (absolute sizes below and above a constant threshold in kpc). In contrast, in our high-mass bin above 10^{11} M_☉, we do not detect any distinct trend for QGs of different sizes. We discuss more in detail below the implications of these different behaviors.
6.1. The Size–Age Relation of $M < 10^{11} \, M_\odot$ QGs at $0.2 \leq z \leq 0.8$

Important implications for the origin of the observed evolution of the average size of QGs at masses $< 10^{11} \, M_\odot$ can be inferred by putting together information on the evolution of the relative number density of smaller and larger galaxies and the fact that, in this mass bin, the former are older than the latter.

In C13, the number density of compact galaxies (i.e., $r_{1/2} < 2$ kpc) remains stable since $z = 1$. This means that either the compact galaxies are growing in size, but some other effect is continuously adding new compact QGs (e.g., Damjanov et al. 2015), keeping the number density constant by replacing the galaxies that grow out of that mass bin, or that the growth of the QG population occurs through the addition of larger, newly quenched galaxies to a stable population of compact galaxies. These scenarios differ in the sizes of the new members of the QG population. In the first case, the new galaxies are compact and the ages of the smaller galaxies should be systematically younger than those of larger galaxies at the same redshift. In the second case, the new galaxies are larger, and the ages of the smaller galaxies would be systematically older.

The results of this paper clearly support the second scenario: compact galaxies are older than larger ones. This spectroscopic result is consistent with the purely photometric trends found by C13 in the same redshift range as here, and also with similar findings of Saracco et al. (2011) at $z \sim 2$, Williams et al. (2016) at $z \sim 1.2$, and of van der Wel et al. (2009) at $z = 0$ (see however Yano et al. 2016 for opposite results obtained with photometry).

Summarizing, the younger ages of larger galaxies plus the evidence that their number density is increasing with cosmic time together push toward the conclusion that progenitor bias is driving a large part of the apparent evolution of the median size of passive galaxies below $10^{11} \, M_\odot$. This in itself reflects the changes in the size–mass relation of their star-forming progenitors (Toft et al. 2007; Buitrago et al. 2008; Kriek et al. 2009). As shown in Lilly & Carollo (2016), half of the size difference between star-forming galaxies and QGs of a given mass is explained by progenitor bias effects in which the sizes of star-forming galaxies of a given mass increase with cosmic time, and the other half by differential radial fading after quenching of the stellar populations of the star-forming progenitors (see also Tacchella et al. 2015a; Carollo et al. 2016). About half of the average size evolution of QG thus emerges naturally by considering progenitor effects.

Our results therefore imply that the increase with cosmic time of the number density of QGs at these “low” masses, which is well quantified in the observed mass functions of QGs at different redshifts (Williams et al. 2009; Ilbert et al. 2010, 2013; Domínguez Sánchez et al. 2011), is mostly driven by the addition of the population of larger, later quenched objects, with smaller galaxies keeping a stable number density though cosmic time.
This is further substantiated by the fact that, between our highest and lowest redshift bins, the Δage of the compact galaxies (selected at constant absolute size through the horizontal cut) is consistent with the passive evolution of their stellar populations, and therefore consistent with the idea that this population has hardly added any new members. In contrast, the smaller difference in ages for larger galaxies is consistent with the idea that new, younger, and, i.e., newly quenched galaxies have been added to this population.

6.2. The Size–Age Relation of $M > 10^{11} \, M_\odot$ QGs at $0.2 \leq z \leq 0.8$

The signatures of progenitor bias are less apparent (if at all) at higher masses above $10^{11} \, M_\odot$ (for us, specifically, in our $11 < \log M_\odot/M_\odot < 11.5$ mass bin). This difference has already been pointed out by C13, and is not unexpected. There is a host of evidence that this mass represents a threshold in the importance of merging; above and below this threshold, QGs have respectively boxy or disky isophotes, cores or cusps, flat or steep metallicity gradients and are slow or fast rotators (see e.g., Bender 1988; Carollo et al. 1993; Faber et al. 2007; Cappellari 2013). Similar results are also found in hydrodynamical simulations within a ΛCDM universe (Hopkins et al. 2009; Feldmann et al. 2010). Furthermore, as stressed in Peng et al. (2010), imposing that galactic evolution obeys a continuity equation requires that very massive passive galaxies at $10^{11} \, M_\odot$ must have generally undergone significant merging after quenching (see their Figure 16).

Also, at these very high masses, C13 report a decrease in the number density of the most compact objects between the high- and low-redshift bin of $\sim$30%, suggesting that at these masses the contribution of newly quenched objects is less important. Using our spectroscopic sample, we (confirm the same trends of C13 for the number density evolution of large and small QGs in the low mass bin, but) find that the number density of small galaxies at high masses remains stable as well, as at the low masses. Our definition of small galaxies at higher masses differs, however, from that in C13, as we adopt a threshold of $<4.5 \, \text{kpc}$ (instead of 2.5 kpc as in C13) to identify small systems. Another element for caution in handling our spectroscopically estimated number densities comes from the possible incompleteness biases of the spectroscopic survey, although zCOSMOS was designed to yield a high and fairly uniform sampling rate across most of the field (about 70%), and has delivered a high success rate in measuring redshifts (Lilly et al. 2009).

6.3. Timescales for Stellar Mass Growth in QGs of Different Masses and Sizes

Our investigation of the [$\alpha$/Fe] abundance ratios for the different subsamples of QGs in our sample, i.e., QGs at different masses below and above $10^{11} \, M_\odot$ and, within the same mass bin, of different sizes, shows that these subsamples have generally enhanced [$\alpha$/Fe] values relative to the solar value. This is a well-known property of massive QGs at redshift $z = 0$ (Carollo et al. 1993; Carollo & Danziger 1994; Lee & Worthey 2005; Thomas et al. 2005; Choi et al. 2014), and has been hinted to be true also at higher redshifts in very small samples (Onodera et al. 2015). Our study proves that, at least at masses above $10^{10.5} \, M_\odot$, such enhanced abundance...
ratios, which support short timescales for the buildup of the stellar populations of QGs, are already achieved at least since $z = 0.6$.

Furthermore, within the errors and in the mass and redshift range of our analysis, we find that the abundance ratios of QGs do not show any substantial dependence on either mass, size or redshift. This suggests rather universal formation timescales for the stellar populations of massive QGs of all sizes over a broad range of stellar masses. This implies short gas consumption timescales for the formation of the bulk of the stellar populations. (Note that use of $\alpha$/Fe to infer quenching timescales depends strongly on the amount of mass produced during quenching.) The result of Thomas et al. (2005), for a dependence of the abundance ratios on stellar mass, could nevertheless be recovered considering two points: (i) their analysis extends down to significantly lower masses, and (ii) that they analyze the $z = 0$ population, which will have added a non-negligible number of galaxies at the masses of our study in the intervening several billion years.

The constancy of the $\alpha$/Fe ratios of the QG populations above and below the $10^{11} M_\odot$ threshold at the redshifts of our study is in contrast to the different behaviors seen for their age-size relation, and is well explained if the most massive population originates from dry mergers of the less massive population. This further strengthens the argument for a major role of such mergers in leading to the emergence with cosmic time of the most massive QGs in the universe at masses $>10^{11} M_\odot$.

7. SUMMARY AND CONCLUDING REMARKS

The observed average size of QGs is about $\sim3$–5 times larger today than at $z \sim 2$. There are two main scenarios which have been proposed to explain this evolution: the size growth of individual galaxies or the progenitor bias introduced if newly formed members of the population are larger than the previous members. In this work, we measure the stellar ages of QGs in order to distinguish between these two scenarios which make quite different predictions for the variation of stellar population age with size. If the driver of this evolution is the addition of large newly quenched objects then larger QGs will be younger. If the driver is the size growth of individual galaxies, and small galaxies are being replaced by newly quenched objects, then the larger galaxies will be older.

In light of the fact that, at any epoch and stellar mass, star-forming galaxies are on average larger than passive galaxies, purity (and completeness) of the QG samples is clearly crucial when attempting the measurement of their stellar ages (see, e.g., Keating et al. 2015). The polluting presence of star-forming galaxies in QG samples would bias the latter toward larger sizes, while also biasing their age estimates toward younger values. In this work we have selected our QG samples using galaxies securely identified as quiescent. Starting from the 20k zCOSMOS-bright catalog, we selected galaxies with absent, or very weak, emission lines. We stacked the spectra in bins defined in size, stellar mass and redshift, to study the average stellar population properties of the sample galaxies. Two binning schemes were used in size, a relative one normalized to the evolving mean size–mass relation, and an absolute one constructed at fixed physical size. We then used pPXF to derive best-fit ages from BC03 solar SSP templates. To further check our age results, we also computed absorption line strengths following the wavelength definitions of feature and pseudo-continua of the Lick system of spectral line and derived values for age, [Z/H], and $\alpha$/Fe, comparing our results to the Thomas et al. (2011) models. Our robust spectroscopic selection for QGs is much cleaner than the more frequently adopted color–color selections; in addition, we carefully checked the spectrum of each object by visual inspection, to ensure the absence of star formation tracers. We are thus confident that contamination by star-forming galaxies is negligible in the results that we have discussed above.

Reassuringly, the average age of QGs (not discriminating in size) increases with cosmic time. Turning to the ages as a function of size, we find that the $10^{11} M_\odot$ mass scale is a “threshold” above and below which the size–age relation changes behavior, as already pointed out in C13. Below $10^{11} M_\odot$, larger galaxies are systematically younger than smaller ones. The $\Delta$age between small and large galaxies becomes more significant toward lower redshift. The $\Delta$age from the highest to the lowest redshift bin in the small-size QG population is in good agreement with a passive evolution of its stellar populations. The younger ages of the larger galaxies at each redshift argue for newly quenched objects to be systematically larger at later epochs. This trend is visible using both our size binning schemes. We conclude that progenitor bias is a major and possibly the dominant component of the observed evolution in the average sizes of QG at these masses.

Above $10^{11} M_\odot$, where dry mergers are expected to play a major role in imprinting the well-known “dissipationless” features that are observed at $z = 0$ in this ultra-massive population, there is indeed no clear trend between ages and sizes. Size growth of individual galaxies through dry mergers is the most likely channel for the observed growth of the average size of the QG population at this top-mass end.

The confirmation of a “transition” mass around $10^{11} M_\odot$ for the size–age behavior—and thus for the dominant role of progenitor bias at low masses and dry mergers at high masses in driving the observed average size growth of QGs with time—highlights the fundamental importance of sample selection and of tuning the interpretation of the data to the specific sample selection. For example, Zanella et al. (2016) report that the size–age relation in their sample of, in quotation, $M > 4.5 \times 10^{10} M_\odot$ QGs at $z \sim 1.5$ supports the mergers interpretation. A quick inspection of their analysis shows, however, that $\sim80\%$ of their galaxies actually have masses above $10^{11} M_\odot$. Therefore, their result is better commented on as holding for this top-mass end population, which puts their result in agreement with our work.

Interestingly, the $\alpha$-to-iron abundance ratio of the stellar populations of QGs at all masses within the $10^{10.5}$–$11.5 M_\odot$ window is rather constant since $z = 0.6$. This ratio should reflect the formation timescales for the stellar populations in these systems. The constancy of the measured $[\alpha$/Fe] ratio thus suggests similar timescales, independent of galaxy size, across the whole $10^{10.5}$–$11.5 M_\odot$ mass range for the galaxy population that has already quenched by our lowest redshift bin at $z \sim 0.3$, consistent with the idea that the most massive galaxies above $10^{11} M_\odot$ are formed by mergers of lower-mass galaxies.

We acknowledge support from the Swiss National Science Foundation.
