The Astrophysical Journal, 876:135 (26pp), 2019 May 10 https://doi.org/10.3847/1538-4357/ab14f2

Optically Faint Massive Balmer Break Galaxies at $z > 3$ in the CANDELS/GOODS Fields

Belén Alcalde Pampliega[1,2], Pablo G. Pérez-González[1,3], Guillermo Barro[4], Helena Domínguez Sánchez[5], M. Carmen Eliche-Moral[6], Nicolás Cardiel[1], Antonio Hernán-Caballero[1], Lucía Rodriguez-Muñoz[7], Patricia Sánchez Blázquez[8], and Pilar Esquej[9]

[1] Departamento de Física de la Tierra y Astrofísica, Faultad de CC Físicas, Universidad Complutense de Madrid, E-2840 Madrid, Spain
[2] Isaac Newton Group of Telescopes (ING), Apto. 321, E-38700 Santa Cruz de la Palma, Canary Islands, Spain
[3] Centro de Astrobiología (CAB, INTA-CSIC), Carretera de Ajalvir km 4, E-28850 Torrejón de Ardoz, Madrid, Spain
[4] Astronomy Department, Department of Physics, University of the Pacific, Stockton, CA 95211, USA
[5] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[6] Instituto de Astrofísica de Canarias, Calle Vía Láctea, s/n E-38205, La Laguna, Tenerife, Spain
[7] Dipartimento di Fisica e Astronomia, Università di Padova, vicolo dell'Osservatorio 2, I-35122 Padova, Italy
[8] Departamento de Física Teórica, Universidad Autónoma de Madrid, E-28049 Cantoblanco, Spain
[9] Herschel Science Centre, ESA, Villafranca del Castillo, Apartado 78, E-28691 Villanueva de la Cañada, Spain
Received 2018 June 5; revised 2019 March 13; accepted 2019 March 26; published 2019 May 13

Abstract

We present a sample of 33 Balmer break galaxies (BBGs) selected as *Hubble Space Telescope*/F160W dropouts in the deepest CANDELS/GOODS fields ($H \gtrsim 27.3$ mag) but relatively bright in *Spitzer*/IRAC ([3.6], [4.5] < 24.5 mag), implying red colors (median and quartiles: $\langle H - [3.6] \rangle = 3.1^{3.4}_{2.8}$ mag). Half of these BBGs are newly identified sources. Our BBGs are massive ($\langle \log(M/M_\odot) \rangle = 10.8^{11.0}_{10.4}$), high-redshift ($\langle z \rangle = 4.8^{5.1}_{4.4}$), dusty ($\langle A(V) \rangle = 2.0^{2.0}_{1.5}$ mag) galaxies. The spectral energy distributions (SEDs) of half of our sample indicate that they are star-forming galaxies with typical specific star formation rates (SFRs) of 0.5–1.0 Gyr^{-1}, qualifying them as main-sequence (MS) galaxies at $3 < z < 6$. One-third of these SEDs indicate the presence of prominent emission lines (Hβ + [O III], Hα + [N II]) boosting the IRAC fluxes and red colors. Approximately 20% of the BBGs are very dusty ($A(V) \sim 2.5$ mag) starbursts with strong mid-to-far-infrared detections and extreme SFRs ($\mathrm{SFR} > 10^3\, M_\odot\, \mathrm{yr}^{-1}$) that place them above the MS. The rest, 30%, are post-starbursts or quiescent galaxies located $>2\sigma$ below the MS with mass-weighted ages older than 700 Myr. Only two of the 33 galaxies are X-ray-detected active galactic nuclei (AGNs) with optical/near-infrared SEDs dominated by stellar emission, but the presence of obscured AGNs in the rest of the sources cannot be discarded. Our sample accounts for 8% of the total number density of $\log(M/M_\odot) > 10$ galaxies at $z > 3$, but it is a significant contributor (30%) to the general population of red $\log(M/M_\odot) > 11$ galaxies at $4 < z < 6$. Finally, our results point out that one of every 30 massive $\log(M/M_\odot) > 11$ galaxies in the local universe was assembled in the first 1.5 Gyr after the big bang, a fraction that is not reproduced by state-of-the-art galaxy formation simulations.

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

Supporting material: figure set

1. Introduction

Understanding how galaxies form and evolve is one of the central challenges of modern astronomy. In the current ΛCDM paradigm, dark matter halos are the primary structures that provide seeds for gas collapse and allow the baryonic growth of galaxies. Dark matter halos assemble primarily in a hierarchical manner, with low-mass halos forming early and merging to produce more massive halos as they move down to lower redshifts (Kauffmann et al. 1993; Reed et al. 2003). Consequently, the most massive galaxies are expected to appear in massive halos at lower redshifts following a similar hierarchical assembly. However, observational studies suggest that massive galaxies ($\log(M/M_\odot) > 11$) in the local universe form rapidly in strong bursts of star formation at early times (e.g., Pérez-González et al. 2008; Mancini et al. 2009; Ilbert et al. 2010, 2013; Brammer et al. 2011; Caputi et al. 2011; Muzzin et al. 2013; Tomczak et al. 2014; Grazian et al. 2015). Similarly, many surveys have identified a substantial population of massive galaxies at redshifts up to $z \sim 4$, when the universe was only 1.5 Gyr old (Mobasher et al. 2005; Wiklind et al. 2008; Caputi et al. 2012). Although some of those studies present evidence for evolved stellar populations or suppressed star formation (Fontana et al. 2009; Nayyeri et al. 2014; Spitler et al. 2014; Straatman et al. 2014), the existence of fully quiescent galaxies at very high redshifts is still controversial (Straatman et al. 2016; Glazebrook et al. 2017; Hill et al. 2017; Marsan et al. 2017; Simpson et al. 2017; Schreiber et al. 2018). Determining when the first massive galaxies emerged and characterizing the evolution of their number density is particularly important to improve our picture of galaxy evolution and constrain galaxy formation models. A major complication in addressing these questions is gathering a complete, robust, and unbiased census of massive galaxies up to the highest redshifts possible.

The advent of deep surveys with the WFC3 camera on the *Hubble Space Telescope* (*HST*) has significantly expanded our census of distant galaxies up to $z \sim 10$ (e.g., Bouwens et al. 2015; Oesch et al. 2018). However, the use of near-infrared (NIR) observations implies that the sample selection at $z \gtrsim 3$ is based on the rest-frame UV emission of the galaxies, which is particularly sensitive to the effects of dust attenuation and biased toward the detection of blue systems. Consequently, while UV-based selection techniques, such as the Lyman break dropout (LBG; Madau et al. 1996) or the search for Lyα emitters

(LAEs), are very effective at identifying blue, (typically) low-mass star-forming galaxies (SFGs; Steidel et al. 2003; Giavalisco et al. 2004a), these methods are strongly biased against red, dusty, or evolved galaxies, which typically make up most of the massive galaxy population at mid-to-high redshifts (e.g., Brammer et al. 2011; Spitler et al. 2014; Caputi et al. 2015). Thus, rest-frame UV-selected samples at high redshift are likely incomplete, missing massive red galaxies that could potentially be identified with observations at longer wavelengths.

At $z > 3$, the strongest spectral features in the rest-frame optical continuum of galaxies, the 4000 Å and Balmer breaks, are shifted redward of $\sim$1.5 μm, thus making mid-to-far-IR (or even radio) observations essential to identifying the presence of massive galaxies. A number of different selection techniques based on "extremely" red mid-IR (MIR) colors (e.g., Kennicutt 1998; Fontana et al. 2006; Rodighiero et al. 2007; Huang et al. 2011; Caputi et al. 2012; Nayyeri et al. 2014; Schreiber et al. 2016); Wang et al. 2016 and/or bright far-IR (FIR) or submillimeter emission (Casey et al. 2012; Riechers et al. 2013; Vieira et al. 2013) have successfully identified a hidden population of massive galaxies at redshifts $z \gtrsim 3$ that are missing from even the deepest *HST* surveys. The extremely red colors of these galaxies can indicate either a heavily obscured burst of star formation (usually accompanied by strong FIR emission from the heated dust) or a quiescent, passively evolving galaxy (e.g., de Barros et al. 2014). Overall, the intrinsically faint optical-to-NIR fluxes of these galaxies, typically coupled with high dust obscurations, make the modeling of their spectral energy distributions (SEDs) very challenging (see, e.g., Schaerer et al. 2013; de Barros et al. 2014). Consequently, the inferred redshifts, stellar population properties, and star formation rates (SFRs) are quite uncertain (e.g., Michałowski et al. 2010; da Cunha et al. 2015).

A way forward to overcome the mentioned limitations when studying massive galaxies at high redshift is to gather large unbiased samples from the deepest cosmological surveys carried out in the MIR. These surveys must be very deep and cover relatively wide areas, since $z > 3$ massive galaxies are relatively scarce systems (the typical number density is $\sim$0.1 arcmin^{-2}; Caputi et al. 2012). This methodology would provide more robust, statistically significant constraints on the overall properties of the oldest massive galaxies, including the contribution of red MIR-detected sources to the massive end of the $z > 3$ galaxy population. Simultaneously, by using MIR deep surveys, we can also characterize the incompleteness of our current mass-limited samples, which are typically based on NIR selections with *HST*. In addition, follow-up observations of bona fide candidates for massive high-z galaxies in new spectral ranges (e.g., with ALMA or the upcoming *James Webb Space Telescope* (*JWST*)) can alleviate the SED fitting limitations and provide more precise values of their redshifts and stellar population properties, which are essential to have a complete view of the very early phases of massive galaxy formation.

In this context, in the present paper, we aim to obtain a (more) complete sample of massive galaxies at $z \gtrsim 3$. In order to achieve this goal, we focus our analysis on the search for and characterization of MIR-bright, NIR-faint galaxies that might have been missed in the *HST*-based, NIR-selected catalogs presented by the Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS; Guo et al. 2013) and 3D-*HST* (Skelton et al. 2014) surveys. In particular, we present the results of an Infrared Array Camera (IRAC) 3.6 + 4.5 μm selection and multiwavelength analysis of a sample of red massive galaxies at $z > 3$, i.e., probing the massive galaxy population formed roughly between the first and second Gyr in the lifetime of the universe. Hereafter, we will call these objects Balmer break galaxies (BBGs). Our sample of BBGs has been built by searching for extremely red objects (EROs) in the *Spitzer* 3.6 and 4.5 μm IRAC images that are not detected in the F160W CANDELS deep observations carried out over the 330 arcmin^2 in the GOODS-N and GOODS-S fields. At $3 < z < 7$, the wavelengths in the 3–5 μm range (and, therefore, the 3.6 and 4.5 μm IRAC bands) are a robust proxy for a deep and roughly constant stellar mass cut (Fontana et al. 2006; Caputi et al. 2009; Ilbert et al. 2010), allowing us to built a mass-complete sample of BBGs. After presenting our method to identify these IRAC-selected, NIR-faint massive galaxies at $z > 3$, we compare their colors, redshifts, and other stellar properties to those of NIR-faint and $H - [3.6]$ color- and mass-selected galaxies in the CANDELS catalog to compare and characterize the regions of the redshift—stellar mass parameter space that is being populated by our newly identified BBGs. Lastly, we study the SFRs and stellar population properties of all of the BBGs (which are hard to constrain due to their intrinsically faint nature in all but the MIR bands), and we analyze the role of these galaxies in the context of galaxy evolution, especially at the high-mass end.

This paper is organized as follows. In Section 2, we present the data set available in the GOODS fields. The procedure followed to select our sample of IRAC-bright BBGs and the methods applied for searching for counterparts in other bands are presented in Sections 3 and 4, respectively. In Section 4 we describe our estimations of the photometric redshifts and stellar property derivation, together with the SED fitting procedure. In Section 5 we present our results and discuss the properties of different subsamples of BBGs, including a comparison with the literature. Finally, in Section 6, we summarize our findings and present the conclusions. Appendix A contains a detailed description of the photometric measurements in the optical and NIR bands. Appendix B describes the analysis of the comparison samples used throughout the paper. And Appendix C shows the SEDs and postage stamps of all of the BBGs presented in this work.

We adopt a cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. All magnitudes refer to the AB system (Oke & Gunn 1983). The initial mass function (IMF) is assumed to be that presented in Chabrier (2003).

2. Multiwavelength Data Set

This work presents the search for and analysis of a sample of BBG candidates in two of the deepest cosmological fields, namely, GOODS-N (R.A. = $12^h36^m55^s$, decl. = $+62°14'15''$) and GOODS-S (R.A. = $3^h32^m31^s$, decl. = $-27°48'54''$; Giavalisco et al. 2004a). In the following, we describe the multiband data sets available in these fields that we have used in our analysis. We limit our search for BBGs to the area surveyed by CANDELS (see Section 2.2), which is also covered by other surveys probing wavelengths from the UV to the FIR and millimeter. In total, we work with a sky region of 330 arcmin^2 (160 arcmin^2 in GOODS-N and 170 arcmin^2 in GOODS-S).

2.1. Spitzer/IRAC Data

Our BBG candidate search is primarily based on deep MIR images taken in the GOODS fields by the *Spitzer* IRAC from 3.6 to 8.0 μm. Here we make use of the deepest multi-epoch mosaics in these regions, which are based on observations from the GOODS/IRAC survey (Dickinson & Giavalisco et al. 2003) and the *Spitzer* Extended Deep Survey (SEDS; Ashby et al. 2013). The four IRAC bands centered at 3.6, 4.5, 5.6, and 8.0 μm have 5σ limiting magnitudes of 26.1, 25.5, 23.5, and 23.4 mag, respectively.

2.2. *CANDELS* HST *WFC3 NIR Data*

Deep NIR imaging of the GOODS fields was obtained with the *HST*/WFC3 camera as part of CANDELS (Grogin et al. 2011; Koekemoer et al. 2011). Here we make use of the publicly available F105W, F125W (*J*), F140W (only in GOODS-S), and F160W (*H*) mosaics, as well as the *H*-band selected galaxy catalogs in both fields presented in Guo et al. (2013) for GOODS-S and G. Barro et al. (2019, in preparation) for GOODS-N. These catalogs include UV-to-NIR multiband photometry, as well as photometric redshifts and stellar population properties based on the fitting of the SEDs. See also Galametz et al. (2013) for more details on the data reduction and the creation of the catalogs. Note that the GOODS fields are the only two of the five CANDELS fields that have the deepest layer of NIR observations, reaching a 5σ sensitivity limit of 27.6 mag (Grogin et al. 2011).

We also use *K*-band imaging taken with the VLT/HAWK-I instrument in the GOODS-S field (HUGS survey; Fontana et al. 2014). Similarly, in GOODS-N, we use *K*-band imaging from CFHT/WIRCam (Kajisawa et al. 2011). The 5σ depths of these data sets are 26.3 and 24.7 mag, respectively.

2.3. HST/*ACS and SHARDS Optical Data*

In addition to the NIR and MIR imaging described above, we make use of the deep optical mosaics taken with the *HST*/Advanced Camera for Surveys (ACS) in the GOODS fields as part of the GOODS and CANDELS surveys. There are publicly available mosaics in five bands: F435W (*b*), F606W (*v*), F775W (*i*), F850W (*z*), and F814W. They reach 5σ point-source sensitivity limits of 28.5, 28.8, 28.1, 27.6, and 28.4 mag (Giavalisco et al. 2004b).

Furthermore, we also use imaging data from the GTC Survey for High-z Absorption Red and Death Sources (SHARDS; Pérez-González et al. 2013), which consists of observations in 25 contiguous medium-band ($R \sim 50$) filters covering the spectral range 500–950 nm, reaching an AB magnitude of 27.0, at least at the 3σ level.

Apart from the publicly available single-band mosaics, we have also created two stacked images by combining either all of the *HST* bands (optical and NIR) or all of the SHARDS medium bands. The goal of building these stacks is to increase the limiting depth of our search for optical counterparts to our MIR-bright, NIR-faint (or even undetected) BBGs.

2.4. *FIR and Submillimeter Data*

The GOODS fields have also been observed in the FIR by *Spitzer* and *Herschel* as part of the GOODS (Dickinson & Giavalisco et al. 2003), GOODS-*Herschel* (Elbaz et al. 2011), and PACS Evolutionary Probe (PEP; Berta et al. 2011; Lutz et al. 2011) surveys. Here we make use of the *Spitzer*/MIPS 24 and 70 μm mosaics presented in Pérez-González et al. (2008), the *Herschel* PACS 100 and 160 μm, and the SPIRE 250, 350, and 500 μm catalogs described in Elbaz et al. (2011) and Magnelli et al. (2013). The 5σ limits of these surveys are 30 (30) μJy, 1.2 (1.2) μJy, 1.7 (1.5) mJy, 3.6 (3.2) mJy, 9 (8) mJy, 12 (11) mJy, and 13 (11) mJy for 24, 70, 100, 160, 250, 350, and 500 μm in GOODS-N (GOODS-S).

We also search for additional FIR data in the following surveys: SCUBA (Borys et al. 2005; Pope et al. 2005) in GOODS-N and LABOCA (Weiß et al. 2009), LESS, and ALMA follow-up ALESS (Hodge et al. 2013; Karim et al. 2013) in GOODS-S. At millimeter wavelengths, the AzTEC 1.1 mm (Perera et al. 2008; Penner et al. 2011), MAMBO 1.2 mm (Borys et al. 2005), and GISMO 2 mm (Staguhn et al. 2014) surveys are available in GOODS-N. These surveys reach sensitivities of 2–5 mJy corresponding to L(IR) of $\gtrsim 10^{12}$ for $z \sim 4$.

2.5. *X-Ray*

We have used X-ray data from the *Chandra* 2 Ms source catalog by Alexander et al. (2003), covering the entire surveyed region of the F160W mosaic in GOODS-N, and the 4 Ms catalog from Xue et al. (2011) in GOODS-S. The on-axis sensitivity limits in the soft/hard bands are $2.5 \times 10^{-17}/1.4 \times 10^{-16}$ and $9.1 \times 10^{-18}/5.5 \times 10^{-17}$ erg cm^{-2} s^{-1} in 2 and 4 Ms, respectively. These fluxes correspond to X-ray luminosities L_X(2–10keV) $> 10^{43}$ erg s^{-1} (L_X(2–10 keV) $> 10^{44}$ erg s^{-1}) for $z > 3$ ($z > 4$), according to Ueda et al. (2014); see also Padovani et al. (2017).

3. Selection of BBGs at $z > 3$

In this section, we describe the selection technique used to identify candidates for massive high-redshift galaxies. Briefly, candidates are identified by searching for relatively bright IRAC sources that have no optical/NIR counterparts in very deep *HST* imaging (i.e., dropouts). This technique is an extension of the classical ERO (see, e.g., McCarthy 2004) or distant red galaxy (DRG; Franx et al. 2003; Brammer & van Dokkum 2007) methods used to select red galaxies at $z \sim 2$ and has been used in different papers to identify massive galaxies at higher redshifts (Huang et al. 2011; Caputi et al. 2012; Stefanon et al. 2015; Wang et al. 2016). All of these methods use a single-color selection threshold to identify extremely red galaxies at $z \gtrsim 3$ with strong breaks around the Balmer/4000 Å rest-frame region. At these redshifts, the Balmer/4000 Å break lies in between the *HST* F160W and the *Spitzer* IRAC 3.6 μm. Therefore, we use the $H - [3.6]$ color to search for BBGs. Using a single color to identify a spectral feature often exhibits degeneracies. In our case, a strong Balmer or 4000 Å break can be explained by either evolved stellar populations or younger stellar populations with significant dust obscuration (see, e.g., de Barros et al. 2014). Throughout the paper, we will use the name BBG to refer to all of the candidates for massive high-redshift galaxies identified by our color selection, regardless of their intrinsic stellar ages. In Section 5, we infer redshifts and stellar population properties from the fitting of their SEDs and discuss their typical ages and dust attenuations. We will show that the average redshift distribution of the BBGs peaks at $z \gtrsim 4$, in agreement with the prediction from the color selection. Overall, our sample of

sources identified with a red $H-[3.6]$ color exhibits intermediate ages of ∼1 Gyr (with strong Balmer breaks) and relatively high dust attenuations ($A(V) \gtrsim 1.5$ mag). Only a handful of BBGs have ages consistent, within the large uncertainties, with having more evolved stellar populations. We note, however, that at $z \gtrsim 4$, where most of our BBGs seem to lie, there is little room for stellar populations older than 1 Gyr (the age of the universe, $z = 4$, is around 1.5 Gyr), so the observed colors are most likely due to Balmer jumps rather than 4000 Å breaks (justifying our nomenclature). See Dunlop et al. (2013) and references therein for a discussion about BBGs at high redshift.

3.1. The F160W Dropout Search

Our selection technique is based on two conditions. The BBG candidates are required to be bright in the first two channels of IRAC, [3.6] and [4.5] ⩽ 24.5 mag, and they must be undetected (dropouts) in the *HST*/F160W ($H \gtrsim 27$ mag) mosaics (according to publicly available catalogs). The search for dropouts in F160W relies on the multiband catalogs published by Guo et al. (2013) and G. Barro et al. (2019, in preparation) for the CANDELS GOODS-S and GOODS-N regions, respectively (we also checked the catalogs published by the 3D-*HST* team; Skelton et al. 2014). The IRAC photometry presented in these works for *H*-band detected sources is based on a point-spread function (PSF) matching technique, TFIT (Laidler et al. 2007), which is also used and described extensively in Galametz et al. (2013). Briefly, TFIT is used to generate a model of the IRAC image by convolving the high spatial resolution *HST*/F160W mosaic with the appropriate PSF transformation kernel. Then, the fluxes on the resulting "template" image are scaled to those of the galaxies in the IRAC frame on a galaxy-by-galaxy basis, taking into account the contamination by neighboring sources. The individual scaling factors provide PSF-matched $H-[3.6]$ colors for all *H*-band detected galaxies. Lastly, TFIT subtracts the scaled "template" image from the original IRAC mosaic, creating a residual frame that is used to verify the quality of the source extraction and flux measurements. In our case, we use these residual images to search for potential *H*-band dropouts with bright IRAC magnitudes. Figure 1 illustrates this procedure, highlighting the detection of a BBG candidate.

3.2. Masking and BBG Candidate Extraction

Before searching for BBG candidates in the residual IRAC image, we performed a series of iterative masking procedures to smooth the image. This cleaning procedure is necessary because the IRAC residual image often contains spurious flux coming from saturation artifacts, as well as from the wings and cores of bright sources that are not properly subtracted. This problem is typically caused by slight changes (at the 5% flux level) in the IRAC's PSF along the mosaic.

We applied three different cleaning masks to the residual IRAC images. First, to avoid detecting *H*-band bright sources, we created a mask including pixels above a threshold flux (6×10^{-3} and 3×10^{-2} Jy pixel^{-1} in GOODS-N and GOODS-S, respectively) in the convolved images, which is equivalent to masking *H*-band bright pixels. Second, we masked the regions contaminated by the brightest ($[3.6] < 20$) stars in the field using

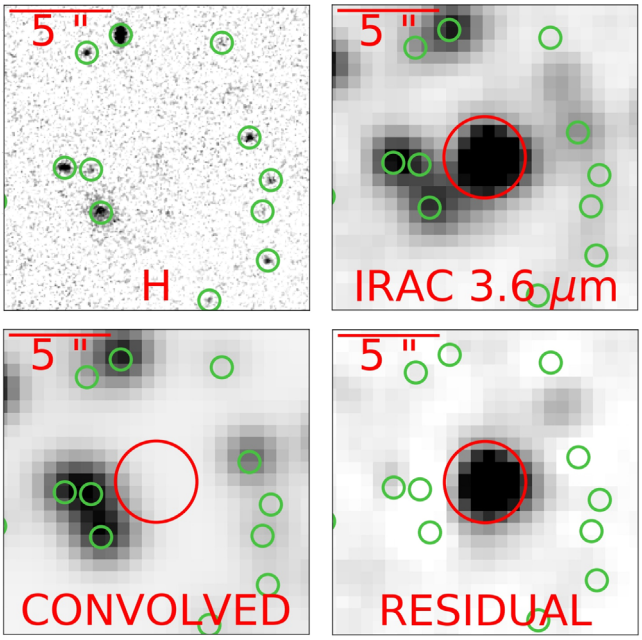

Figure 1. Example of the deconvolution process followed in the IRAC 3.6 and 4.5 μm images (see G. Barro et al. 2019, in preparation). Top left: cutout of the *H*-band image centered in the position of our source GDN_BBG02 (see Table 1). Its position is also marked in all panels with a 2″ radius red circle. The *H*-band detected sources from the CANDELS public catalog (G. Barro et al. 2019, in preparation; see also the 3D-*HST* catalog in Skelton et al. 2014) are circled in green. Top right: IRAC 3.6 μm image showing all of the *H*-band sources detected in the region. Bottom left: TFIT "template" image built by convolving the *HST* image to the 3.6 μm resolution and then scaling all *H*-band detected sources to reproduce the flux measured in the original IRAC image. Bottom right: residual image obtained by subtracting the scaled "template" from the IRAC science frame. Note how the *H*-band dropout appears clearly defined in this residual image (jointly with another fainter dropout to the NW beyond our IRAC magnitude cut).

circular masks with magnitude-dependent radii,

$$r = -21.8 \times H + 380.8, \quad \text{for } 14 < H < 16, \tag{1}$$

$$r = -4.6 \times H + 122.4, \quad \text{for } 16 < H < 20, \tag{2}$$

with r expressed in arcseconds.

Lastly, we masked artifacts (negative fluxes in the convolved images) that appeared as a result of the TFIT convolution process. These three masks were applied to the IRAC residual image, replacing the affected pixels by the median background calculated in a 1′ region around each source.

We also applied a mathematical morphology (MM) method to the regions around *H*-band bright sources to avoid extra flux arising from their wings. We iteratively generated 1 pixel width outlines applying dilation (Lea & Kellar 1989; Lybanon et al. 1994; Maccarone 1996), subtracted the median flux, and added the median background value to each contour. Figure 2 exemplifies our cleaning procedure, showing the environment of one of the *H*-band dropouts in the raw and residual IRAC images, as well as in the final cleaned image.

3.3. Dual Detection in IRAC 3.6 and 4.5 μm

After applying the masks to the IRAC residual images, we used SExtractor (Bertin & Arnouts 1996) to search for BBG candidates. We required the BBG candidate to be detected in both IRAC 3.6 and 4.5 μm. Thus, we ran SExtractor separately

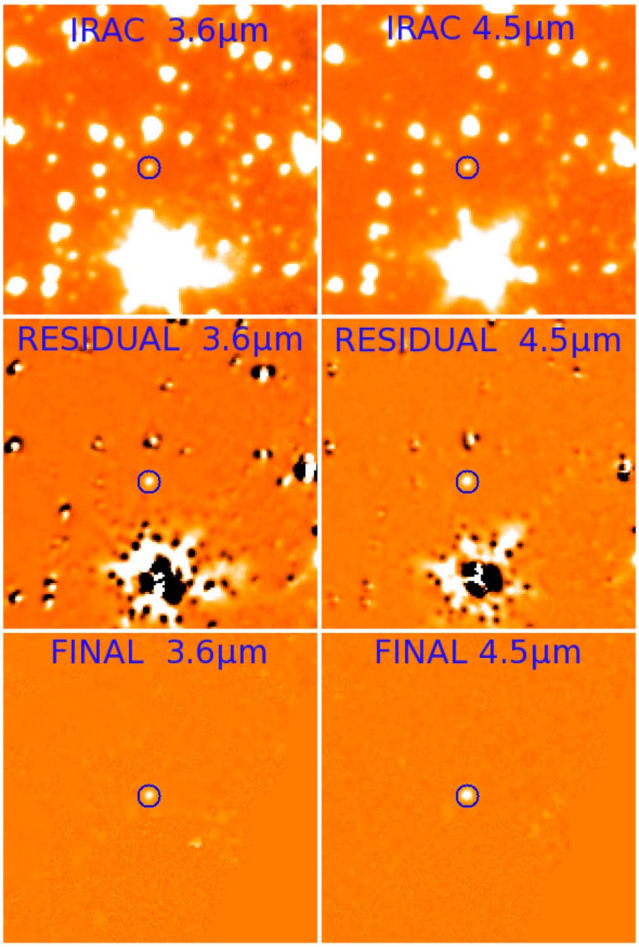

Figure 2. Example of the detection of one of the BBGs as an *H*-band dropout using our technique, which combines convolution of *H*-band images to the PSF of IRAC images using TFIT, followed by a cleaning algorithm and a scaling on a galaxy-by-galaxy basis. The left and right columns show this procedure on the IRAC 3.6 μm and 4.5 μm residual images, respectively. The $40'' \times 40''$ cutouts are centered in the position of GDN_BBG03 (blue circle). In each column, the first row shows the original IRAC image. The second row presents the residual image obtained by subtracting the "template" built by TFIT from the IRAC science frame. The third row shows the final residual images after masking artifacts.

on the two cleaned residual images, and then we cross-matched the resulting catalogs within a $1''$ radius, keeping only the sources in common. The dual 3.6 + 4.5 μm detection provides a more robust selection and further reduces the impact of spurious detections around artifacts.

In addition, we required all BBG candidates to be IRAC-bright ([3.6] and [4.5] $\leqslant$ 24.5 mag) and not included in the *H*-band selected CANDELS catalogs. To account for the latter, we removed all of the sources in the dual 3.6 + 4.5 μm catalog with an F160W counterpart in any of the GOODS-N or GOODS-S CANDELS catalogs (and 3D-*HST* catalogs presented in Skelton et al. 2014) identified within a search radius of $0\farcs5$. Therefore, all of our final selected sources are, in principle, dropouts in the *H* band.

Finally, we visually inspected all of the remaining BBG candidates in the IRAC residual images to remove any possible remaining artifacts due to contamination from bright and/or nearby objects. We also discarded three sources that qualified as dropouts but were found to lie at $z < 2$, while in this paper, we are interested in $z > 3$ galaxies. The final catalog contains 33 bona fide BBG candidates (17 in GOODS-N and 16 in GOODS-S). By sample construction, our sources are relatively bright in the first two IRAC channels and extremely faint (dropouts) blueward of 2 μm. The coordinates and magnitudes (see next section) of these sources are given in Table 1.

4. SEDs, Photometric Redshifts, and Stellar Population Properties of the IRAC BBG Candidates

In this section, we describe the multiwavelength characterization of the BBG candidates presented in the previous section. This characterization consists of the measurement of the SED of each source using the data described in Section 2. We also discuss the estimate of their photometric redshifts and stellar population properties based on the fitting of these multiband SEDs to stellar population synthesis models.

4.1. Multiband Photometry from NIR-to-FIR Data and Deep Optical Stacks

We measured multiwavelength photometry for the 33 IRAC-selected BBGs following the methods described in depth in Pérez-González et al. (2008) and Barro et al. (2011a). All 33 BBGs are in fact detected in the IRAC catalogs of Pérez-González et al. (2008) for the GOODS regions. We chose not to adopt the SEDs of Pérez-González et al. (2008) for the BBGs. Instead, we repeated the same photometric procedure to take advantage of the new and deeper mosaics in the region (see Section 2 for more details).

First, we measured the photometry in the four IRAC bands using the residual IRAC images provided by TFIT and described in the previous section. The use of these "cleaned" images reduces the flux contamination due to bright neighbors, stars, or image artifacts. For each object, we considered several aperture radii, ranging from $0\farcs75$ to $2''$, in order to to maximize the signal-to-noise ratio (S/N) of the measurement and reduce the contamination from nearby sources. We applied the appropriate aperture correction to the measurement for each radius. The typical scatter between measurements using different apertures is 0.13, 0.16, 0.16, and 0.15 mag in IRAC 3.6, 4.5, 5.8, and 8.0 μm. The new photometry is fully consistent with the values in Pérez-González et al. (2008). By definition of the sample, our BBGs are detected in the first two IRAC bands, and ~80% of them are detected in the four IRAC channels.

Redward of the IRAC channels ($\lambda > 8$ μm), we measured the photometry in the *Spitzer*/MIPS, *Herschel*/PACS, and SPIRE FIR bands. As discussed in Section 4.4, these fluxes can be used to characterize the dust emission properties of the BBGs. A total of eight sources (~25% of the sample) were detected in MIPS 24; five of those MIPS sources were detected by PACS and four by SPIRE. We also searched for counterparts in the submillimeter catalogs available in the GOODS fields (see Section 2). Among the eight MIPS emitters, three BBGs were detected at 850 μm, and one of them was also detected at 1200 μm. Additionally, two BBGs were detected in X-rays.

Last, we focus on the more difficult task of trying to characterize the SEDs blueward of $\lambda \lesssim 3.6$ μm and measuring the photometry in the optical, NIR, MIR, and FIR bands. We find that ~40% of the BBGs are detected in the *K*-band images, presenting

Table 1
Observed Properties of Our Sample of BBGs at $z > 3$

	ID	R.A. J2000	Decl. J2000	F160W (AB mag)	IRAC CH1 (AB mag)	IRAC CH2 (AB mag)	IRAC CH3 (AB mag)	IRAC CH4 (AB mag)	MIPS 24 (μJy)	Comments[a]
1	GDN_BBG01	189.23577500	62.20206944	>27.73	24.51 ± 0.14	23.76 ± 0.07	23.57 ± 0.12	23.09 ± 0.09	69.9 ± 6.0	FIR
2	GDN_BBG02	189.30782083	62.30743889	26.51 ± 0.16	21.77 ± 0.04	22.00 ± 0.04	21.39 ± 0.04	20.91 ± 0.04	69.1 ± 6.2	FIR, SMG, W16
3	GDN_BBG03	189.18331250	62.32746389	27.25 ± 0.25	22.86 ± 0.05	22.32 ± 0.04	21.82 ± 0.04	21.41 ± 0.04	29.5 ± 5.8	W16
4	GDN_BBG04	189.43552083	62.29016111	>27.20	24.03 ± 0.09	23.42 ± 0.06	22.92 ± 0.07	22.61 ± 0.07	⋯	
5	GDN_BBG05	189.14454583	62.10413611	>26.47	24.06 ± 0.10	23.55 ± 0.07	23.11 ± 0.08	22.22 ± 0.06	36.1 ± 7.3	
6	GDN_BBG06	189.08689167	62.29081389	27.59 ± 0.18	24.98 ± 0.21	24.39 ± 0.13	24.08 ± 0.18	23.55 ± 0.12	⋯	
7	GDN_BBG07	189.10292917	62.31471944	26.11 ± 0.16	24.65 ± 0.16	24.34 ± 0.13	⋯	⋯	⋯	ME
8	GDN_BBG08	189.39490000	62.31689167	27.13 ± 0.16	23.95 ± 0.09	23.75 ± 0.07	23.33 ± 0.10	23.03 ± 0.09	⋯	W16
9	GDN_BBG09	189.40792083	62.21698889	>27.49	24.43 ± 0.13	24.19 ± 0.11	⋯	⋯	⋯	
10	GDN_BBG10	189.02397083	62.22303333	>27.31	24.01 ± 0.09	23.68 ± 0.07	23.39 ± 0.10	23.02 ± 0.09	⋯	W16
11	GDN_BBG11	189.10607083	62.15669722	>27.08	24.10 ± 0.10	23.91 ± 0.09	23.0 ± 0.08	22.44 ± 0.06	⋯	X-ray
12	GDN_BBG12	188.96070833	62.18147500	>27.22	24.46 ± 0.14	24.21 ± 0.12	⋯	⋯	⋯	
13	GDN_BBG13	189.25702917	62.25032778	>26.45	23.79 ± 0.08	23.82 ± 0.08	23.96 ± 0.17	23.14 ± 0.09	21.3 ± 5.2	W16
14	GDN_BBG14	189.14461667	62.11762500	26.43 ± 0.20	24.37 ± 0.12	24.39 ± 0.13	24.22 ± 0.21	23.51 ± 0.11	⋯	ME
15	GDN_BBG15	189.16360000	62.12178333	26.34 ± 0.10	24.25 ± 0.11	24.18 ± 0.11	⋯	⋯	⋯	B15, ME
16	GDN_BBG16	189.42832083	62.26589167	>27.00	23.52 ± 0.07	23.61 ± 0.07	23.27 ± 0.09	22.31 ± 0.06	301.0 ± 8.4	FIR, W16
17	GDN_BBG17	189.38280000	62.33746667	26.49 ± 0.11	24.78 ± 0.18	23.97 ± 0.09	⋯	⋯	⋯	B15, ME
18	GDS_BBG01	53.13474583	−27.90747222	26.40 ± 0.11	23.36 ± 0.08	23.69 ± 0.10	22.58 ± 0.07	22.45 ± 0.08	⋯	W16
19	GDS_BBG02	53.19989167	−27.90467500	26.78 ± 0.17	22.40 ± 0.05	22.23 ± 0.05	21.58 ± 0.04	21.13 ± 0.04	52.2 ± 4.9	FIR, SMG, X-ray, B15, W16
20	GDS_BBG03	53.04758333	−27.86863611	26.15 ± 0.11	23.39 ± 0.08	23.30 ± 0.07	22.48 ± 0.07	22.36 ± 0.07	⋯	W16
21	GDS_BBG04	53.21421250	−27.85935833	>27.18	24.42 ± 0.17	24.57 ± 0.19	24.68 ± 0.39	⋯	⋯	
22	GDS_BBG05	53.04210833	−27.84253333	>27.38	23.99 ± 0.13	24.12 ± 0.14	23.01 ± 0.10	22.65 ± 0.09	⋯	
23	GDS_BBG06	53.03234167	−27.83515833	26.73 ± 0.16	24.51 ± 0.17	24.40 ± 0.17	⋯	⋯	⋯	
24	GDS_BBG07	53.11909167	−27.81397778	26.71 ± 0.10	23.65 ± 0.10	23.48 ± 0.08	22.81 ± 0.09	22.61 ± 0.09	⋯	W16, ME
25	GDS_BBG08	53.16725833	−27.71545278	>27.68	24.23 ± 0.16	23.95 ± 0.12	22.99 ± 0.10	23.00 ± 0.11	⋯	H11
26	GDS_BBG09	53.06088750	−27.71842778	27.10 ± 0.12	23.83 ± 0.11	23.66 ± 0.09	22.93 ± 0.09	22.33 ± 0.07	24.4 ± 6.0	FIR, H11, W16
27	GDS_BBG10	53.13275833	−27.72019444	>27.02	23.79 ± 0.11	23.73 ± 0.10	22.65 ± 0.08	22.23 ± 0.07	⋯	H11, W16
28	GDS_BBG11	53.08477083	−27.70801944	>27.51	23.67 ± 0.10	23.21 ± 0.07	22.38 ± 0.06	22.01 ± 0.06	⋯	H11, W16
29	GDS_BBG12	53.19106667	−27.69395833	26.02 ± 0.09	24.01 ± 0.13	23.83 ± 0.11	23.08 ± 0.11	22.68 ± 0.09	⋯	
30	GDS_BBG13	53.02082083	−27.69909722	>27.09	24.02 ± 0.13	23.79 ± 0.10	22.92 ± 0.09	22.49 ± 0.08	⋯	W16
31	GDS_BBG14	53.12760833	−27.70668611	25.72 ± 0.07	22.79 ± 0.06	22.68 ± 0.05	22.28 ± 0.06	21.86 ± 0.06	⋯	W16, ME
32	GDS_BBG15	53.04935000	−27.75788611	26.61 ± 0.17	24.11 ± 0.14	23.73 ± 0.10	22.85 ± 0.09	22.69 ± 0.09	⋯	
33	GDS_BBG16	53.19653333	−27.75699444	>27.24	23.86 ± 0.12	23.81 ± 0.11	22.78 ± 0.08	22.71 ± 0.09	⋯	B15, W16

Note.
[a] H11: reported in Huang et al. (2011). B15: reported in Bouwens et al. (2015). W16: reported in Wang et al. (2016). SMG: source detected at wavelengths longer than 850 μm. FIR: source detected by PACS and/or SPIRE. X-ray: source detected in X-rays. ME: multiple counterparts or extended morphology in the *HST* individual bands or stacks.

magnitudes in the range $K \sim 24.6$–26.7 mag ($S/N > 5$). All 33 BBGs were required to be undetected in the CANDELS (and 3D-*HST*) F160W catalogs. However, in principle, there could still be a weak flux in the image that was missed by the CANDELS source extraction procedure. In order to constrain that potential signal, we first searched for weak detections of the BBG candidates in the deep (nine-band) *HST* and (25-band) SHARDS stacks described in Section 2.3 using a 1″ search radius around the IRAC positions. Interestingly, we found that 11 out of 16 (∼70%) and nine out of 17 (∼50%) BBG candidates are detected in the GOODS-S and GOODS-N *HST* stacks, respectively. In addition, six of the latter (∼35%) are also detected in the SHARDS stack (only available in GOODS-N). Out of our 33 sources, seven are exclusively detected in the IRAC data. These seven IRAC-only sources present magnitudes around 24.0–24.5 mag, ∼0.3 mag fainter than the rest.

For all of the BBG candidates, we measured accurate positions using the following method. First, we tweaked the World Coordinate System (WCS) solution for the IRAC images locally, taking as a reference the F160W CANDELS mosaic. For this purpose, we aligned the IRAC and *HST* images using detected galaxies in a 30″ circle around each BBG candidate. Typically, this implies corrections in the WCS of the IRAC images by offsets smaller than 0.″2. The rms of the comparison between source centroids calculated in *HST* and IRAC images is 0.″3. For the BBGs with counterparts in the *HST* stack, we adopted the more robust and reliable astrometric positions based on the *HST* imaging. Using these positions, we measured the fluxes in the F160W mosaic. For the galaxies with extended or multiple-knots morphology (as inferred from the stacks), we used elliptical apertures with semimajor axes ranging between 0.″6 and 0.″9. For the rest of the sample, we used apertures of 0.″4 radius, which maximized the S/N in the *HST* images for pointlike ultra-faint sources. We also considered larger apertures and checked the consistency of our results regarding the photometric procedure. Please refer to Appendix A for a detailed description of the method used in this work to derive consistent and reliable photometry in the optical and NIR *HST* bands.

We obtained weak but reliable fluxes ($H = 25.7$–27.6; $S/N > 5$) for 17 of the 33 galaxies. This clearly indicates that many BBG candidates are (weakly) detected in the F160W mosaics, but they were missed in the CANDELS (and 3D-*HST*) catalogs due to the lower completeness level of the selection procedure at faint magnitudes. Finally, in all cases where we could not recover a positive flux in our measurements, we measured upper limits. The upper flux limit was computed as 5σ of the sky noise measured in an empty region around the source with the same aperture size as that used for the photometric measurement and taking into account pixel-to-pixel noise correlations (Pérez-González et al. 2008).

In summary, most BBGs have fluxes in five to six optical-to-MIR bands. The complete SEDs were then used to estimate photometric redshifts and characterize the stellar emission (see next subsections). Around 20% of the sample also has FIR detections that were used to analyze the dust emission and (obscured) SFR properties (see Section 4.4).

4.2. Photometric Redshifts

We estimated photometric redshifts for the 33 BBGs using two different codes, namely, `pzeta` (Pérez-González et al. 2008) and `EAzY` (Brammer et al. 2008). Both codes estimate the redshift by fitting the SEDs in the spectral range where the emission is most probably dominated by stellar emission, i.e., at wavelengths bluer than rest-frame ∼3 μm, using a limited set of galaxy and active galactic nucleus (AGN) templates. Figure 3 illustrates the fitting procedure with the two codes and shows the resulting photometric redshift probability distribution functions (zPDFs).

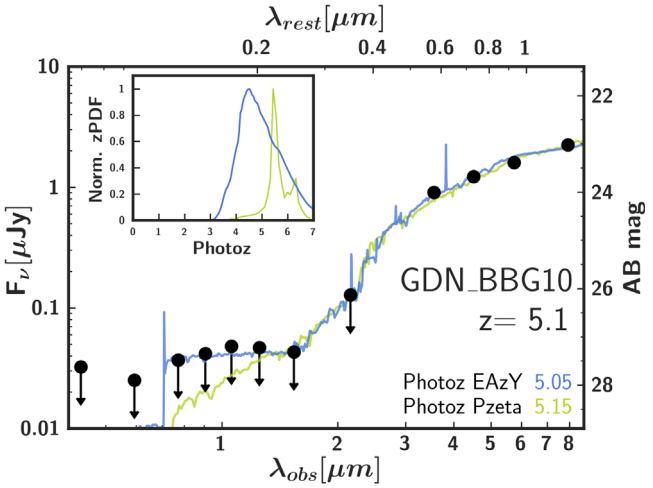

Figure 3. The SED of one of our sources in GOODS-N (black circles, including measured fluxes and 5σ upper limits), jointly with the best-fitting models used by our two photo-z codes (`EAzY` in blue and `pzeta` in green). The inset shows the zPDF, and the legend gives the two redshift estimates obtained with each code, as well as the final assumed value.

Although the majority of the sources are extremely faint at short wavelengths ($\lambda < 3.6\ \mu$m), we used the low-S/N data points (both from individual images and stacks) to impose upper limits and better constrain the photo-z. Moreover, we visually inspected the results for each galaxy to verify that the fit relies primarily on the high-S/N data and that the best-fit template was realistic. When necessary, we reran the code, excluding the lowest-S/N data ($S/N \lesssim 5$) or any template that was not consistent with being undetected at longer wavelengths or with all of the upper limits. In general, the best-fit photo-z from both codes and the zPDFs is consistent. Thus, we adopted the result with the most conservative uncertainty (i.e., the broader zPDF). In the cases where the results did not agree or none of the solutions properly fitted the SED, we analyzed the zPDF and adjusted the permitted redshift range interval to obtain the most reliable value.

Obtaining reliable photometric redshifts obviously benefits from having high-quality photometry in many bandpasses. Given the difficulties in deriving reliable fluxes for BBGs at short wavelengths, we repeated the whole photo-z estimation process for each galaxy with the three different sets of SEDs explained in Appendix A. Hence, we investigated the error in redshift estimate including the contribution of the photometric uncertainties. Relatively small (considering the type of very faint galaxies that we are dealing with) photometric redshift differences were found when using the different photometric approaches, typically below $\delta z = 0.2$. In any case, we checked the consistency of the results presented in the following sections using the zPDFs for the three SED types presented in Appendix A. These different zPDFs were also used to estimate uncertainties in those results, for example, in the photometric redshift distribution of our sample of BBGs.

Table 2
Parameter Space (Star Formation Timescale, Age, Dust Attenuation, and Metallicity) Allowed in the Stellar Population Synthesis Fitting Procedure

Parameter	Range	Units	Step
Timescale (τ)	8.0–10.0	log(yr)	0.1 dex
Age (t)	6.0–9.5	log(yr)	0.1 dex
Dust attenuation $A(V)$[a]	0.0–4.0	mag	0.1 mag
Metallicity (Z)	1.0	$Z_\odot$	Fixed

Note.
[a] For non-IR emitters, we only allowed dust attenuation to vary between 0 and 2 mag.

4.3. Stellar Population Properties

Based on the best-fit photometric redshift computed in the previous section, we again fitted the SEDs of the BBGs using stellar population synthesis models in order to characterize their stellar masses, dust attenuations, and mass-weighted ages (t_m). We performed this fit using two different codes, namely, `synthesizer` (Pérez-González et al. 2008) and `FAST` (Kriek et al. 2009). We used the same set of stellar population and dust modeling assumptions for both codes. We considered the stellar population models from Bruzual & Charlot (2003), assuming a delayed exponential star formation history (SFH; $\mathrm{SFR}(t) \propto te^{-t/\tau}$) and Chabrier (2003) IMF, and we adopted the Calzetti et al. (2000) dust attenuation law. Each stellar population model is characterized by four parameters: timescale τ, age t, metallicity Z, and dust attenuation $A(V)$. We assumed solar metallicity and allowed the other three free parameters to vary within the ranges presented in Table 2. Figure 4 shows an example of the fitting procedure and the resulting stellar population properties.

Overall, the results for stellar masses, dust attenuations, and mass-weighted ages using the two codes are roughly consistent within the typical uncertainties (e.g., ∼0.3 dex for the stellar mass). Nonetheless, we further verified the robustness of the estimated parameters by analyzing possible degeneracies (clusters of likely solutions) in the full parameter space using a Monte Carlo (MC) algorithm included in `synthesizer` (see, e.g., Domínguez Sánchez et al. 2016 for more details). Briefly, we generated 1000 random variations of the SED of each galaxy assuming Gaussian photometric errors, and we explored the resulting set of best-fit parameters in the age-versus-τ space. The 1000 MC particles typically form one to three clusters of solutions. In most cases, there is always one cluster with a much higher likelihood (defined from the number of MC particles belonging to that cluster). However, a small fraction of the galaxies (∼15%) exhibit two or more solutions with similar likelihoods. In those cases, we compared each set of results with those from as in FAST (Kriek et al. 2009) at the beginning of the section, and we visually inspected the best-fit result for each cluster to identify the most reliable solution. We also used the MC simulations to assign uncertainties to the best-fit parameters based on the 68% probability contours around the median result for each cluster. Given the faintness of the BBGs, the uncertainties in their derived physical parameters are relatively large. The statistical effect of the uncertainties in the photometric redshifts on the stellar masses was also considered. Stellar mass probability distribution functions (smPDF) for each galaxy were constructed from the zPDFs described in Section 4.2. These smPDFs have been used to estimate the uncertainties of the results presented in the following sections.

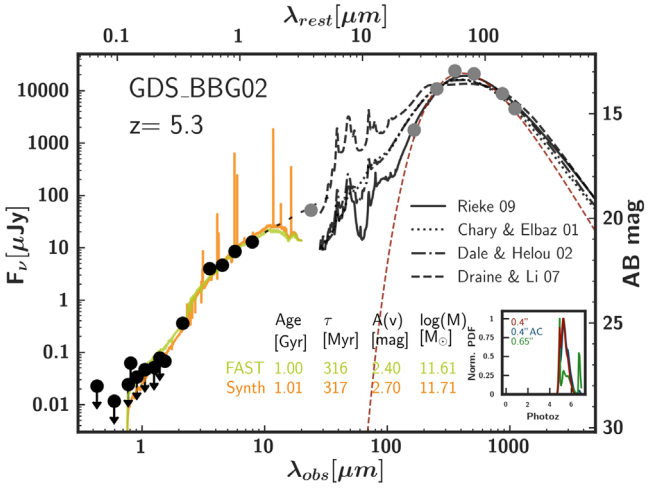

Figure 4. The SED fitting example for one of our BBG candidates. Bands (most likely) dominated by stellar emission are plotted with black circles, and bands probing the dust emission are plotted with gray circles. The inset shows the photo-z zPDFs from `EAzY` for the three different sets of SEDs obtained with distinct photometric approaches (red, blue, and green). The best fit and the inferred parameters from `synthesizer` and `FAST` are shown in orange and green, respectively. The best-fitting solutions for different dust emission libraries (Chary & Elbaz 2001; Dale & Helou 2002; Draine & Li 2007; Rieke et al. 2009) used to characterize MIR/FIR and submillimeter fluxes are shown with different black lines. The best dust emission model is the one shown with a solid black line. The red dashed line corresponds to the modified blackbody model fitting the data above 20 μm.

4.4. SFRs

We computed SFRs for the BBGs using the most reliable SFR tracer available for each galaxy. This approach is similar to the SFR "ladder" method described in Wuyts et al. (2011). In brief, we rely on the IR-based SFR estimates for galaxies detected at MIR-to-FIR wavelengths, and we used the best-fitting SPS model to estimate the SFR for the rest. As shown in Wuyts et al. (2011), the agreement between these estimates for galaxies with a moderate attenuation (faint-IR fluxes) ensures continuity between SFR indicators.

For IR-detected galaxies (∼25% of the sample), the total SFRs, $\mathrm{SFR_{UV+IR}}$, are computed from a combination of IR and rest-frame UV luminosities (uncorrected for attenuation) applying the Kennicutt (1998) equation normalized to a Chabrier (2003) IMF:

$$\mathrm{SFR_{UV+IR}} = 1.1 \times 10^{-10}(L_{\mathrm{IR}} + 3.3 \times L_{\mathrm{UV}})\ [M_\odot\ \mathrm{yr}^{-1}]. \quad (3)$$

The UV luminosity traces the (typically small) fraction of ionizing photons that are not absorbed by the dust. The IR luminosity is determined from the fitting of the available MIR/FIR and submillimeter data to four dust emission models (Chary & Elbaz 2001; Dale & Helou 2002; Draine & Li 2007; Rieke et al. 2009) following the methods described in Pérez-González et al. (2008, 2010) and Barro et al. (2011a). For each IR-detected BBG, we used the L_{IR} of the model that more accurately fit the data for the estimate of $\mathrm{SFR_{UV+IR}}$. In all cases, the typical scatter of L_{IR} estimations based on different template libraries (including the fit to a modified blackbody) is below 0.1 dex. See Figure 4 and Appendix C for some examples of the IR SED fitting.

For IR-undetected galaxies (∼75% of the sample), we used the SFR values inferred from the SED fit. This is also the case for MIPS emitters at $z \gtrsim 5$, for which MIPS 24 μm shifts out of

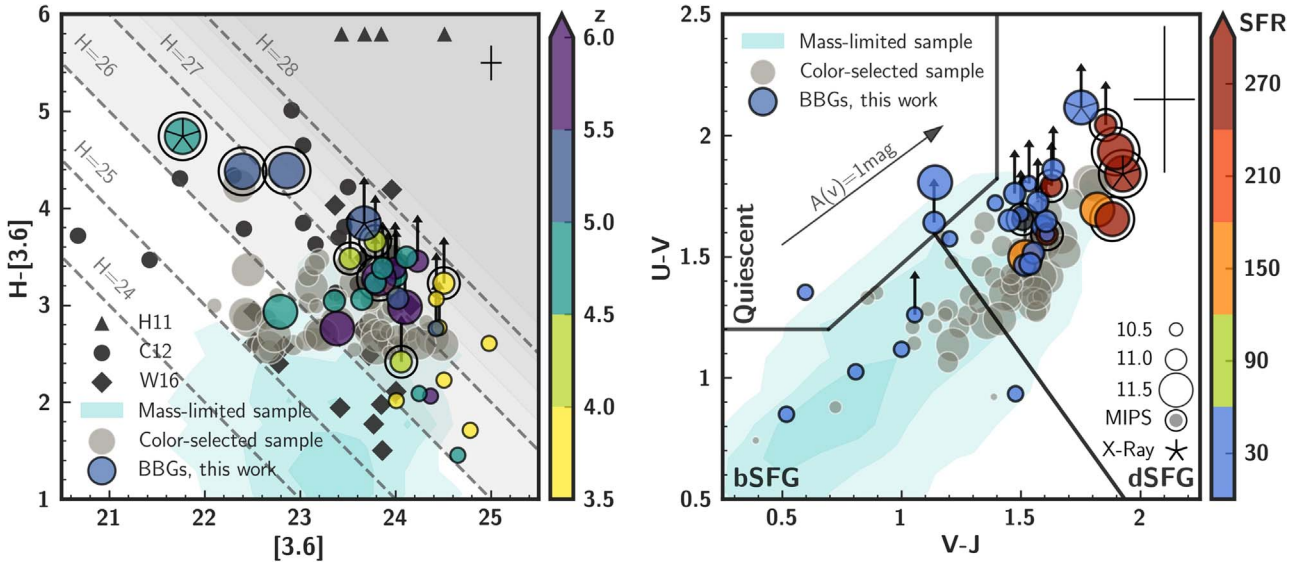

Figure 5. Left panel: observed-frame $H - [3.6]$ color plotted vs. the observed [3.6] magnitude for our sample of BBGs, color-coded by their photometric redshift and scaled in size as a function of their stellar mass (legend shown in right panel). The CANDELS color-selected sample is plotted in gray with the size also scaled according to the masses. The 0.3σ, 1σ, and 2σ distributions of the mass-limited sample are shown by the light blue density regions. The black triangles represent the four massive galaxies at $z > 4.5$ found by Huang et al. (2011), black filled circles show the extremely red sources from Caputi et al. (2012), and black diamonds correspond to the H-band dropouts reported by Wang et al. (2016). Lines (dashed) of constant H values are also shown. The darker background indicates the regions of lower completeness for the H band (Guo et al. 2013). Individual error bars are not plotted for clarity, but the average values for our sample of BBGs are shown in the top right corner of each panel. The H-band upper limits ($H - [3.6]$ lower limits) are also shown. Right panel: rest-frame $U - V$ vs. $V - J$ color–color plot, where BBGs are color-coded by SFR and scaled by stellar mass. The SFR lower limits are shown in dark gray. Sources from the CANDELS color-selected sample scaled by mass are also shown. The 0.3σ, 1σ, and 2σ distributions of the mass-limited sample are shown by the light blue areas. The MIPS-detected galaxies are surrounded by a circle, while the X-ray detected galaxies are highlighted with an asterisk inside the symbol. The Whitaker et al. (2011) upper boundary (black wedge) separates quiescent galaxies (top left) from SFGs (bottom). The black diagonal line denotes an additional criterion proposed in our work (perpendicular to the attenuation vector) to separate bSFGs from dSFGs. The plot also includes a 1 mag attenuation vector (which assumes a Calzetti et al. 2000 law). The H-band upper limits ($U - V$ lower limits) are also shown.

the rest-frame MIR region ($\lambda = 4$–$20\,\mu$m) where dust emission models are not defined and would produce highly uncertain $L_{\rm IR}$ values due to the large extrapolation involved. For these two types of sources (28 BBGs in total), we used the SFR averaged over the last 100 Myr of the SFH. We adopted this measurement over an $L_{\rm UV}$-based value corrected for extinction due to the faintness of the BBGs. Indeed, the rest-frame UV is redshifted into the ACS and WFC3 bands, where our galaxies are extremely faint or even undetected. Consequently, UV slope measurements cannot be performed properly or would be highly uncertain. While some BBGs exhibit marginal optical detections, the SED-based SFRs can be computed uniformly for all galaxies, and they include the F160W fluxes when available.

In summary, we estimate SFRs using a variety of methods for all BBGs. The reliability of these SFR estimations is, in general, not high, mostly because of the faintness of our galaxies, which is especially extreme in the rest-frame UV (given that BBGs were selected as red objects missed by the deepest optical and NIR surveys). Nevertheless, the SFRs for the several FIR emitters in our sample (five out of 33) are more robust, given that the dust emission is very well constrained. For the sources with only an MIPS MIR detection or undetected, the SFRs, based on SED fitting, should be considered as a lower limit, since they are typically significantly smaller than the SFRs calculated from dust emission probed by the FIR data. We also remark here that the SFRs for the galaxies in the comparison samples have been estimated in a similar way, so our results for the BBGs relative to the known (cataloged) populations of massive galaxies at high-z are robust. Further observations capable of measuring emission lines or fainter MIR/FIR fluxes are needed for better accuracy in the SFR analysis.

5. Physical Properties of BBGs

Here we analyze the distribution of the observed colors, photometric redshifts, and stellar population properties of the 33 BBGs using the results from the UV-to-FIR SED fitting techniques described in Section 4. We will also compare BBGs with two samples constructed with the CANDELS GOODS-S and GOODS-N H-band selected catalogs presented in Guo et al. (2013) and G. Barro et al. (2019): a mass-limited and a color-selected sample. The mass-limited sample is composed of massive ($M > 10^{10}\,M_{\odot}$) galaxies at $z > 3$. We remark that this is a sample constructed with a simple cut in redshift and mass rather than a mass-complete sample. The color-selected sample, aimed at reproducing our BBG selection, is composed of red ($H - [3.6] > 2.5$ mag), faint ($H > 25$ mag) galaxies. These two samples are characterized in Appendix B. In addition, we compare our sample of BBGs with the samples of red galaxies of nature presented in Huang et al. (2011, hereafter H11), Caputi et al. (2012, hereafter C12), and Wang et al. (2016, hereafter W16).

5.1. Observed IR Colors and Photometric Redshifts

The left panel of Figure 5 shows the $H - [3.6]$ versus [3.6] color–magnitude diagram for our 33 BBGs color-coded by redshift. For this and the rest of the figures in the following sections, we use our fiducial photometry. For reference, we

Table 3
Statistical Properties of the Different Samples

Sample	z	H (mag)	[3.6] (mag)	[4.5] (mag)	$H-[3.6]$ (mag)	M ($M_\odot$)
BBGs	$4.8^{5.1}_{4.4}$	$27.1^{27.3}_{26.5}$	$24.0^{24.4}_{23.7}$	$23.8^{24.1}_{23.6}$	$3.1^{3.4}_{2.8}$	$10.8^{11.1}_{10.4}$
Mass-lim.[a]	$3.8^{4.7}_{3.3}$	$24.9^{25.7}_{24.2}$	$23.3^{23.8}_{22.8}$	$23.2^{23.8}_{22.6}$	$1.6^{2.1}_{1.1}$	$10.4^{10.6}_{10.1}$
Color-sel.[b]	$4.7^{5.3}_{4.1}$	$26.5^{26.8}_{25.9}$	$23.6^{23.9}_{23.1}$	$23.4^{23.9}_{22.7}$	$2.8^{3.1}_{2.6}$	$10.8^{11.1}_{10.4}$
H11[c]	5.7	$29.8^{30.2}_{29.5}$	$23.8^{24.0}_{23.6}$	$23.5^{23.7}_{23.4}$	$6.1^{6.6}_{5.5}$	⋯
C12[d]	$4.1^{4.9}_{3.4}$	$26.9^{27.5}_{26.5}$	$23.2^{23.5}_{22.4}$	$22.5^{23.13}_{22.1}$	$3.7^{4.2}_{3.5}$	⋯
W16[e]	$4.7^{5.7}_{4.2}$	$26.0^{26.7}_{25.4}$	$23.7^{23.9}_{23.2}$	$23.4^{23.6}_{22.9}$	$2.6^{3.0}_{2.1}$	⋯

Notes. Median values, first and third quartiles of their redshift, magnitudes, colors, and masses are shown. The statistics have been calculated by assuming 5σ values for the H-band undetected galaxies.
[a] CANDELS mass-limited sample.
[b] CANDELS color-selected sample.
[c] H11: Huang et al. (2011).
[d] C12: Caputi et al. (2012).
[e] W16: Wang et al. (2016).

describe how the different photometric methods described in Appendix A affect the results.

As demonstrated in Appendix B, a red $H-[3.6]\gtrsim 2$ mag color is a good proxy to identify massive red galaxies (dusty or evolved) at $z>3$ (BBGs). These galaxies present strong Balmer (or D4000) breaks, which produce the red colors (sometimes in combination with dust attenuation). Thus, H-band dropouts in the CANDELS catalogs ($H(5\sigma)\sim 27$ mag) with bright [3.6] magnitudes are excellent high-redshift BBG candidates. However, as discussed in Section 4, some of our BBGs have weak H-band detections. They are undetected in the CANDELS catalogs mostly due to the increasing catalog incompleteness at $H>25$ mag (see Figure 4 in Guo et al. 2013). As a result, a small fraction (∼20%) of our BBGs have faint H-band magnitudes but colors slightly bluer than $H-[3.6]\sim 2.5$ mag (see the bottom right corner of the color–magnitude diagram in Figure 5). Note that the bluest BBGs have faint IRAC magnitudes, $[3.6]=24$–25. Likewise, one can also identify the opposite situation, i.e., galaxies with relatively bright H and IRAC magnitudes but red $H-[3.6]\gtrsim 2.5$ colors, that are also good BBG candidates. This is the case for the gray circles in Figure 5, which depict the color-selected sample (see Appendix B for more details). The distribution of the color-selected sample in the diagram shows that our BBG dropout criterion is essentially a faint-end extension of a color- and H-band-limited sample. Table 3 summarizes the average properties of our sample of BBGs and the galaxies selected by color from the CANDELS catalog. Our BBGs are typically galaxies with a very faint or nonexistent detection in the H band, $\langle H\rangle\sim 26.5$ mag (for the 60% of sources with detections in the CANDELS data). These galaxies are remarkably bright in IRAC, $\langle[3.6]\rangle\sim 24$ mag, which converts them in very red sources, $\langle H-[3.6]\rangle\sim 3.1$ mag. For reference, these statistical properties change by 0.3–0.5 mag when considering the other two photometric methods described in Appendix A; for example, for the photometric apertures of size $r=0.''65$, our sample of BBGs presents $\langle H\rangle\sim 26.7$ and $\langle H-[3.6]\rangle\sim 2.8$ mag. Comparing our BBGs with other samples, we find that the CANDELS color-selected sample has ∼0.4 mag brighter [3.6] and H-band magnitudes and a slightly bluer $H-[3.6]$ color. We note that both samples have very similar photometric redshift distributions peaking at $z\sim 5$ (see discussion below).

Figure 5 also compares the 33 BBGs to other samples of red, massive, high-z candidates (black symbols) from the works of H11, C12, and W16. In particular, H11 and W16 identified four and 16 BBG candidates, respectively, in the GOODS fields. Our selection method recovers all of these galaxies (as indicated in Table 3). Our photometry is consistent with these works in the IRAC bands ($\langle\Delta[3.6]\rangle\sim 0.06$ mag) and slightly brighter in the H band, particularly relative to H11 ($\langle\Delta H_{\rm H11}\rangle\sim 2.67$ mag). This difference suggests that our forced photometric measurement is very effective at recovering flux for faint H-band sources. This difference translates to H11 reporting much redder colors for the galaxies in common with our sample. We also find significant differences in some cases ($\langle\Delta H_{\rm W16}\rangle\sim 0.69$ mag) between our H-band magnitudes and those reported by W16. Overall, the mean color of our BBGs ($H-[3.6]\sim 3$ mag) is similar to the values reported in W16 and C12. However, the C12 sample is slightly redder (∼0.4 mag), but we must also take into consideration that their sample was built in a field with shallower *HST* and IRAC data (the UDS; see Galametz et al. 2013). This translates to the C12 sample being around 1 mag brighter at 3.6 and 4.5 μm than our BBGs.

Table 3 also indicates a remarkably good agreement in the average photometric redshift values for all of the color-selected BBG samples. Typically, these MIR-bright red galaxies lie at $\langle z\rangle\sim 4$–6. The redshift distributions of our galaxies and those in H11, C12, and W16 are shown in Figure 6. The uncertainties in the photo-z histogram for our BBGs were derived as the $\pm 1\sigma$ deviation in each bin arising from building 1000 photo-z histograms based on our BBGs' zPDF for the three different sets of SEDs (refer to Appendix A for more details). While a few BBGs are found at $z<4$, the majority of the sample is skewed to higher redshifts, $z\gtrsim 4$ ($80\%\pm 1\%$ of the sample lies at $4<z<6$). We remark that, as we anticipated at the beginning of this section, the red colors of our BBGs strongly point to a $z>3$ redshift. As shown in this histogram, and considering the uncertainties in the optical/NIR photometry, a small fraction (<5%) of the total sample would present photometric redshifts below $z=3$.

5.2. *Rest-frame* UVJ *Colors, Stellar Masses, and Dust Attenuations*

The right panel of Figure 5 shows a *UVJ* diagram for our sample of BBGs, color-coded by SFR and sized by mass. We compare our BBGs with the CANDELS color-selected and

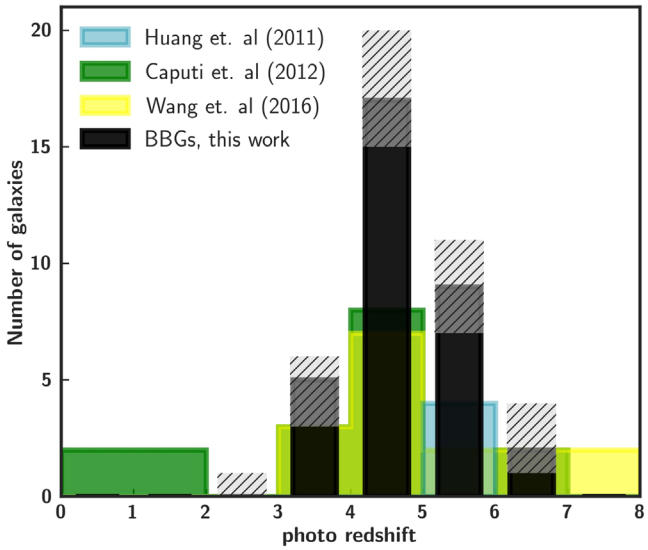

Figure 6. Photometric redshift distribution of our sample of BBGs, together with previous samples of similar galaxies reported in the literature and explained in the text. The BBGs are shown with black bars, and the uncertainty (see text for details) is highlighted with black hatched bars. The photo-z distributions of galaxies reported in Caputi et al. (2012) and Wang et al. (2016) are shown by green and yellow histograms, respectively. The four sources presented in Huang et al. (2011), all assumed to be at $z \sim 5.7$, are shown in blue. Photo-z statistics for each sample are presented in Table 3.

mass-limited samples. The black lines indicate the quiescent region (upper left), and the low- and high-extinction star-forming regions, bSFG (lower left) for blue SFGs and dSFG (upper right) for dusty systems, are also shown (see, e.g., Brammer et al. 2011 or Whitaker et al. 2012 for a detailed discussion of the age and extinction patterns in the *UVJ* diagram).

Figure 5 shows that the BBGs overlap with the most extinguished, dustier galaxies in the CANDELS color-selected sample. Note also that the color-selected sample identifies redder, more massive dusty or evolved galaxies than a pure mass-limited sample with no color constraints (located in the shaded region). This is highlighted by the blue regions, which show the 0.3σ, 1σ, and 2σ color distributions of massive ($M > 10^{10}\ M_{\odot}$) galaxies at $z > 3$ drawn from the CANDELS catalogs (see Appendix B for more details). The predominantly dust-obscured nature of the BBGs is further confirmed by the strong MIR-to-FIR detections of the galaxies in the upper right corner of the *UVJ* diagram ($\sim$30% of the BBGs in the dSFG region), which also imply large SFRs (see next section).

Nonetheless, about $\sim$20% of the BBG sample exhibits blue *UVJ* colors at the opposite extreme of the dSFGs. These galaxies are also among the bluest ($H - [3.6] \lesssim 2.5$ mag) and brightest ($H \sim 26$ mag) of the sample, located at the bottom right corner of the color–magnitude diagram in the left panel of Figure 5 (see discussion in Section 5.4). This suggests that they are included in the otherwise red sample of H-band dropouts due to incompleteness in the CANDELS catalog.

Interestingly, Figure 5 also indicates that there are few quiescent galaxies in both the mass-limited and BBG samples and none in the color-selected and BBG samples. We remark, however, that several galaxies count with lower limits in $U - V$, so their colors are consistent with the quiescence locus. The lack of quiescent galaxies at $z > 3$ (and we remind the reader that we are more efficient in detecting galaxies at $4 < z < 6$) is consistent with recent works that have identified several massive quiescent galaxies at $z = 3$–4 (e.g., Straatman et al. 2014, 2015). These quiescent galaxies have lower redshifts and therefore brighter H-band magnitudes and bluer $H - [3.6]$ colors than our BBGs. For example, the six quiescent galaxies identified by Straatman et al. (2014) in GOODS-S have a median color of $H - [3.6] = 2.2$ mag and a redshift of $z = 3.7$, whereas 80% of our BBGs are at $z > 4$. All together, the color distribution of the BBGs suggests that it is increasingly more difficult to find bona fide fully quenched galaxies beyond $z \sim 4$. Possibly this is because, at such high redshifts, even the most evolved galaxies did not have time to reach a mass-weighted age older than $t \sim 1$ Gyr, which is the approximate threshold for a single stellar population (SSP) to make it into the quiescent region of the *UVJ* diagram (see Figure 17 in Appendix B).

Figure 7 shows the photometric redshift and stellar mass distributions of the BBGs compared to the CANDELS color-selected and mass-limited samples. As shown in the previous section, an H-band dropout selection in the GOODS/CANDELS field (implying an "extremely" red $H - [3.6]$ color) identifies galaxies at $z \gtrsim 4$ (Figure 6), whereas the CANDELS color-selected and, especially, mass-limited samples have a more pronounced tail at $z \sim 3$. The stellar mass histograms confirm the intuition we had when examining the *UVJ* diagram in Figure 5: both the color-selected and BBG samples are biased toward the identification of more massive galaxies than those from the bluer, mass-limited sample. This correlation between color, mass, and dust attenuation is fully consistent with previous results at lower redshifts (e.g., Brammer et al. 2011; Straatman et al. 2016; Wang et al. 2017; Fang et al. 2018). Typically, our BBGs have stellar masses around or above $\log(M/M_{\odot}) = 11$. In absolute terms, the BBG sample recovers a number of massive obscured galaxies similar to that of the CANDELS color-selected sample and one-tenth of the fraction of mass-limited galaxies (but the latter is more biased toward lower redshifts). This implies that the completeness level of the deepest H-band selected catalogs (e.g., CANDELS, 3D-*HST*) is not very high, and they are missing a significant fraction (around 40%) of massive red galaxies at $z > 4$ (see discussion in Section 5.6).

5.3. BBGs and the Star-forming Main Sequence

Figure 8 shows the SFR–versus–stellar mass diagram for the BBGs and the CANDELS color-selected and mass-limited samples (main statistical properties given in Table 4). The gray lines show the star-forming "main sequence" (MS; see, e.g., Noeske et al. 2007; Elbaz et al. 2011) at $z = 4$ (from Speagle et al. 2014; Salmon et al. 2015; Schreiber et al. 2015; and Davidzon et al. 2017). The MS illustrates the known correlation between SFR and mass for typical SFGs that has been shown to exist even up to $z = 7$–8 (Stark et al. 2009, 2013; Salmon et al. 2015). The MS inferred for the CANDELS mass-limited sample ($\langle M \rangle \sim 10^{10.3}\ M_{\odot}$ and $\langle z \rangle \sim 3.8$; black line) is consistent with the MSs from the literature, in spite of the typically high uncertainties in the SFR and stellar mass estimations for these redshifts. The gray shaded region shows 2σ around the MS ($\sigma \sim 0.3$ dex). The color code indicates galaxies in the MS (green) and above and below the MS 2σ band, namely, starburst (red) and sub-MS (orange) galaxies.

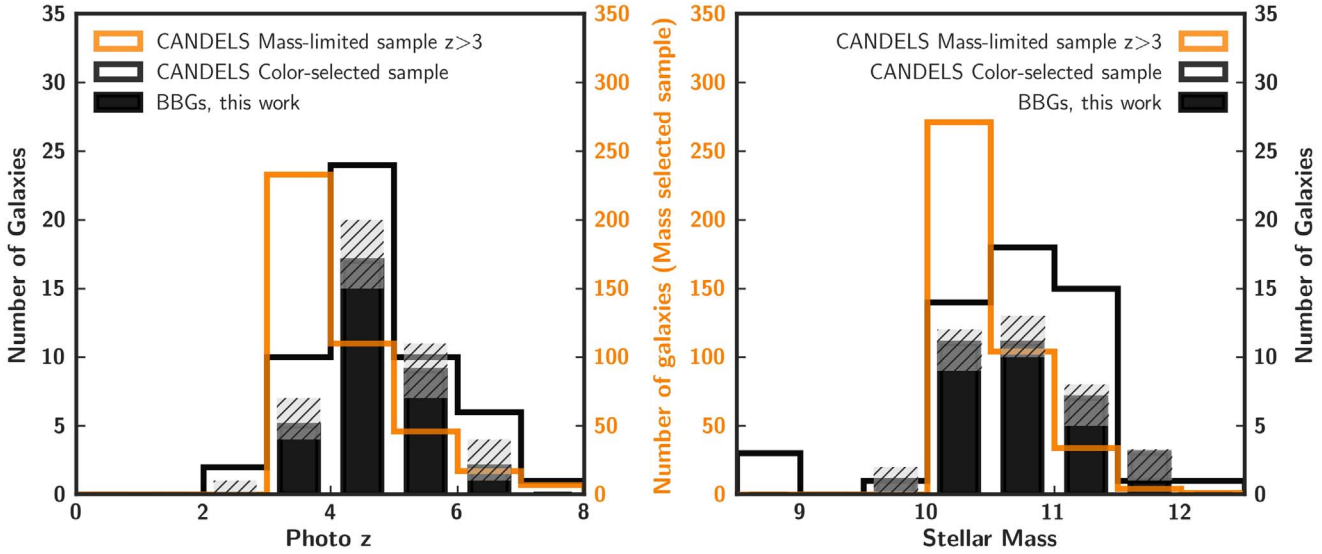

Figure 7. Photometric redshift (left panel) and stellar mass (right panel) distributions of our sample of BBGs compared to those of the color-selected and mass-limited CANDELS samples explained in the text (see also Appendix B). The photo-z distribution of the mass-limited and color-selected samples are shown by orange and black histograms, respectively. The BBGs are shown with black bars, and their uncertainties are highlighted with black hatched bars. Given the substantially higher number of galaxies in the mass-limited sample, a different axis has been used for it, as indicated by the orange labels.

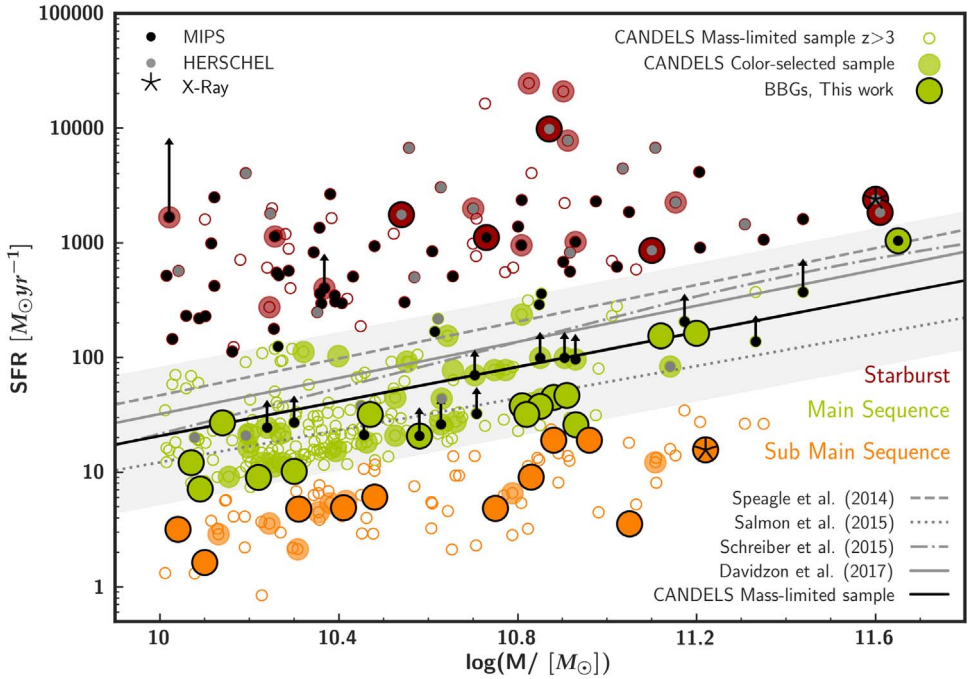

Figure 8. The SFR–vs.–stellar mass plane for the CANDELS comparison samples (color-selected: filled symbols; mass-limited: open symbols) and the BBGs (filled symbols enclosed by a black circle) reported in this work, color-coded according to their position with respect to the MS: starburst galaxies are shown in red, MS galaxies in green, and sub-MS galaxies in orange. The MSs at $z = 4$ according to Speagle et al. (2014), Salmon et al. (2015), Schreiber et al. (2015), and Davidzon et al. (2017) are shown with different gray lines (see the legend in the panel). The MS inferred for the CANDELS mass-limited sample is shown with a black solid line. The gray shaded region delimits the 2σ area around the MS. The MIR/FIR emitters are marked with an enclosed black/gray circle. The galaxies with MIPS detections but no IR-derived SFRs (because they lie at $z > 5$; see text for details) are shown as lower limits. The X-ray emitters are highlighted with an asterisk.

Despite the overall high scatter around the MS at $M \lesssim 10^{11.2}\,M_\odot$, the few galaxies at the high-mass end, $M \gtrsim 10^{11.4}\,M_\odot$, are located above the MS. At lower masses, $M \lesssim 10^{10.5}\,M_\odot$, however, all of the BBGs are either in the MS or below. This result is fully consistent with the inferences from the *UVJ* diagram in Figures 5 and 10; i.e., their red intrinsic colors are mainly due to the presence of strong bursts of obscured star formation, with a few galaxies being consistent with harboring more evolved stellar populations.

This result is further confirmed by Figure 9, which shows the mass-weighted age and attenuation histograms for the three samples. The distributions for the BBGs are clearly skewed toward higher attenuations, up to $A(V) \sim 3$ mag for some starbursts, in contrast with the mass-limited sample, typically characterized by attenuations around $A(V) \sim 1$ mag or below. The BBGs show older mass-weighted ages (typically older than 300 Myr) than the bluer, mass-limited sample (which peaks below 300 Myr). In line with the general location of BBGs in

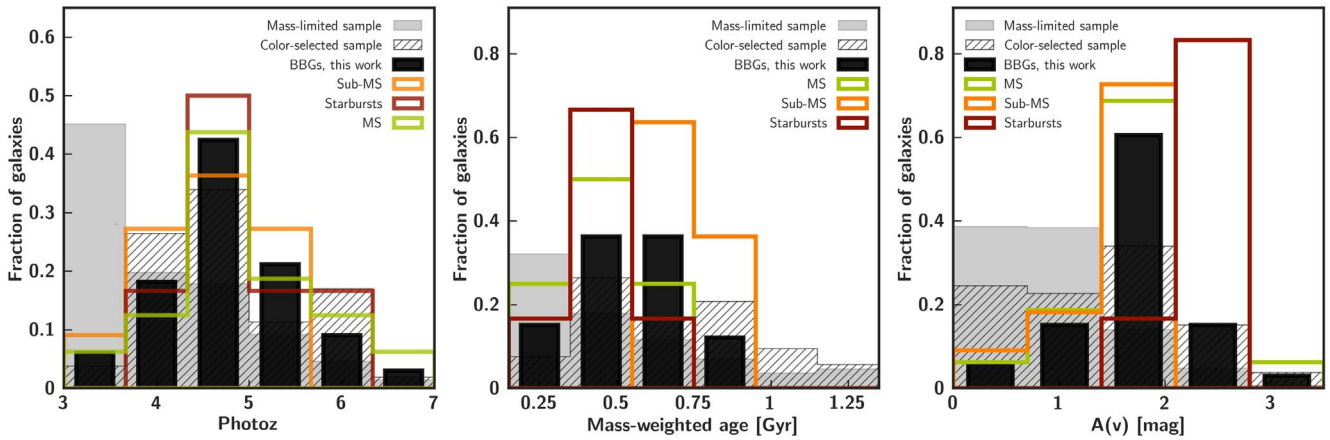

Figure 9. Distributions of the redshift (left panel), mass-weighted age (middle panel), and attenuation (right panel) for the different star formation subsamples. The black filled histograms represent the fraction of BBGs, while the distribution of the mass-limited and color-selected samples are shown with filled and hatched gray histograms, respectively. Starbursts, MS, and sub-MS BBGs are shown in red, green, and orange, respectively. Each histogram is normalized to its sum.

Table 4
Statistical Properties of Starburst, MS and Sub-MS BBGs, and Color-selected and Mass-limited Samples

Sample			z	t_m (Gyr)	$A(V)$ (mag)	SFR ($M_\odot$ yr^{-1})	sSFR (Gyr^{-1})	Mass ($M_\odot$)
Mass-limited	All		$3.8^{4.8}_{3.3}$	$0.3^{0.6}_{0.2}$	$0.7^{1.3}_{0.4}$	26^{124}_{13}	$1.2^{6}_{0.6}$	$10.4^{10.6}_{10.2}$
	Starburst	25%	$3.8^{5.2}_{3.2}$	$0.3^{0.5}_{0.1}$	$1.0^{1.6}_{0.2}$	737^{1622}_{356}	27^{63}_{12}	$10.5^{11.0}_{10.3}$
	MS	56%	$4.1^{4.8}_{3.3}$	$0.2^{0.5}_{0.2}$	$0.7^{1.1}_{0.4}$	23^{41}_{15}	$1.0^{1.7}_{0.7}$	$10.3^{10.5}_{10.1}$
	Sub-MS	19%	$3.5^{4.0}_{3.3}$	$0.7^{1.0}_{0.3}$	$0.5^{1.0}_{0.2}$	5^{11}_{4}	$0.2^{0.2}_{0.1}$	$10.3^{10.5}_{10.1}$
Color-selected	All		$4.7^{5.3}_{4.1}$	$0.7^{0.8}_{0.4}$	$1.5^{2.0}_{0.7}$	67^{236}_{12}	$1.3^{10}_{0.6}$	$10.8^{11.1}_{10.5}$
	Starburst	22%	$5.0^{5.6}_{4.6}$	$0.7^{0.8}_{0.5}$	$1.7^{1.9}_{1.6}$	1077^{2185}_{478}	46^{143}_{16}	$11.0^{11.1}_{10.8}$
	MS	58%	$4.8^{5.3}_{4.4}$	$0.5^{0.8}_{0.3}$	$1.3^{2.0}_{0.7}$	44^{90}_{19}	$1.2^{1.4}_{0.6}$	$10.8^{11.1}_{10.5}$
	Sub-MS	20%	$3.9^{4.0}_{3.7}$	$0.8^{1.0}_{0.4}$	$0.8^{2.0}_{0.3}$	5^{6}_{3}	$0.2^{0.2}_{0.1}$	$10.3^{10.5}_{10.2}$
BBGs	All		$4.8^{5.1}_{4.3}$	$0.5^{0.6}_{0.5}$	$2.0^{2.0}_{1.5}$	26^{154}_{9}	$0.5^{1.1}_{0.2}$	$10.8^{11.0}_{10.4}$
	Starburst	19%	$4.7^{5.2}_{4.4}$	$0.5^{0.5}_{0.5}$	$2.5^{2.6}_{2.4}$	1803^{2258}_{1275}	14^{43}_{6}	$11.0^{11.6}_{10.9}$
	MS	48%	$4.8^{5.3}_{4.5}$	$0.5^{0.6}_{0.4}$	$1.9^{2.0}_{1.6}$	32^{45}_{19}	$0.6^{1.0}_{0.5}$	$10.8^{11.0}_{10.3}$
	Sub-MS	33%	$4.7^{5.0}_{4.0}$	$0.7^{0.8}_{0.7}$	$1.7^{2.0}_{1.4}$	5^{12}_{4}	$0.2^{0.2}_{0.1}$	$10.8^{11.0}_{10.4}$

Note. Median values, first and third quartiles of their redshift, mass-weighted age, extinction, SFR, sSFR, and mass are shown. The percentage of sources belonging to each star formation mode is also shown. Sources with SFR lower limits have not been taken into account.

these histograms, the color-selected sample is built up by marginally older and less extincted galaxies. The average mass-weighted age of our BBGs is 0.5 Gyr, slightly older than the typical age for the mass-limited sample (0.3 Gyr) and similar to or younger than the average for the color-selected sample (0.7 Gyr). These young ages point out that our color and magnitude selection was preferentially identifying the Balmer and not the 4000 Å break (typical of more evolved populations, >1 Gyr). This is the justification for the choice of name (BBGs) for the new population of high-z massive galaxies presented in this paper. The oldest ages can be found among the sub-MS BBGs, which can be as old as 0.9 Gyr. The redshift distribution is roughly homogeneous for the three subsamples of BBGs divided by star formation activity, peaking at $z \sim 4.5$, although starbursts dominate at higher redshifts and sub-MS galaxies at lower redshifts with a flatter distribution between $z \sim 3.5$ and 4.5. This could be identified as the assembly of the first quiescent massive galaxies, which would be actively forming stars in the MS or even very actively forming stars in the starburst region at $z \gtrsim 4.5$ and may start to evolve more passively or even quench by $z \sim 4$ (after more than ∼1 Gyr).

Concerning attenuations, typical attenuations are around 2 mag and as large as 3.4 mag for starbursting BBGs. Both the mass-limited and color-selected galaxies present significantly smaller attenuations, around 1 mag.

The most relevant conclusion from this comparison is that the magnitude-limited F160W samples based on CANDELS data miss ∼15% of the massive ($M > 10^{10.5}\,M_\odot$) galaxies at $z \gtrsim 4$. The fraction increases as we move to higher masses (see Section 5.6). In relative terms, the impact of adding missed BBGs to a mass-limited sample is larger on the starburst region. We also remark that ∼20% of the total number of BBGs, ∼30% if we only consider $M \geqslant 10^{10.5}\,M_\odot$, are starbursts. If we combine the mass-limited sample and our BBGs, these fractions are 25% and 35%, respectively. These figures are much larger than those found at $z \sim 2$–3, 2%–4% for $M \gtrsim 10^{10}\,M_\odot$ (Rodighiero et al. 2011; Schreiber et al. 2015). Our fraction is slightly larger than the 15% reported by Caputi et al. (2017) for a sample of galaxies at $z \sim 4$–5 and $\langle M \rangle \sim 10^{9.5}\,M_\odot$.

Interestingly, very few of our BBGs lie far enough below the MS to be considered as completely quiescent. If we consider

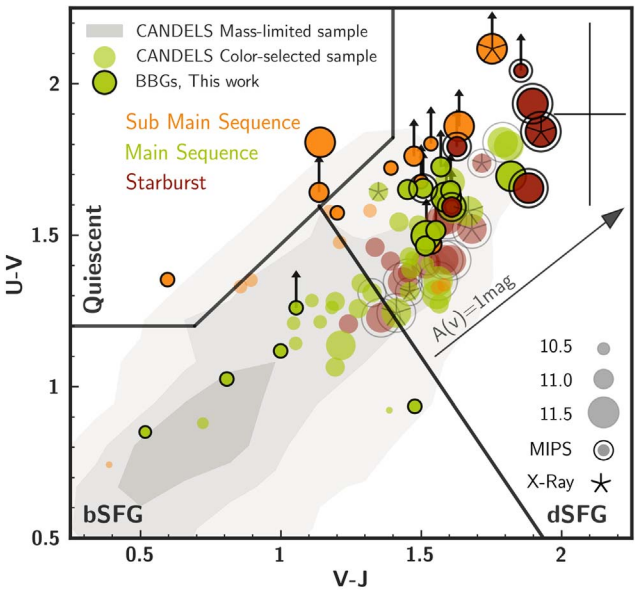

Figure 10. Rest-frame $U - V$ vs. $V - J$ colors for BBGs and the CANDELS color-selected comparison sample color-coded by star formation subsamples (with respect to the MS; see the color code in Figure 8) and sized by stellar mass. The H-band nondetections are shown as $U - V$ lower limits. The MIPS detections are enclosed by a circle, and X-ray-detected galaxies are highlighted with an asterisk inside the symbol. The 0.3σ, 1σ, and 2σ distributions of the mass-limited sample are shown by the light gray areas. The Whitaker et al. (2011) upper boundary (black wedge) separates quiescent galaxies (top left) from SFGs (bottom). The black diagonal line denotes an additional criterion to separate bSFGs from dSFGs. The 1 mag attenuation vector computed assuming a Calzetti et al. (2000) reddening law is also shown. Error bars are not plotted for clarity, but the average values are shown in the top right corner.

the average redshift of our sample, $\langle z \rangle = 4.8$, corresponding to a universe age of 1.2 Gyr, there is not much time for a massive galaxy to completely quench unless its star formation is very short and starts very early. This result is consistent with what we found with the *UVJ* diagram. For convenience, we show another version of it in Figure 10, but this time the BBGs and color-selected galaxies are color-coded by star formation subsamples. We remind the reader again that the $U - V$ color for our sample has a relatively high uncertainty (0.2 mag), given that our sources are extremely faint in the bands probing the U band. Therefore, some of our BBGs are consistent with being quiescent (see next subsection), or at least have moved to a post-starburst phase.

In any case, four sub-MS BBGs (GDN_BBG08, GDN_BBG09, GDN_BBG12, and GDS_BBG14) present mass-weighted ages around ~0.9 Gyr (with times from the start of their SFH around $t \sim 1.5$ Gyr; see Table 5), which translates to $z \gtrsim 18$ for the onset of their star formation.

5.4. Stacked SEDs of the BBGs

In this section, we further compare the colors of BBGs and galaxies in the color-selected and mass-limited samples grouped by their location relative to the MS. Given the intrinsically faint fluxes of BBGs at nearly all wavelengths, here we analyze average rest-frame SEDs of multiple BBGs to obtain a better and finer sampling of their average SEDs, and we compare them to the other two samples of (brighter) galaxies. The SEDs are normalized at 0.7 μm rest frame, which roughly corresponds to the IRAC 3.6 and 4.5 μm bands, where BBGs are brightest. We only include photometric points with reliable detections ($S/N > 3$) at wavelengths shorter than IRAC. We also removed from the stacks a small number of galaxies with unconstrained SFRs or redshifts $z \gtrsim 6$, which cannot be properly normalized at 0.7 μm rest frame if they are not detected beyond 4.5 μm. Sources from the mass and color-selected samples with uncertain IRAC photometry ($<5\sigma$ in [3.6], [4.5]) were also removed. Furthermore, we visually inspected their *HST* images to remove potentially blended objects or sources dominated by a central pointlike emission (AGN candidates).

Figure 11 shows the individual and average SEDs (constructed by fitting stellar population models to the stacked photometry) in the sub-MS, MS, and starburst regions. Quite reassuringly, we find that BBGs exhibit very similar average SEDs to the color-selected sample in all three regions, confirming again that the former are the faint-end extension of the latter. The comparison in the starburst region is particularly revealing, as both samples show clear detections at wavelengths longer than $\lambda_{r-f} > 2$ μm. This is what would be expected for dust emission in heavily enshrouded galaxies. Indeed, dust emission is starting to dominate the integrated SED at $\lambda_{r-f} \sim 2$ μm, and by $\lambda_{r-f} > 3$ μm, dust emits more than 50% of the integrated light for ~35% of the sources. In addition, the best-fit stellar population model to the stack for these starbursts indicates a large attenuation, $A(V) \sim 2$ mag. The comparison in the MS region shows that some BBGs have slightly bluer colors than the color-selected sample at short wavelengths, $\lambda_{r-f} < 400$ nm. These galaxies are the smaller-mass, bluer ($H - [3.6] \lesssim 2$) BBGs discussed in the previous sections, which are indeed more similar to the overall bluer stack of the mass-limited sample. This is confirmed by the *UVJ* diagram (Figure 10), where the few green circles characterized by bluer *UVJ* colors lie in the region with the highest density of galaxies from the mass-limited sample. Most of the BBG sample in this MS region, however, is characterized by a red SED reproduced by a stellar population with a very similar $A(V)$ compared to the starburst sample but considerably older age. We note that these two quantities are highly degenerated in the models fitting the stacked photometry, but, should the attenuation be smaller, the age should be older, so there must be a real difference between the starburst and MS subsamples. This is also confirmed by the lower fraction of MIPS emitters in the MS (100% of the starburst BBGs are detected in the MIR/FIR and 9% for the MS BBGs). Finally, there are very few BBGs or color-selected galaxies in the sub-MS region to infer any statistically significant result. However, as expected, both populations appear to have redder SEDs than those in the mass-limited sample, which favors older (or more extincted) galaxies.

In addition, the bluest BBGs could have been selected due to the presence of a strong (i.e., high equivalent width (EW)) emission line (such as Hα at $z \sim 4.5$) in the [3.6] band (see, e.g., Kashikawa et al. 2012; Smit et al. 2014; Mármol-Queraltó et al. 2016). Flux contamination by a strong emission line in the IRAC bands would lead to a brighter [3.6] magnitude and a redder $H - [3.6]$ color for a galaxy with relatively low stellar mass, which would push it into the color selection region for BBGs. We show this effect in Figure 12, where we depict IRAC colors [3.6] − [4.5] versus redshift for our BBGs compared to the mass-limited and color-selected sample. Typically, the BBGs present colors around [3.6] – [4.5] = 0.2 mag (which are consistent with the average stellar population models shown in

Table 5
Stellar Properties of Our Sample of BBGs at $z > 3$

	ID	z	SFRsed ($M_\odot$ yr^{-1})	SFR_{2800} ($M_\odot$ yr^{-1})	SFR_{IR} ($M_\odot$ yr^{-1})	SFR ($M_\odot$ yr^{-1})	Mass $\log[M/M_\odot]$	τ (Myr)	Age (Gyr)	t_m (Gyr)	$A(V)$ (mag)	*UVJ*[a]	SFR–M[b]
1	GDN_BBG01	3.8	17	6	1757	1763	$10.4^{10.5}_{10.4}$	324^{379}_{286}	$1.0^{1.1}_{0.9}$	0.5	$2.4^{2.5}_{2.4}$	dSFG	Starburst
2	GDN_BBG02	4.8	292	14	2381	2395	$11.7^{11.8}_{11.7}$	315^{353}_{280}	$1.0^{1.1}_{0.9}$	0.5	$2.7^{2.8}_{2.7}$	dSFG	Starburst
3	GDN_BBG03	5.2	318	1	1039	1040	$11.8^{11.8}_{11.7}$	313^{358}_{281}	$1.0^{1.1}_{0.9}$	0.5	$3.4^{3.4}_{3.3}$	dSFG	MS
4	GDN_BBG04	5.1	26	3	⋯	26	$11.0^{11.1}_{11.0}$	204^{230}_{177}	$1.0^{1.1}_{0.9}$	0.6	$1.9^{2.0}_{1.8}$	dSFG	MS
5	GDN_BBG05	4.4	23	3	1108	1111	$10.8^{10.9}_{10.8}$	248^{279}_{222}	$1.0^{1.1}_{0.9}$	0.6	$2.0^{2.0}_{2.0}$	dSFG	Starburst
6	GDN_BBG06	4.2	10	1	⋯	10	$10.3^{10.4}_{10.3}$	504^{569}_{444}	$1.4^{1.6}_{1.3}$	0.7	$2.0^{2.1}_{1.9}$	dSFG	MS
7	GDN_BBG07	4.8	12	10	⋯	12	$10.0^{10.1}_{9.9}$	634^{872}_{430}	$0.9^{1.1}_{0.7}$	0.4	$0.6^{0.8}_{0.5}$	bSFG	MS
8	GDN_BBG08	3.8	6	0	⋯	6	$10.5^{10.5}_{10.4}$	324^{380}_{284}	$1.4^{1.6}_{1.2}$	0.9	$1.5^{1.7}_{1.4}$	dSFG	Sub-MS
9	GDN_BBG09	4.2	4	2	⋯	4	$10.4^{10.5}_{10.4}$	316^{356}_{280}	$1.4^{1.6}_{1.3}$	0.9	$1.5^{1.6}_{1.4}$	dSFG	Sub-MS
10	GDN_BBG10	5.1	4	37	⋯	4	$10.8^{10.8}_{10.7}$	151^{180}_{105}	$1.0^{1.1}_{0.9}$	0.7	$1.2^{1.4}_{1.0}$	Quiescent	Sub-MS
11	GDN_BBG11	6.6	154	1927	⋯	154	$11.1^{11.1}_{11.0}$	338^{407}_{288}	$0.7^{0.8}_{0.6}$	0.3	$2.0^{2.0}_{1.9}$	dSFG	MS
12	GDN_BBG12	3.8	4	5	⋯	4	$10.4^{10.4}_{10.3}$	317^{354}_{281}	$1.4^{1.6}_{1.3}$	0.9	$2.0^{2.1}_{1.9}$	dSFG	Sub-MS
13	GDN_BBG13	4.4	20	5	⋯	20	$10.6^{10.6}_{10.5}$	315^{354}_{281}	$1.0^{1.1}_{0.9}$	0.5	$2.0^{2.0}_{1.9}$	dSFG	MS
14	GDN_BBG14	5.9	31	38	⋯	31	$10.4^{10.5}_{10.3}$	323^{385}_{286}	$0.7^{0.8}_{0.6}$	0.3	$1.2^{1.3}_{1.1}$	bSFG	MS
15	GDN_BBG15	5.0	3	3	⋯	3	$10.2^{10.3}_{10.1}$	195^{222}_{169}	$1.0^{1.1}_{0.9}$	0.6	$0.2^{0.2}_{0.1}$	Quiescent	Sub-MS
16	GDN_BBG16	4.6	48	0	9828	9828	$10.9^{11.0}_{10.9}$	328^{401}_{286}	$1.0^{1.1}_{0.9}$	0.5	$2.3^{2.5}_{2.3}$	dSFG	Starburst
17	GDN_BBG17	4.2	7	3	⋯	7	$10.0^{10.1}_{10.0}$	346^{415}_{294}	$1.0^{1.1}_{0.9}$	0.5	$0.7^{0.9}_{0.6}$	dSFG	MS
18	GDS_BBG01	4.8	38	2	⋯	38	$10.8^{10.9}_{10.8}$	318^{355}_{281}	$1.0^{1.1}_{0.9}$	0.5	$1.7^{1.7}_{1.6}$	dSFG	MS
19	GDS_BBG02	5.3	290	3	1839	1842	$11.7^{11.8}_{11.7}$	317^{357}_{283}	$1.0^{1.1}_{0.9}$	0.5	$2.7^{2.7}_{2.7}$	dSFG	Starburst
20	GDS_BBG03	6.3	162	5	⋯	162	$11.1^{11.2}_{11.1}$	315^{354}_{278}	$0.7^{0.8}_{0.6}$	0.3	$1.8^{1.8}_{1.7}$	dSFG	MS
21	GDS_BBG04	4.9	8	1	⋯	8	$10.3^{10.3}_{10.2}$	324^{371}_{291}	$1.1^{1.2}_{1.0}$	0.6	$1.2^{1.3}_{1.0}$	bSFG	MS
22	GDS_BBG05	5.5	44	0	⋯	44	$10.9^{11.0}_{10.9}$	317^{358}_{282}	$1.0^{1.1}_{0.9}$	0.5	$2.0^{2.1}_{1.9}$	dSFG	MS
23	GDS_BBG06	3.5	1	1	⋯	1	$10.2^{10.2}_{10.2}$	805^{913}_{740}	$1.5^{1.6}_{1.3}$	0.7	$2.0^{2.0}_{1.9}$	dSFG	Sub-MS
24	GDS_BBG07	4.6	37	2	⋯	37	$10.8^{10.9}_{10.8}$	323^{375}_{285}	$1.0^{1.1}_{0.9}$	0.5	$1.9^{2.0}_{1.8}$	dSFG	MS
25	GDS_BBG08	5.5	19	2	⋯	19	$11.0^{11.0}_{10.9}$	198^{225}_{177}	$1.0^{1.1}_{0.9}$	0.6	$1.7^{1.8}_{1.7}$	dSFG	Sub-MS
26	GDS_BBG09	5.8	161	4	856	861	$11.1^{11.2}_{11.0}$	329^{396}_{289}	$0.7^{0.8}_{0.6}$	0.3	$2.5^{2.6}_{2.4}$	dSFG	Starburst
27	GDS_BBG10	4.8	46	0	⋯	46	$10.9^{11.0}_{10.9}$	316^{360}_{283}	$1.0^{1.1}_{0.9}$	0.6	$2.0^{2.1}_{1.9}$	dSFG	MS
28	GDS_BBG11	4.9	15	0	⋯	15	$11.2^{11.3}_{11.2}$	158^{177}_{141}	$1.0^{1.1}_{0.9}$	0.7	$2.0^{2.1}_{1.9}$	dSFG	Sub-MS
29	GDS_BBG12	3.4	26	0	⋯	26	$10.3^{10.3}_{10.2}$	392^{441}_{349}	$0.7^{0.8}_{0.6}$	0.3	$2.0^{2.1}_{1.9}$	bSFG	MS
30	GDS_BBG13	5.1	19	0	⋯	19	$11.0^{11.1}_{10.9}$	200^{224}_{178}	$1.0^{1.1}_{0.9}$	0.6	$2.0^{2.0}_{1.9}$	dSFG	Sub-MS
31	GDS_BBG14	4.5	3	0	⋯	3	$11.1^{11.1}_{11.0}$	125^{142}_{110}	$1.0^{1.1}_{0.9}$	0.8	$0.9^{1.0}_{0.8}$	Quiescent	Sub-MS
32	GDS_BBG15	4.7	9	3	⋯	9	$11.0^{11.0}_{10.9}$	157^{177}_{140}	$1.0^{1.1}_{0.9}$	0.7	$2.0^{2.1}_{2.0}$	dSFG	Sub-MS
33	GDS_BBG16	4.5	31	1	⋯	31	$10.8^{10.8}_{10.7}$	316^{353}_{280}	$1.0^{1.1}_{0.9}$	0.5	$2.0^{2.0}_{1.9}$	dSFG	MS

Notes.
[a] Type of galaxy according to the *UVJ* diagram: bSFG or dSFG.
[b] Type of galaxy according to its position with respect to the MS in the SFR–vs.–stellar mass plot: starburst, MS, or sub-MS galaxy.

Figure 11 for MS systems). Remarkably, half a dozen galaxies are characterized by colors as blue as [3.6] – [4.5] = −0.4 mag, and a similar number presents very red colors, [3.6] – [4.5] = +0.5 mag. The blue colors can be attributed to the presence of prominent emission lines, more specifically, Hα + [N II] entering the 3.6 μm filter at $z = 4$–5 and presenting an $EW_0 \sim 300$–1000 Å (see, e.g., Smit et al. 2014, 2015). The blue IRAC colors could also be reproduced with very young stellar populations. As an example, in the plot, we show the colors expected for constant SFH starbursts with ages 1 and 100 Myr and no extinction. We note, however, that those models would not be compatible with the very red colors observed for H − [3.6]. Concerning the galaxies with very red IRAC colors, some of them could be explained with Hα + [N II] entering the 4.5 μm filter at $z \sim 5.2$ but also with a very red model dominated by stellar continuum (such as the one shown in Figure 11 corresponding to a sub-MS galaxy). Some examples of this type of galaxy with strong emission lines are GDN_BBG07, as well as GDN_BBG15 and GDN_BBG17, which were cataloged as LBGs by Bouwens et al. (2015). The average stellar mass ($\log M/M_\odot$) and their quartiles for bSFG (6/33) are $10.2^{10.1}_{10.1}$, less massive compared to dSFG (25/33), $10.8^{11.0}_{10.6}$. Remarkably, these galaxies present similar [3.6] and [4.5] magnitudes compared to dSFGs, but they are not detected at wavelengths longer than 5 μm. This would be consistent with their SED being flatter, corresponding to blue sources with emission lines.

In Figure 10, we showed that the dSFG region (containing 73% of the BBGs) is populated by mainly starburst and MS galaxies (which are also located within the bSFG region). Only three sub-MS galaxies lie in the quiescent region, and the rest are located in the dSFG region while very close to the boundary. The few galaxies consistent with being evolved or in a post-starburst state at $z \gtrsim 4$ are mainly located in the sub-MS or MS regions. As discussed in the previous sections, it is difficult to identify these candidates reliably because most

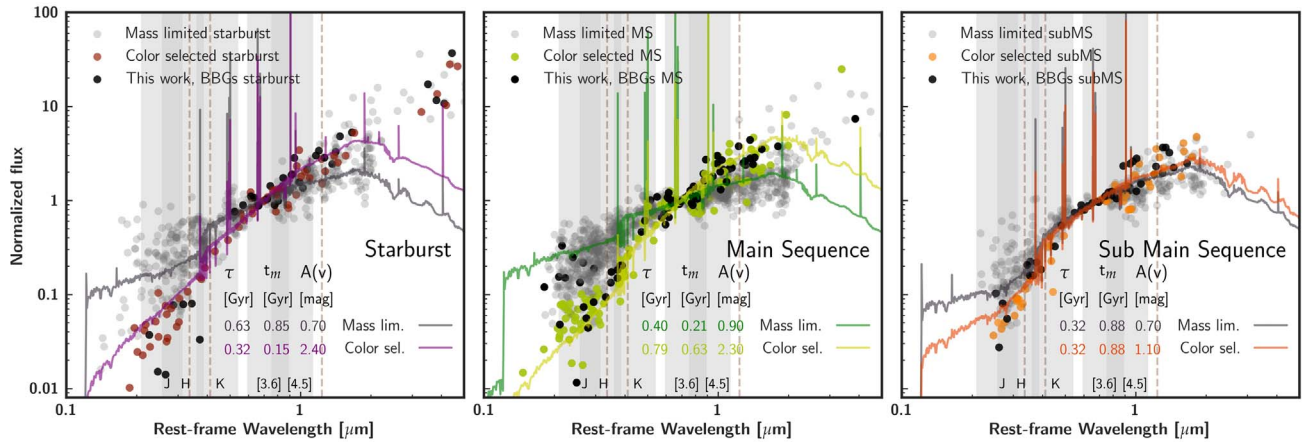

Figure 11. Rest-frame average SEDs for the CANDELS comparison samples and our BBG sources grouped by the different galaxy types according to their position with respect to the MS in the SFR–vs.–M plane (see Figure 8). The SEDs are normalized at 0.7 μm rest frame (observed photometry between the IRAC [3.6] and [4.5] selection bands). Sources with uncertain (S/N < 5) photometry were excluded from the stacks. Left panel: rest-frame SEDs of starburst galaxies from the mass-limited sample (gray symbols), the color-selected sample (colored symbols), and the BBGs reported in this work (black symbols). We also show two best-fit stellar population models fitted to the median photometry for the mass-limited and color-selected subsamples. The mass-weighted age and the attenuation of these fitted models are given in the panel. Gray shaded regions delimit the rest-frame wavelength ranges probed by the observed J, H, K, 3.6 μm, and 4.5 μm bands for $z = 4$. Brown dashed lines mark the locations of rest-frame UVJ bands. Middle panel: same as the left panel but for MS galaxies. Right panel: same as the other two panels but for sub-MS galaxies.

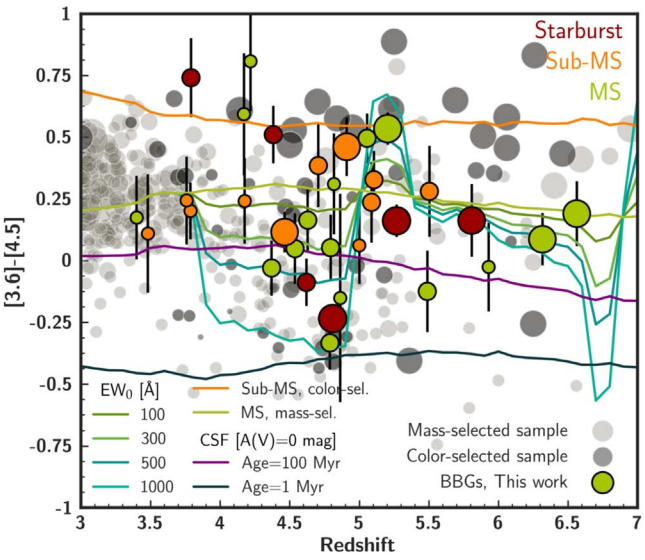

Figure 12. IRAC [3.6]–[4.5] colors vs. redshift for our BBGs compared to the mass-limited and color-selected samples, shown by light and dark gray circles, respectively. The BBGs are color-coded by star formation subsamples, and all of the galaxies are scaled by stellar mass. The green and orange lines show the colors of the MS and sub-MS best-fitting templates (see Figure 11 and main text for details). The effect of including emission lines (Hβ+[O III], Hα +[N II]) of different EWs in the MS best-fitting template is shown by different blue and green lines (see legend). For example, Hα+[N II] at $z \sim 4.5$ produces bluer [3.6]–[4.5] colors (when the lines are in the 3.6 μm IRAC channel) and redder colors for $z \sim 5.2$ (when the same lines enter the 4.5 μm band). Similar effects could be expected for other templates. The colors for no-extinction, constant SFH starbursts of 1 and 100 Myr are shown in dark blue and purple, respectively.

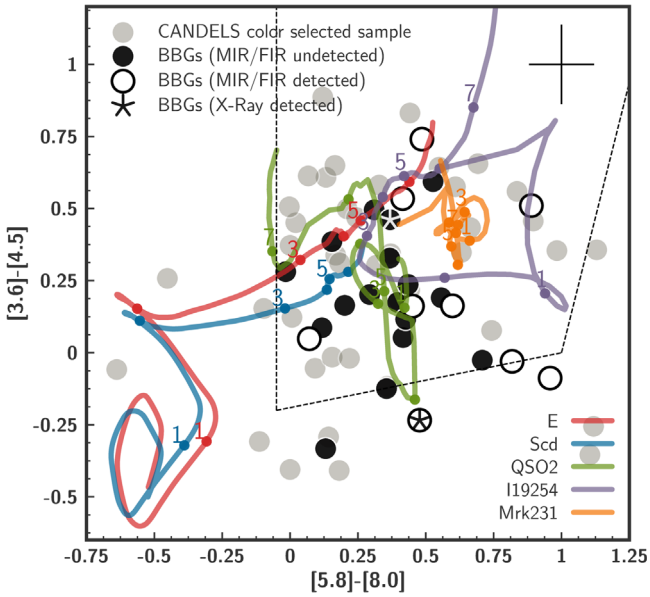

Figure 13. IRAC [3.6]–[4.5] vs. [5.8]–[8.0] color–color diagram and the corresponding AGN selection wedge (dashed line) from Stern et al. (2005; see also Donley et al. 2012). Only seven sources are not shown due to the nondetection in the IRAC 5.8 and/or 8.0 μm channels. We plot the expected colors for an elliptical and a late-type spiral galaxy (assuming the templates found in Coleman et al. 1980 at different redshifts). We also depict the colors for an IR-bright Sy2 galaxy (I19254; Berta et al. 2011), a Type 2 QSO (Polletta et al. 2007), and a Type 1 Seyfert galaxy (Mrk 231; Berta et al. 2011). These templates have been attenuated using a Calzetti et al. (2000) law and $A(V) = 1$ mag. Error bars are not plotted for clarity, but the average values are shown at the top right corner.

BBGs and color-selected galaxies are located near the quiescence boundary, and uncertainties in the observed photometry and photometric redshifts can easily scatter them in or out of the dead galaxy region. In Figure 11, we marked the pivot points of the UVJ colors and rest-frame wavelength ranges probed by the observed JHK, 3.6 μm, and 4.5 μm bands to show the main challenges in characterizing the rest-frame colors at $z \gtrsim 3$ with our current data sets, especially for red sources. Indeed, the rest-frame U band is probed by the H band, where our BBGs are, by definition, undetected or very faint. Consequently, those galaxies undetected in the H band were considered to have $U - V$ lower limits. In addition, the rest-frame J band lies in the IRAC 5.8 μm band, which is significantly shallower ($\sim$2 mag) than 3.6 or 4.5 μm.

Table 6
Number Densities for the BBGs Presented in This Work and the Color- and Mass-limited Comparison Samples Provided for Different Ranges of Redshift and Mass

Redshift		Φ [10^{-6} Mpc^{-3}]			
		Total	Mass-sel.	Color-sel.	BBGs
$z > 3$	$M/M_\odot > 10^{11}$	12.6 ± 2.0	10.4 ± 1.8	4.9 ± 1.2	2.5 ± 0.9
		(42)	(81% ± 15%)	(38% ± 11%)	(19% ± 7%)
	$M/M_\odot = 10^{10–11}$	220.1 ± 6.1	112.7 ± 5.9	9.5 ± 1.7	7.7 ± 1.5
		(393)	(94% ± 5%)	(8% ± 1%)	(6% ± 1%)
$4 < z < 6$	$M/M_\odot > 10^{11}$	10.6 ± 2.0	7.3 ± 2.0	6.1 ± 1.9	3.4 ± 1.4
		(19)	(68% ± 21%)	(58% ± 19%)	(32% ± 14%)
	$M/M_\odot = 10^{10–11}$	90.0 ± 7.1	79.4 ± 6.7	11.7 ± 2.6	11.2 ± 2.5
		(162)	(88% ± 8%)	(13% ± 3%)	(12% ± 3%)
$3 < z < 4$	$M/M_\odot > 10^{11}$	14.4 ± 3.7	14.4 ± 3.7	…	…
		(15)	(100% ± 26%)	(…)	(…)
	$M/M_\odot = 10^{10–11}$	214.1 ± 14.3	209.3 ± 14.2	8.6 ± 2.9	4.8 ± 2.1
		(223)	(98% ± 7%)	(4% ± 1%)	(2% ± 1%)
$4 < z < 5$	$M/M_\odot > 10^{11}$	11.7 ± 3.5	8.5 ± 3.0	6.4 ± 2.6	3.2 ± 1.8
		(11)	(73% ± 28%)	(55% ± 25%)	(27% ± 17%)
	$M/M_\odot = 10^{10–11}$	120.8 ± 11.3	107.0 ± 10.6	17.0 ± 4.2	14.8 ± 4.0
		(115)	(88% ± 9%)	(14% ± 4%)	(12% ± 3%)
$5 < z < 6$	$M/M_\odot > 10^{11}$	9.5 ± 3.3	5.9 ± 2.6	5.9 ± 2.6	3.5 ± 2.0
		(8)	(62% ± 31%)	(62% ± 31%)	(38% ± 24%)
	$M/M_\odot = 10^{10–11}$	55.6 ± 8.1	48.5 ± 7.6	5.9 ± 2.6	7.1 ± 2.9
		(47)	(87% ± 14%)	(11% ± 5%)	(13% ± 5%)

Note. For each range, the total number of galaxies and the relative percentages are also given in parentheses. Uncertainties in the densities and percentages have been calculated assuming Poisson statistics and taking into account the photometric redshift and smPDFs.

Summarizing our results in this section, the vast majority of BBGs correspond to dSFGs, a smaller fraction correspond to bSFGs, and only three BBGs are identified as quiescent galaxies. A few other BBGs are still marginally consistent with being quiescent based on the *UVJ* diagram due to the relatively large photometric uncertainties. Particularly, other quiescent galaxies might be those having older mass-weighted ages (Figure 9) and no FIR detections.

5.5. AGNs in the BBG Sample

Here we study whether some of the BBGs could host an obscured AGN. As discussed in Section 4.1, ∼25% (8/33) of the sample in MIPS 24, and ∼15% (5/33) is detected at longer wavelengths. At the average redshift of the BBGs, $z \sim 4$, these IR detections probe the rest-frame MIR emission, which can be linked to star formation or obscured AGN activity. Figure 13 shows the distribution of the BBGs in the Stern et al. (2005) IRAC color–color diagram, which is widely used to study the likelihood of AGN emission. It is clear from the figure (see Donley et al. 2012) that while this color–color plot is very efficient for identifying strong MIR emission in low- and intermediate-redshift galaxies, it is more ambiguous at $z \gtrsim 3$–4, given that the evolutionary tracks for all of the galaxy templates, from ULIRGs to elliptical galaxies, have colors inside the AGN selection wedge (dashed line). Thus, it is not surprising that most of the BBGs and the galaxies in the color-selected sample are found in the AGN region. We conclude that this color–color diagram is not a reliable diagnostic to understand the nature of the IR emission in our sample of high-z faint sources. Therefore, we can only indicate that, based on the IR detections, around 25%–30% of the BBGs may harbor (bright) obscured AGNs, although no conclusive proof can be presented at this stage. Many more galaxies (up to 75%) present colors that are consistent with Type 2 AGNs (similar to the I19254 or QSO2 templates) but also with obscured/evolved star formation. In addition to the IR data, two of the BBGs are detected in X-rays (asterisks in Figure 13). Given the high redshift of these galaxies, such detection implies a large intrinsic luminosity (L_X(2–10 keV) $> 10^{43}$ erg s^{-1}) typically associated with the presence of a Type 2 or even Type 1 AGN. Note that ∼25% of the BBGs that indeed correspond to the bluest sub-MS and MS galaxies are not shown in the diagram due to their nondetection beyond 4.5 μm.

5.6. Quantification of the Role of BBGs in Galaxy Evolution

Our understanding of the $z > 3$ galaxy population relies largely on samples of UV-selected galaxies typically characterized by blue colors and prominent Lyman breaks (Bouwens et al. 2015). However, it is currently unknown if these galaxies are representative of the massive $z > 3$ galaxy population, or even if any of these galaxies harbor evolved stellar populations outshined in the UV/optical by recent bursts. In this sense, it is important to analyze the contribution of our sample of BBGs to the known population of $z > 3$ galaxies. The most relevant statistical numbers are given in Table 6, and we discuss them in the following paragraphs.

The CANDELS mass-limited sample presented at the beginning of Section 5 comprises 414 galaxies (53 of them also belong to the color-selected sample). The 33 BBGs introduced in this paper only represent a small fraction (8% ± 1%) of the general population (i.e., adding up the mass-limited sample and the BBGs) of massive ($M > 10^{10}$ $M_\odot$), $z > 3$ galaxies in the

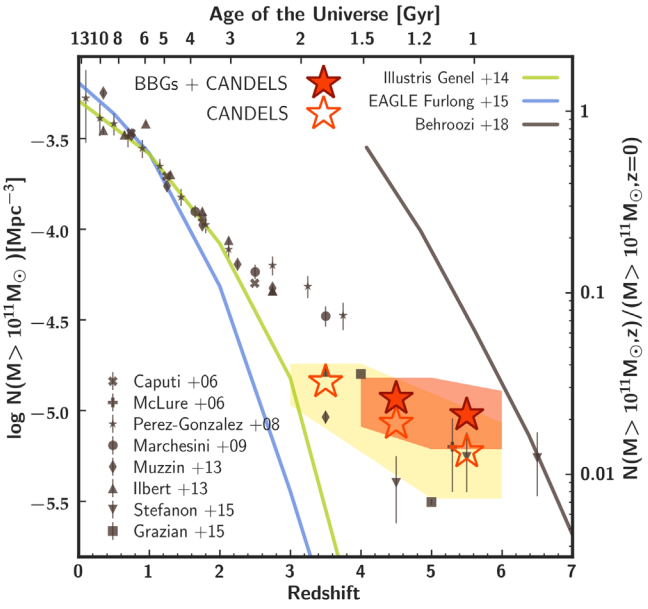

Figure 14. Number density of massive ($M > 10^{11} M_{\odot}$) galaxies as a function of redshift (shown on the bottom horizontal axis; the corresponding age of the universe is shown on the top axis). The fraction of the local number density (computed as the average value of the numbers given in Baldry et al. 2012 and Bernardi et al. 2013) is also shown in the secondary vertical axis. Values from the literature are plotted with different gray symbols, as detailed in the legend. The values inferred from the CANDELS sample at redshifts $z = 3$–4, 4–5, and 5–6 are shown with open red stars, with errors depicted as a yellow shaded area. The total number densities, including the BBGs reported in this work, at the same redshift intervals are shown with filled red stars, with uncertainties plotted as a red shaded area. The gray line corresponds to the number density threshold evolution for a stellar mass of $10^{11}\ M_{\odot}$ presented in Behroozi & Silk (2018). The blue and green lines represent the number densities predicted by the EAGLE (Furlong et al. 2015) and Illustris (Genel et al. 2014) simulations.

GOODS fields. Nonetheless, they do represent a significant fraction ($33\% \pm 13\%$) of the reddest subpopulation of massive galaxies at $z > 3$. Furthermore, our selection technique is especially effective at selecting galaxies at $z = 4$–6, recovering a fraction of $23\% \pm 5\%$ and $43\% \pm 9\%$ of the mass- and color-selected samples, respectively.

The analysis of our sample of BBGs and the comparison samples also points out that 80%–100% of the most massive ($\log M > 10^{11}\ M_{\odot}$) galaxies at $4 < z < 6$ are red ($H - [3.6] > 2.5$ mag). This percentage decreases to 25%–30% for $\log M = 10^{10-11}\ M_{\odot}$. With the sample of BBGs presented in this paper, we have doubled the number of known red massive galaxy candidates at $4 < z < 6$: the CANDELS catalog includes 32 $\log M > 10^{10}\ M_{\odot}$ galaxies, and we detect 26 more. We remark that our sample of BBGs is biased toward the massive end of the stellar mass function: it accounts for $27\% \pm 17\%$ of the total number density of galaxies at $4 < z < 5$ and $38\% \pm 19\%$ at $5 < z < 6$, in both cases for $\log M > 10^{11}\ M_{\odot}$. For lower-mass galaxies, our sample, and red galaxies in general, are minor contributors (~10%) to the global population.

In absolute number density numbers, presented in Figure 14, and for $4 < z < 6$ and $M > 10^{11}\ M_{\odot}$, the mass-limited sample extracted from the CANDELS catalog presents a number density $(7.3 \pm 0.2) \times 10^{-6}$ galaxies Mpc^{-3}. This is consistent with the estimations presented in Stefanon et al. (2015), which range between 5×10^{-6} and 8.3×10^{-6} galaxies Mpc^{-3} (depending on assumptions on the calculation of photometric redshifts) and take into account very faint (or even undetected) NIR sources in the ULTRAVista field detected in IRAC (down to $[3.6] \sim 23.4$, 1 mag brighter than our analysis). These number densities are also consistent with the ones obtained by integrating the Schechter functions presented in Grazian et al. (2015; 5.9×10^{-6} galaxies Mpc^{-3}) and Duncan et al. (2014; 5.0×10^{-6} galaxies Mpc^{-3}) for stellar mass functions based on H-band selected samples. The slight discrepancy can be attributed in part to the systematic differences between the Schechter function fits in those papers and the stellar mass function data points at the massive end. It is worth noticing that, although models of galaxy evolution are capable of properly reproducing the number densities at low redshift ($z \lesssim 2$), current simulations underpredict the observed values of massive galaxies ($\log M > 10^{11}\ M_{\odot}$) at higher redshifts. As shown in Figure 14, EAGLE (Furlong et al. 2015) values plunge at $z \sim 1.5$, while in Illustris (Genel et al. 2014), it occurs at $z \sim 2.5$. The reason for this mismatch between models and observations is unclear. But our results, which are consistent with other estimations at $z > 2$ shown in Figure 14, clearly point to a rapid early evolution of the star formation in some halos resulting in the appearance of very massive galaxies in the first Gyr of the history of the universe, resembling more of a quick monolithical collapse than a gentle hierarchical assembly. The observed values are, however, well below the threshold calculated by Behroozi & Silk (2018) as the limit of number densities for massive galaxies imposed by the current cosmological paradigm, which, according to these authors, could not be surpassed with our current knowledge of the physics governing the evolution of the universe. From the observational point of view, in order to place more robust constraints on modern theoretical models, we need to better constrain the stellar masses and overcome the spatial resolution limitations that our current MIR data have.

To wrap up our results, taking into account our BBGs, we have obtained a more complete census of massive galaxies at $z > 3$ by adding IRAC sources undetected (or faint) in the H and K bands. We conclude that the total number density of $M > 10^{11}\ M_{\odot}$ at $4 < z < 6$ is 1.1×10^{-5} galaxies Mpc^{-3}. This corresponds to 3% of the total number of local $M > 10^{11}\ M_{\odot}$ galaxies (considering the median value of those obtained by integrating the local stellar mass functions in Baldry et al. 2012 and Bernardi et al. 2013); i.e., nearly one of every 30 massive galaxies in the local universe must have assembled more than $10^{11}\ M_{\odot}$ of its mass in the first 1.5 Gyr of the universe.

6. Summary and Conclusions

Combining ultradeep data taken in the *HST*/WFC3 F160W and *Spitzer*/IRAC 3.6 and 4.5 μm bands, we have identified a sample of 33 IRAC-bright/optically faint BBGs at high redshift within the two GOODS fields. Our sample is composed of extremely red ($H - [3.6] \gtrsim 2.5$ mag) and relatively bright MIR ($[3.6] < 24.5$ mag) galaxies. This translates to the following physical properties, according to our analysis of X-ray–to–radio SEDs: the typical BBG is a massive galaxy with a stellar mass $\langle \log(M/M_{\rm sun}) \rangle = 10.8$ lying at redshift $\langle z \rangle = 4.8$, and BBGs harbor relatively young stellar populations (mass-weighted age $\langle t_m \rangle = 0.6$ Gyr) with significant amounts of dust ($\langle A(V) \rangle = 2$ mag), although the range of stellar properties is wide.

We have analyzed the sample of BBGs by comparing them with mass-limited ($M > 10^{10}\ M_{\odot}$ and $z > 3$) and color-selected ($H - [3.6] \geqslant 2.5$) samples extracted from the CANDELS

catalogs published for these fields. We have found that our BBGs substantially differ from the galaxies in the mass-limited sample, which are bluer in general, while their physical properties are quite similar to those in the color-selected sample. However, our BBGs are too faint in the rest-frame UV and optical to be included in typical NIR-selected samples in this redshift range.

The $H-[3.6]$ red colors of most of our sources are compatible with heavily extincted starbursts or relatively evolved populations. However, our sample includes a distinct population of blue (in both their observed $H-[3.6]$ and their *UVJ* rest-frame colors) galaxies. This population has similar SEDs to galaxies from the mass-limited sample. They indeed present uncommon blue [3.6]–[4.5] colors that might be caused by the presence of an emission line in the [3.6] band (converting them in red sources in our selection color $H-[3.6]$). We also note that these possible line emitters correspond to some of the less massive ($M<10^{10.5}\ M_{\odot}$) BBGs in our sample. Therefore, their detection might be a consequence of our improved photometric technique to recover faint sources and reliable upper limits.

We have also demonstrated that $H-[3.6]$ color and IRAC magnitude cuts imply a redshift selection. The redshift distributions of both the BBGs and the color-selected sample peak at $z=4$–5, while the mass-limited sample presents an exponentially decreasing redshift distribution (typical of flux-limited samples). We are also more effective at selecting galaxies at $z=4$–5 than any other sample of *H*-band faint galaxies in the literature. Our selection criterion is also adequate to uncover the high-mass end of the stellar mass function ($M\gtrsim 10^{10}\ M_{\odot}$). The BBG stellar mass distribution peaks at $M\sim 10^{10.5}M_{\odot}$. The color-selected sample presents a comparable histogram with a longer tail at higher masses due to their brighter IRAC magnitudes. The mass-limited sample, in contrast, presents a distribution that decreases with increasing masses.

From the SED modeling, we find strong evidence that massive red galaxies at $z=3$–6 span a diverse range in stellar population properties. In order to understand the nature of the heterogeneous sample of BBGs, we have divided the sources in three star formation regimes according their position with respect to the MS: starbursts, MS, and sub-MS galaxies. Analyzing the average SEDs of BBGs, we confirm that, in general, mass-limited galaxies present bluer SEDs than those in the BBG and color-selected samples. However, as mentioned before, we identify a subsample of BBGs in the MS that are blue and harder to separate from the general population probed by a mass-limited sample. In addition, we find a considerable number of sub-MS galaxies (33% of our sample), most of them with $M<10^{10.5}\ M_{\odot}$, characterized by lower attenuations ($A(V)\sim 1.5$ mag) and older mass-weighted ages ($t_m\sim 0.7$ Gyr). On the other hand, starbursts are found among the most massive ($M>10^{10.5}M_{\odot}$) galaxies in the BBG (20% of the total number of BBGs are starbursts) and color-selected (15%) samples. Starbursts are characterized by high attenuations ($A(V)\sim 2.5$ mag) and young ages ($t_m\sim 0.5$ Gyr). We remark that the total IR luminosity for five of the six starbursts has been calculated with three to five data points (within the wavelength range that dominates the integrated IR luminosity), which translates to relatively small errors. This suggests that a significant fraction of the BBGs (∼25 and up to ∼75%) might host an obscured AGN. The MS galaxies represent a constant proportion of BBGs (∼50%) and color-selected (∼60%) up to the highest masses, $M\sim 10^{11.5}\ M_{\odot}$. However, an important fraction (25%) of the MS galaxies from the color-selected sample have been assigned an SFR lower limit (given their detection by MIPS but their high redshift that prevents them from obtaining a robust SFR estimation) and may correspond to starburst galaxies. The BBGs in the MS present a larger scatter in their attenuations, mass-weighted ages, and *UVJ* colors.

Subdividing BBGs by their rest-frame *UVJ* colors, we find that most of the BBGs correspond to dSFGs (80% of the sample), a smaller fraction to bSFGs (10%), and the rest to quiescent galaxies. Although several studies have reported the existence of galaxies with suppressed star formation at that epoch (e.g., Straatman et al. 2014), just three of our BBGs lie within the quiescent wedge, and three more BBGs have mass-weighted ages that are old enough ($t_m\geqslant 0.9$ Gyr) to be consistent with evolved galaxies. However, 50% of our sample (16 galaxies) is not detected in the *H* band down to magnitudes fainter than ∼27 mag and, therefore, only count with $U-V$ lower limits. Out of those, 10 galaxies (∼30% of the entire sample) not detected by MIPS or *Herschel* could still be identified as quiescent galaxies in the *UVJ* diagram, although no conclusive proof of their nature can be inferred given the high uncertainties.

We have found that the red BBGs presented in this work account for 8% of the total number density of $\log(M/M_{\odot})>10$ galaxies at $z>3$ found by public catalogs such as those from CANDELS or 3D-*HST*. Our BBGs are, however, a significant contributor (30%) to the general (adding cataloged galaxies and our BBGs) population of $\log(M/M_{\odot})>11$ galaxies at $4<z<6$. Our analysis also reveals that while 80%–100% of the most massive ($M>10^{11}\ M_{\odot}$) galaxies at $4<z<6$ are red, this percentage decreases for lower-mass galaxies. We remark that with the sample of BBGs presented in this paper, we have doubled the number of known red massive galaxy candidates at $4<z<6$: the CANDELS catalog includes 32 $M>10^{10}M_{\odot}$ galaxies, and we have detected 26 more. Hence, accounting for this kind of object is key to understanding the population of massive galaxies at high redshift. Adding the BBGs presented in this work to the known population of $4<z<6$ and $M>10^{11}M_{\odot}$, we have found a total number density of 1.1×10^{-5} galaxies Mpc^{-3}. This represents 3% of the number density of local $M>10^{11}M_{\odot}$; i.e., nearly one in 30 massive galaxies in the local universe must have assembled more than $10^{11}M_{\odot}$ in the first 1.5 Gyr of the universe. We compare these numbers with state-of-the-art galaxy formation simulations, such as Illustris and EAGLE, finding that, while the models do a reasonably good job up to $z\sim 2$, they fail to predict very massive ($M>10^{11}M_{\odot}$) galaxies at $z\gtrsim 3.5$, such as those presented in this paper by orders of magnitude.

Spectroscopic follow-up observations in both the optical and NIR are critical to confirm the redshifts and better characterize the properties of this heterogeneous population of red massive galaxies at $z=3$–6 missed by the deepest (mainly NIR-selected) studies. Imaging in NIR and MIR wavelengths, together with spectroscopy from the *JWST*, will be essential to understanding their nature, and ALMA will be crucial in constraining the amount of dust and gas in these systems, as well as discriminating between dust-enshrouded star formation and obscured AGN activity.

We acknowledge support from the Spanish Programa Nacional de Astronomía y Astrofísica under grants AYA2015-63650-P and BES-2013-065772. N.C. acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (MINECO) under grant No. AYA2016-75808-R, which is partly funded by the European Regional Development Fund (ERDF). This work has made use of the Rainbow Cosmological Surveys Database, which is operated by the Universidad Complutense de Madrid (UCM) partnered with the University of California Observatories at Santa Cruz (UCO/Lick, UCSC). This research has made use of the software packages SExtractor, IRAF (Tody 1993), DAOPHOT routine, and STILTS (http://www.starlink.ac.uk/stilts/) software, provided by Mark Taylor of Bristol University, England. This work also employed Astropy, a community-developed core Python package for astronomy (Astropy Collaboration et al. 2013); APLpy, an open-source plotting package for Python (Robitaille & Bressert 2012); Matplotlib (Hunter 2007) and Numpy; and Photutils (Bradley et al. 2016).

Appendix A
HST-based Photometric Measurements for BBGs in the Optical and NIR Bands

In this section, we present the method used to measure consistent and reliable photometry in the optical and NIR *HST* bands for the BBGs presented in this work. The intrinsically faint nature of these galaxies in the optical and NIR makes the construction of robust SEDs an extremely challenging task. Indeed, by definition, BBGs are very faint and, in many cases, even undetected in the optical and NIR bands. Only their MIR fluxes are strongly detected, with all the sources being clearly detected by IRAC. Twenty percent of the sources are also detected by MIPS and 15% are also detected by *Herschel* in the FIR.

As described in Section 4.1, our BBGs cannot be found, by definition of the selection, in the CANDELS and 3D-*HST* photometric catalogs published by Guo et al. (2013) and Skelton et al. (2014), respectively. These galaxies were most likely missed in the catalogs due to limitations (incompleteness) in the source detection method. However, a forced photometric measurement using small apertures does recover reliable fluxes in the optical and NIR *HST* bands for some of our BBGs.

In order to measure those fluxes with the highest S/N possible (or at least set upper limits based on the background noise), the choice of an appropriate photometric aperture is critical. The CANDELS and 3D-*HST* photometric catalogs are based on isophotal magnitudes corrected to Kron (1980) magnitudes (see Section 3 in Guo et al. 2013), with an imposed minimum aperture size of 2.08 pixels (0.″125). Only a small fraction (<2.5%) of the sources in the CANDELS catalog have isophotal radii smaller than 2 pixels, and we find that all of our BBGs that are marginally detected in the *HST* stacked images exhibit radii larger than ∼3–4 pixels (see below).

We focus first on understanding and characterizing the brightness profile of the BBGs to identify the most appropriate aperture size for the flux measurement that maximizes the S/N. To do so, we created several different stacks of BBGs and also of unresolved (stellar) pointlike sources and (faint) color-selected galaxies in the CANDELS sample in order to compare their average properties.

First, we created individual stacks for each one of the 33 BBGs by combining all of the deep *HST* optical and NIR imaging. These stacks are shown in Appendix C (postage stamp in the lower left corner). Twenty of the 33 BBGs are clearly ($>5\sigma$) detected in these stacks, and six of those 20 (e.g., GDN_BBG07 or GDN_BBG15) are marginally resolved or exhibit multiple knots in their morphology. For these six resolved galaxies, the photometric apertures are determined based on their isophotal sizes measured in the *HST* stack (i.e., as in the CANDELS catalog). Their radii range from $r = 0.''6$ to $0.''9$ (see Table 7). For the other 14 galaxies detected in the stacks, we analyze their growth curves, and we find that the S/N of the recovered flux is maximum for apertures of $r \sim 0.''4$. For comparison, the median and quartiles for the Kron-based apertures for faint ($H > 26$ mag) galaxies present in the CANDELS catalogs are $0.4^{0.5}_{0.''3}$. Similarly, this value is also consistent with the aperture size of the photometric measurements in the 3D-*HST* catalog, $r = 0.''35$ (Skelton et al. 2014).

We further study the average brightness profile of the BBGs by creating a stack of all of the galaxies. We first stack all of the *HST* bands for each galaxy, and then we stack the galaxies together. For clarity, we create two such stacks, one for the six extended galaxies and another for the remaining 27, as we expect their profiles to be slightly different. The galaxy stacks are computed following the method of Dole et al. (2006). Briefly, the procedure consists of three steps. First, we create a WFC3 stack in each field. Second, we extract ($3'' \times 3''$, as shown in Figure 15) square images centered around each source and mask all known sources (i.e., those in the CANDELS catalog). Finally, we sum up all postages after applying different rotations to them. More precisely, we sum each image and its horizontally flipped analogous, and we also add each one of these images rotated by 90°, 180°, and 270°; i.e., we use the image for each source eight times in total. During the last stacking step, outlier pixels are rejected. The last steps increase the S/N of the stack and provide more accurate average light profiles. The final average light profile of the BBGs is shown in Figure 15.

For comparison purposes, we also create two additional stacks, one for clean, well-detected, pointlike sources (stars) in the field and another for faint galaxies in the CANDELS color-selected sample described in Section 5. The sample of pointlike sources is selected based on photometric and morphological criteria (see Pérez-González et al. 2008; Barro et al. 2011a, 2011b). Specifically, we selected stars in the fields with an FWHM value smaller than 4.2 pixels; i.e., we rejected sources with FWHM larger than $0.''25$ (the nominal FWHM in F160W is $0.''17$–$0.''19$; Koekemoer et al. 2011; Guo et al. 2013). Then, we created three different stellar stacks by combining the individual postages in different ways, namely, (1) a direct stack of all of them, (2) a rotated and flipped stack such as the one described above for BBGs, and (3) a stack in which the center of each individual postage is shifted randomly within the typical rms error in the position of BBGs without *HST* detection ($0.''3$; see Section 4.1). Figure 15 shows the stellar stacks computed with methods 1 and 3, as well as two stacks for the color-selected sample computed using either the rotation+flip method or the random variation of the centering (*blurring*) within the typical astrometric precision for the BBGs. The stacks based on the first two methods yield very similar FWHMs of $0.''17$ and $0.''20$, almost identical to the nominal FWHM in F160W. The third stack, however, exhibits

Table 7
Properties of the Photometric Apertures Used for the BBGs in This Work

	ID	Fiducial[a] (arcsec)	*a* (arcsec)	*b* (arcsec)	Angle (deg)	Other Apertures Considered (arcsec)	Morphology in the *HST* Stacks[b]
1	GDN_BBG01	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
2	GDN_BBG02	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
3	GDN_BBG03	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
4	GDN_BBG04	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
5	GDN_BBG05	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
6	GDN_BBG06	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
7	GDN_BBG07	⋯	0.75	0.53	0	⋯	*HST* extended
8	GDN_BBG08	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
9	GDN_BBG09	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
10	GDN_BBG10	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
11	GDN_BBG11	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
12	GDN_BBG12	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
13	GDN_BBG13	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
14	GDN_BBG14	⋯	0.8	0.5	90	⋯	*HST* extended
15	GDN_BBG15	⋯	0.6	0.5	60	⋯	*HST* extended
16	GDN_BBG16	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
17	GDN_BBG17	⋯	0.6	0.5	80	⋯	*HST* extended
18	GDS_BBG01	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
19	GDS_BBG02	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
20	GDS_BBG03	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
21	GDS_BBG04	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
22	GDS_BBG05	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
23	GDS_BBG06	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
24	GDS_BBG07	⋯	0.7	0.36	115	⋯	*HST* extended
25	GDS_BBG08	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
26	GDS_BBG09	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
27	GDS_BBG10	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
28	GDS_BBG11	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
29	GDS_BBG12	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
30	GDS_BBG13	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
31	GDS_BBG14	⋯	0.9	0.34	55	⋯	*HST* extended
32	GDS_BBG15	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint
33	GDS_BBG16	0.4	⋯	⋯	⋯	0.65, 0.4+AC	*HST* ultra-faint

Notes.
[a] Fiducial photometric aperture radius (no aperture correction was applied).
[b] Type of galaxy according to the *HST* stacks.

a noticeable broadening of the light profile ($0\farcs43$) very similar to that for the BBGs, although the latter present brighter wings.

Based on the comparison of all of the stacks, we conclude that the bulk of the BBG sample (27 galaxies, ∼80%) consists of unresolved or marginally resolved (at the resolution of F160W) galaxies, while the other six BBGs, which are detected individually, might present some extended emission (e.g., a faint disk). The FWHM obtained from the stack of unresolved BBGs ($r = 0\farcs42$) is fully consistent with the value obtained by fitting the profile of the color-selected stack ($r = 0\farcs46$), and it is also similar to the typical sizes of the faint galaxies in the CANDELS catalog, whose aperture sizes are given above.

Based on the results from the analysis of the light profiles, we decided to use three different methods to measure the *HST*-based photometry of the BBGs. This means that for each galaxy, we analyzed three different SEDs with the same IRAC fluxes but varying optical/NIR photometry. We use these three different SED types to quantify the impact of the intrinsic faintness of the objects and the uncertainties in size measurements on the SED-derived properties.

The first photometric method is based on fixed circular apertures with a radius of $r = 0\farcs4$, and we do not apply any aperture correction; i.e., we assume that the BBGs are small, faint sources similar to the color-selected sample in the CANDELS catalog (which does not apply any aperture correction to the Kron apertures of similar size to ours). As mentioned above, this aperture size provides the highest S/N in the individual *HST* stacks; therefore, we chose it as our fiducial value for analysis purposes throughout the paper. The second method is based on the same aperture photometry with $r = 0\farcs4$, but in this case, we do apply an aperture correction to account for possible missing flux. We compute this correction from the average BBG profile (building a growth curve with it) measured in the galaxy stack of undetected BBGs (light green curve in Figure 15). For this measurement, we also assume that the full size of the average profile is given by the radius where the flux is equal to 1σ of the background in the stacked image, $r = 0\farcs65$. The aperture correction implied by the average light profile for BBGs from $0\farcs4$ to $r = 0\farcs65$ is 0.25 mag. Note that for a pointlike source (a star) accounting for an uncertainty in the center determination, the aperture correction would be 0.18 mag. For reference, with this aperture, we recover reliable ($>5\sigma$) fluxes in the F160W image for 17 galaxies, and the typical S/N for these measurements is 7.5.

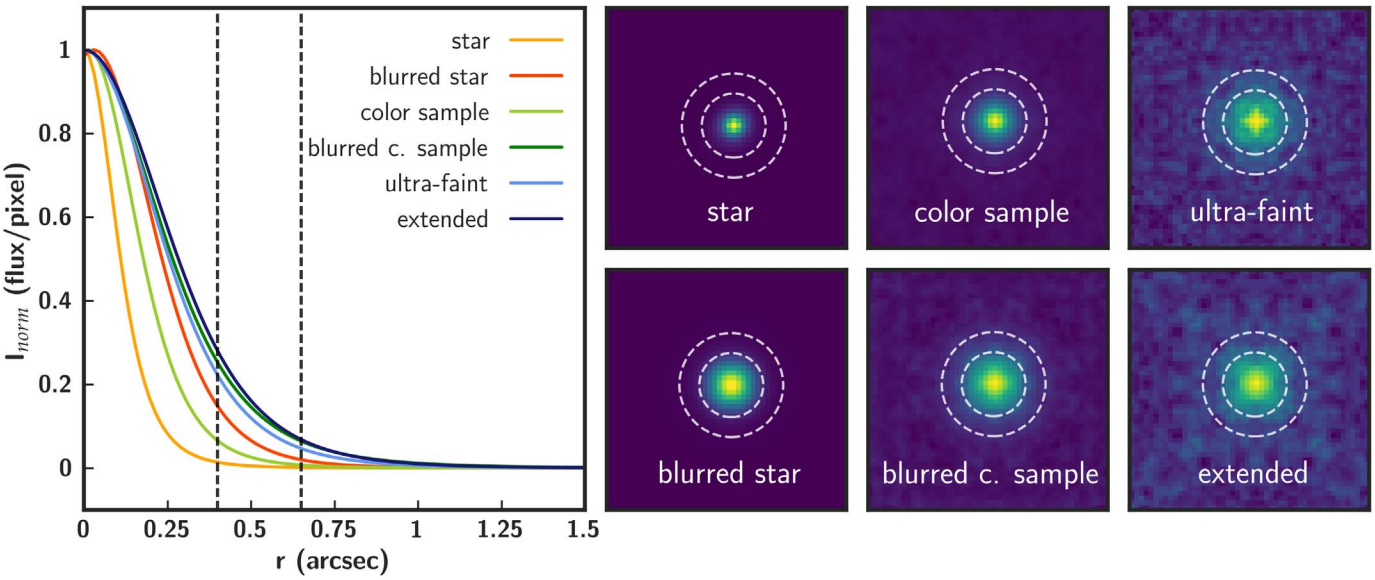

Figure 15. Left: intensity profiles for the different stacked *HST* images described in the text. These stacks include an average stellar profile, a stellar "blurred profile" taking into account random centering errors with rms $0\farcs3$, the average and blurred profiles for the CANDELS color-selected sample, the average profile for the six extended BBGs, and the average profile for the rest (27) of the BBGs. Right: stacked images for the previously mentioned samples of sources. Vertical lines on the left and circles in the postage stamps on the right show the photometric apertures considered in the paper (as discussed in Appendix A).

Lastly, the third method is based on circular apertures with a radius of $r = 0\farcs65$. As mentioned above, this radius roughly corresponds to the full size (i.e., 100% of the flux) of the BBGs as determined from the galaxy stack. Therefore, no aperture correction was applied in this case. For reference, using this aperture, we recovered reliable ($>5\sigma$) fluxes in the F160W band for 14 galaxies, and the typical S/N for these measurements is 6.

The comparison between the integrated F160W magnitudes obtained with the small ($r = 0\farcs4$ with aperture correction) and large ($r = 0\farcs65$) apertures for the 11 unresolved galaxies with reliable ($>5\sigma$) F160W detections yields an average difference and rms of $\Delta m = 0.2 \pm 0.2$; i.e., both types of photometry are consistent within the errors.

In the main text, the $r = 0\farcs4$ aperture photometry without aperture correction is used as the fiducial photometry (unless the source was clearly detected in the *HST* stack; see Table 15). Nonetheless, throughout the paper, we use the two other photometric measurements to study the impact of varying the *HST*-based fluxes on the overall SED of the BBGs and on other SED-based properties, such as redshifts and stellar masses. For example, we use the zPDFs derived from the fitting of the SEDs obtained with the different methods to estimate the uncertainties in the photometric redshift distribution of the BBGs.

Appendix B
Comparison Samples from the Public CANDELS Catalogs

B1. Construction of the CANDELS Samples

The sample of red BBGs presented in this work is built up by extremely faint galaxies at optical wavelengths. In order to understand their nature, we constructed two comparison samples of IRAC-bright ([3.6] and [4.5] < 24.5 mag), *H*-band detected ($H \lesssim 27$ mag) objects based on the GOODS-S and GOODS-N CANDELS catalogs.

We first built a sample of optically faint ($[H > 25$ mag]), extremely red ($H - [3.6] > 2.5$ mag) sources (the color-selected sample). This is directly comparable to our selection biases, since we imposed an IRAC magnitude cut of 24.5 and our sources were (in principle) *H*-band nondetections, which means that they are fainter than ~27 mag. This translates to a color $H - [3.6] > 2.5$ mag. In addition, we constructed a complementary stellar mass–limited ($M > 10^{10} M_{\odot}$) sample cut at $z > 3$ (the mass-limited sample,). As we have shown in Section 5.1, these cuts roughly characterize our sample of BBGs.

In the construction of comparison samples, we discarded sources with uncertain photometry in the 3.6 μm, 4.5 μm, and *H* bands. Specifically, those sources for which the synthetic (inferred from SED fitting) and observed photometry did not match ($m_{\rm synth} - m_{\rm obs} > 0.6$ mag) were removed. This magnitude difference was identified with contamination from nearby objects (e.g., spikes from stars). We note that the synthetic magnitudes could be biased due to the presence of strong emission lines. However, for IRAC, strong emission lines in high-redshift sources typically affect the photometry by less than 0.5 mag (see Schenker et al. 2013; Stark et al. 2013). We also excluded sources located in the edges or regions where the F160W exposure time (<1.5 ks) and limiting magnitude ($5\sigma \sim 27$ mag) are lower. That low accuracy in the fitting could lead to false photo-*z* values and, consequently, to erroneous predicted properties. Those sources with high uncertainties in the F160W band (>0.3 mag) were also discarded. In summary, the mass-limited sample comprises 414 galaxies (i.e., 1.4 sources arcmin^{-2}; 193 in GOODS-N and 221 in GOODS-S) with $M > 10^{10}$ $M_{\odot}$; [3.6], [4.5] < 24.5 mag; and $z > 3$, while the color-selected sample comprises 53 galaxies (i.e., 0.18 sources arcmin^{-2}; 20 in GOODS-N and 33 in GOODS-S) with IRAC [3.6], [4.5] < 24.5 mag; $H > 25$ mag; and H–[3.6] $\geqslant 2.5$ mag. The median values and first and third quartiles of their main properties are summarized in Table 8.

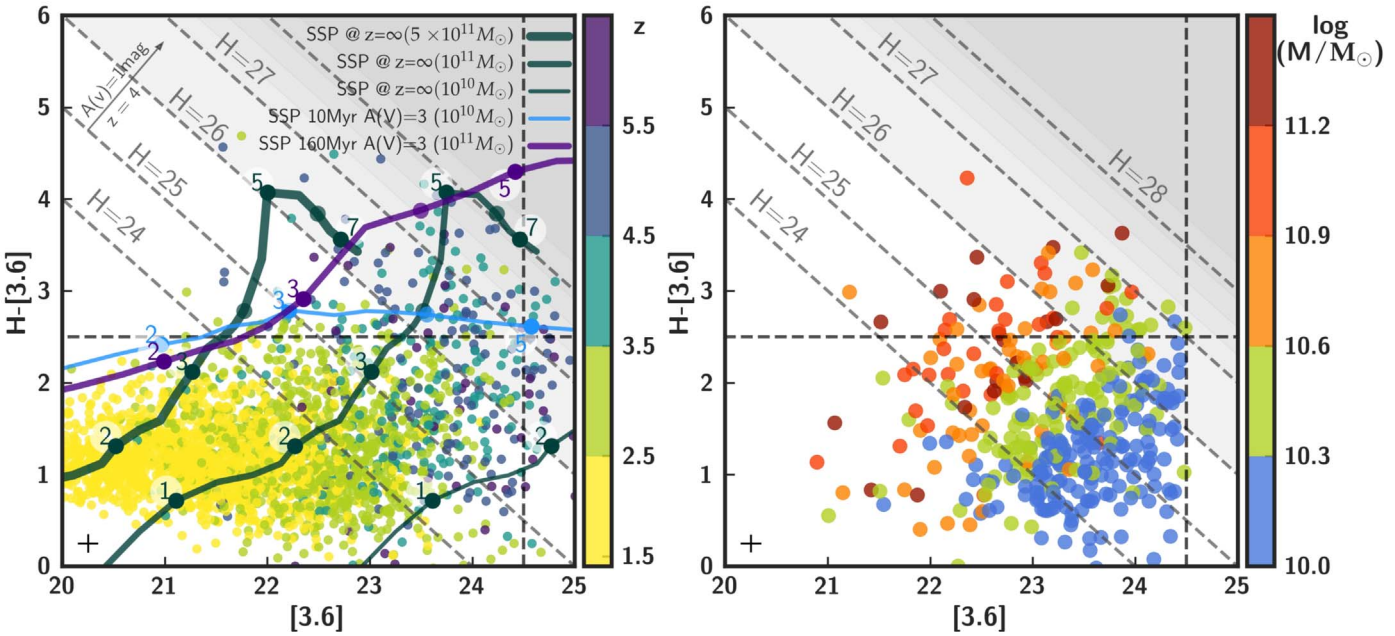

Figure 16. Left panel: the $H - [3.6]$ color plotted vs. the observed [3.6] magnitude for $M > 10^{10}\,M_{\odot}$ CANDELS sources at $z > 1.4$, color-coded by their redshift. Error bars are not plotted for clarity, but the average values are shown in the bottom left corner. We show Bruzual & Charlot (2003) evolutionary tracks of a maximally old SSP for different redshifts (filled circles) with masses 10^{10}, 10^{11}, and $5 \times 10^{11}\,M_{\odot}$. We also plot evolutionary tracks for extincted ($A(V) = 3$ mag) SSPs with age 10 ($M = 10^{10}\,M_{\odot}$) and 160 ($M = 10^{11}\,M_{\odot}$) Myr in cyan and purple, respectively. The 1 mag attenuation vector at $z = 4$ is shown in the upper left corner. Dashed transverse lines depict constant H-band magnitudes. Dashed horizontal and vertical lines show the color and magnitude limits imposed in our selection of the BBG candidates presented in this work and the color-selected sample. Right panel: observed-frame $H - [3.6]$ vs. [3.6] for the mass-limited comparison sample, color-coded by stellar mass. Error bars are not plotted for clarity, but the average values are shown in the bottom left corner. These plots can be directly compared to Figure 5 in the main text, where we plot the results for our sample of BBGs.

Table 8
Statistical Properties of the CANDELS Samples

Sample	z	H (mag)	[3.6] (mag)	[4.5] (mag)	$H - [3.6]$ (mag)	M ($M_{\odot}$)
Mass-limited	$3.8^{4.7}_{3.3}$	$24.9^{25.7}_{24.2}$	$23.3^{23.8}_{22.8}$	$23.2^{23.8}_{22.6}$	$1.6^{1.1}_{2.1}$	$10.4^{10.6}_{10.1}$
Color-selected	$4.7^{5.3}_{4.1}$	$26.5^{26.8}_{25.9}$	$23.6^{23.9}_{23.1}$	$23.4^{23.9}_{22.7}$	$2.8^{2.6}_{3.1}$	$10.8^{11.1}_{10.4}$

Note. Median values and first and third quartiles of their redshift, magnitudes, colors, and masses are shown.

B2. Colors, Redshifts, and Masses of the Comparison Samples

The left panel of Figure 16 shows massive ($M > 10^{10}\,M_{\odot}$) galaxies at $z > 1.4$ from the CANDELS GOODS-N and GOODS-S catalogs. Figure 16 reveals a rough relationship between the position in a color–magnitude diagram ($H - [3.6]$ versus [3.6]) and the redshifts and masses of the galaxies: increasing redshifts lead to redder colors and fainter magnitudes. These red colors can be caused either by the Balmer break/D4000 spectral feature (redshifted beyond the F160W filter at $z > 3$) typical of intermediate and evolved stellar populations or by a dusty starburst with significant UV attenuation. The right panel of Figure 16 shows only massive ($M > 10^{10}$) galaxies at $z > 3$. This mass-limited sample unveils how masses (roughly traced by the IRAC magnitudes) increase parallel to the H-band magnitude constant lines. For galaxies with similar stellar mass, fainter ones present redder colors, while for a given [3.6] magnitude, redder colors will correspond to more massive galaxies and higher redshifts. As indicated by the evolutionary tracks for maximally old quiescent galaxies ($M = 10^{10}$, 5×10^{10}, 10^{11}, and $5 \times 10^{11}\,M_{\odot}$), a red $H - [3.6] \gtrsim 2.5$ mag color is a good proxy to identify massive evolved galaxies or massive dusty galaxies at $z > 3$ (BBGs). Young SFGs would enter the $H - [3.6] \gtrsim 2.5$ mag region at $z \sim 2$ for attenuations of $A(V) \gtrsim 3$ mag. We note that those $z < 3$ galaxies with very high attenuations would most probably be detected by MIPS. We also remark that at $z < 3$, this kind of very dusty starburst would be characterized by relatively bright IRAC magnitudes ($[3.6] \lesssim 22$ mag), and very few sources are found in that region of the color–magnitude diagram for the mass-selected sample. With all this in mind, we conclude that the color and magnitude cuts for both the BBGs and color-selected sample (shown with the dashed vertical and horizontal lines) are effective at selecting massive ($M \gtrsim 10^{10}$) galaxies at high redshift ($z \gtrsim 3$), with little contamination from lower-redshift galaxies or low-mass objects. It is also important to notice that most of the color-selected sources are indeed included in the mass-selected sample.

We also assessed the effect of mass and SFR in the $U - V$ and $V - J$ rest-frame synthetic colors (Figure 17). Both colors tend to increase parallel to the attenuation vector. The *UVJ* diagram has been proven to be very efficient for isolating the red sequence of galaxies at $z < 3$. We use the definition found in Whitaker et al. (2011) to differentiate quiescent galaxies

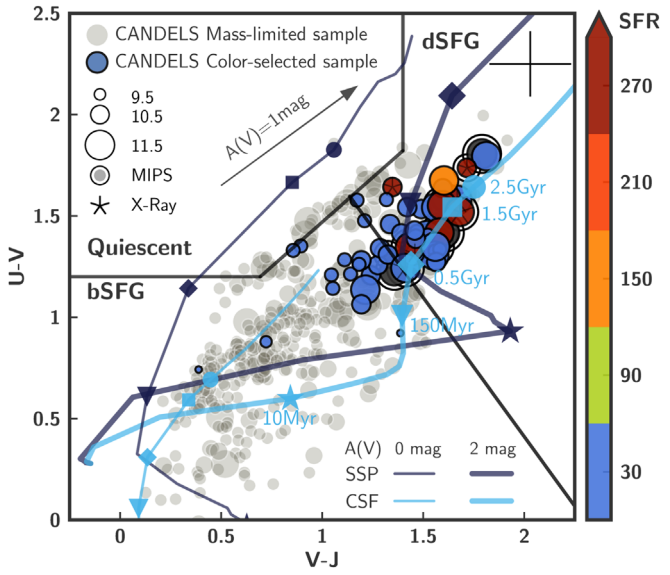

Figure 17. Rest-frame $U - V$ vs. $V - J$ colors for the CANDELS color-selected sample, color-coded by SFR and sized by stellar mass. The SFR lower limits are shown in dark gray. The mass-limited sample is plotted in gray with the size also scaled according the masses. The MIPS-detected galaxies are enclosed by a circle, and the X-ray-detected galaxies show an asterisk inside the symbol. The Whitaker et al. (2011) upper boundary (black wedge) separates quiescent galaxies (top left) from SFGs (bottom). The black diagonal line denotes an additional criterion to separate bSFGs from dSFGs. Error bars are not plotted for clarity, but the average values are shown next to the right boundary of the quiescent region. The 0 and 2 mag extincted SSP and CSF models are shown with thin and thick solid lines, respectively. The 1 mag attenuation vector computed assuming a Calzetti et al. (2000) reddening law is also shown.

from those actively forming stars:

$$\begin{aligned}(U - V) &> 1.2\\ (V - J) &< 1.4\\ (U - V) &> 0.88 \times (V - J) + 0.59.\end{aligned} \quad (4)$$

Within the *UVJ* star-forming region, there exists a wide range of colors that could be caused by differences in the amount of reddening and/or the SFHs. Several criteria have been proposed to distinguish between red and blue SFGs using an additional $U - V$ and/or $V - J$ color criterion (i.e., Wuyts et al. 2007; Fumagalli et al. 2014; Spitler et al. 2014; Straatman et al. 2014). The reddening vector, computed by assuming a Calzetti et al. (2000) reddening law, suggests that the large range of colors observed in galaxies can be explained by different amounts of extinction. But there also exists a degeneracy between age, attenuation, and SFR, as suggested by earlier shallower surveys (e.g., Marchesini et al. 2010). In order to understand the nature of our sample of BBGs, we have tested in Figure 17 the behavior in the *UVJ* diagram of SSP and constant SFR (CSF) models with different amounts of extinction. Figure 17 also shows the implications of our definition of bSFGs and dSFGs. The dSFGs present $U - V$ and $V - J$ colors that can be reproduced by models with $A(V) = 2$ mag of attenuation and ages older than ∼10 and ∼500 Myr for an SSP or CSF SFH, respectively. Younger or less extincted systems would qualify as bSFGs. Figure 17 shows that in order for an object to enter the quiescent galaxy wedge, it must harbor SSPs (or short star-forming bursts, i.e., SFHs that differ significantly from a CSF model) with ages older than 0.5 Gyr and no dust. If some dust is present, younger galaxies might be found in the quiescent locus, and old (>2 Gyr) galaxies might leave that region and be found in the upper right corner of the *UVJ* diagram.

Not surprisingly, in analyzing the comparison samples in terms of the *UVJ* colors, we find that the mass-limited sample presents significantly bluer colors and lower masses than the color-selected one. If we only consider the color-selected sample, the most massive sources are those with the reddest colors (as also reported by Williams et al. 2010) and mainly correspond to dSFGs. We find 22 bSFGs, out of which one (5%) is detected by MIPS 24 μm. We have 31 dSFGs in the color-selected sample, 11 of them (35%) showing IR emission. The upper left sector that delimits the quiescent region contains a few galaxies from the mass-limited sample but none from the color-selected sample. However, their closeness to the boundary, together with the high uncertainties, makes its classification quite uncertain. Moreover, there is a small number of galaxies in the color-selected sample that lie very close to the boundary and are characterized by very low SFRs and high mass-weighted ages. Therefore, they might be either quiescent or post-starburst galaxies.

Not surprisingly, most of the FIR-detected galaxies (26%) in the color-selected sample are found among the most massive sources. They are located within the dSFG locus and present very red *UVJ* colors. Considering the possible presence of an obscured AGN, we note the existence of six X-ray emitters also in the same area. Among them, three (50%) present MIPS 24 μm emission, and two are also detected by PACS and SPIRE or even at 850 or 1200 μm. Compared to the mass-limited comparison sample, it is evident and expected that the $H - [3.6] > 2.5$ mag color cut biases the selection against bSFGs.

Appendix C
SEDs and Postage Stamps of BBG Candidates

In Figure 18, we present SEDs, stellar and dust emission models, and postage stamps in several bands (including stacked images) for all of the BBGs presented in this paper.

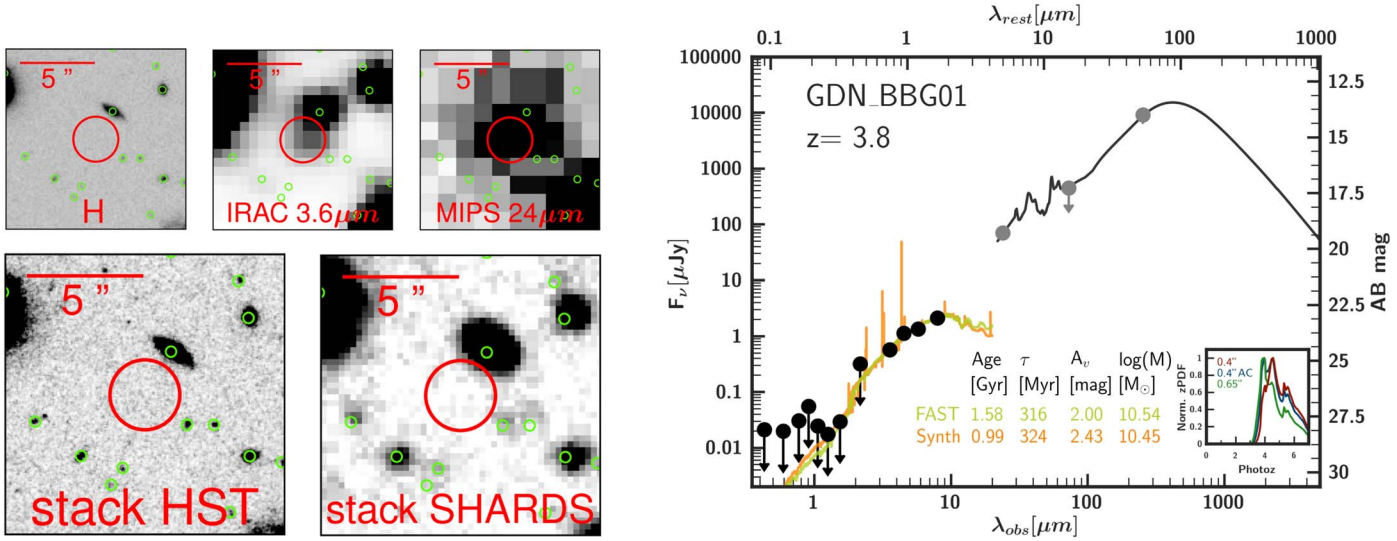

Figure 18. Postage stamps ($12'' \times 12''$) and SEDs for the 33 BBGs presented in this work. For each galaxy, on the left, we plot images taken by *HST* in the *H* band, *Spitzer* in the [3.6] and 24 μm bands, the stack with all *HST* data (including ACS and WFC3 images), and the stack for SHARDS data (in GOODS-N). Green circles in all stamps show sources included in the public catalogs released by the CANDELS team (Guo et al. 2013). The red circle marks the BBG. On the right, we show SEDs including all measured fluxes and upper limits. Bands (most probably) dominated by stellar emission are plotted with black dots, and bands probing the dust emission are plotted with gray dots. These SEDs have been fitted to stellar population models (using two codes, `synthesizer` and `FAST`) whose main physical properties are written in the plot. Dust emission is also fitted with Rieke et al. (2009) templates when the galaxy is detected by MIPS and/or *Herschel*. The inset in these SED plots shows the zPDF obtained for the three photometric methods (based on the different apertures written in the legend) explained in Appendix A, except for extended sources (see Table 7), which only shows one zPDF.

(The complete figure set (33 images) is available.)

ORCID iDs

Belén Alcalde Pampliega https://orcid.org/0000-0002-4140-0428
Guillermo Barro https://orcid.org/0000-0001-6813-875X
Helena Domínguez Sánchez https://orcid.org/0000-0002-9013-1316
Nicolás Cardiel https://orcid.org/0000-0002-9334-2979
Antonio Hernán-Caballero https://orcid.org/0000-0002-4237-5500
Lucía Rodriguez-Muñoz https://orcid.org/0000-0002-0192-5131

References

Alexander, D., Bauer, F., Brandt, W., et al. 2003, AJ, 126, 539
Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539
Ashby, M., Willner, S., Fazio, G., et al. 2013, ApJ, 769, 80
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, MNRAS, 421, 621
Barger, A., Cowie, L., Capak, P., et al. 2003, AJ, 126, 632
Barro, G., Pérez-González, P. G., Gallego, J., et al. 2011a, ApJS, 193, 13
Barro, G., Pérez-González, P. G., Gallego, J., et al. 2011b, ApJS, 193, 30
Behroozi, P., & Silk, J. 2018, MNRAS, 477, 5382
Bell, E. F., Papovich, C., Wolf, C., et al. 2005, ApJ, 625, 23
Bernardi, M., Meert, A., Sheth, R. K., et al. 2013, MNRAS, 436, 697
Berta, S., Magnelli, B., Nordon, R., et al. 2011, A&A, 532, A49
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Borys, C., Smail, I., Chapman, S. C., et al. 2005, ApJ, 635, 853
Bouwens, R., Illingworth, G., Oesch, P., et al. 2012, ApJ, 754, 83
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
Bradley, L., Sipocz, B., Robitaille, T., et al. 2016, Photutils: Photometry tools, Astrophysics Source Code Library, ascl:1609.011
Brammer, G. B., & van Dokkum, P. G. 2007, ApJL, 654, L107
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2011, ApJ, 739, 24
Brinchmann, J., Charlot, S., White, S., et al. 2004, MNRAS, 351, 1151
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, A&A, 554, A70
Calzetti, D. 2001, PASP, 113, 1449
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Caputi, K. I., Cirasuolo, M., Dunlop, J. S., et al. 2011, MNRAS, 413, 162
Caputi, K. I., Deshmukh, S., Ashby, M. L. N., et al. 2017, ApJ, 849, 45
Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, ApJL, 750, L20
Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, ApJ, 810, 73
Caputi, K. I., Lilly, S. J., Aussel, H., et al. 2009, ApJ, 707, 1387
Casey, C. M., Berta, S., Béthermin, M., et al. 2012, ApJ, 761, 139
Chabrier, G. 2001, ApJ, 554, 1274
Chabrier, G. 2003, PASP, 115, 763
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Cimatti, A., Daddi, E., Renzini, A., et al. 2004, Natur, 430, 184
Cole, S. 2011, MNRAS, 416, 739
Coleman, G., Wu, C.-C., & Weedman, D. 1980, ApJS, 43, 393
da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110
Daddi, E., Dannerbauer, H., Stern, D., et al. 2009, ApJ, 694, 1517
Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, ApJ, 626, 680
Dahlen, T., Mobasher, B. & CANDELS Collaboration 2012, AAS Meeting, 220, 133.01
Dale, D. A., & Helou, G. 2002, ApJ, 576, 159
Dale, D. A., Helou, G., Contursi, A., Silbermann, N. A., & Kolhatkar, S. 2001, ApJ, 549, 215
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
de Barros, S., Schaerer, D., & Stark, D. P. 2014, A&A, 563, A81
Dickinson, M., Giavalisco, M., et al. 2003, in The Mass of Galaxies at Low and High Redshift, ed. R. Bender & A. Renzini (Berlin: Springer), 324
Dole, H., Lagache, G., Puget, J.-L., et al. 2006, A&A, 451, 417
Domínguez Sánchez, H., Pérez-González, P. G., Esquej, P., et al. 2016, MNRAS, 457, 3743
Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142
Draine, B., & Li, A. 2007, ApJ, 657, 810
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, MNRAS, 432, 3520
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Fang, J. J., Faber, S. M., Koo, D. C., et al. 2018, ApJ, 858, 100
Fazio, G., Hora, J., Allen, L., et al. 2004, ApJS, 154, 10
Fontana, A., Dunlop, J. S., Paris, D., et al. 2014, A&A, 570, A11
Fontana, A., Salimbeni, S., Grazian, A., et al. 2006, A&A, 459, 745

Fontana, A., Santini, P., Grazian, A., et al. 2009, A&A, 501, 15
Franx, M., Labbé, I., Rudnick, G., et al. 2003, ApJL, 587, L79
Fumagalli, M., Labbé, I., Patel, S. G., et al. 2014, ApJ, 796, 35
Furlong, M., Bower, R. G., Theuns, T., et al. 2015, MNRAS, 450, 4486
Galametz, A., Grazian, A., Fontana, A., et al. 2013, ApJS, 206, 10
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
Giavalisco, M., Dickinson, M., Ferguson, H., et al. 2004a, ApJL, 600, L103
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004b, ApJL, 600, L93
Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, Natur, 544, 71
Grazian, A., Fontana, A., Santini, P., et al. 2015, A&A, 575, A96
Greve, T. R., Pope, A., Scott, D., et al. 2008, MNRAS, 389, 1489
Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3
Grogin, N. A., Kocevski, D. D., Faber, S., et al. 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Heavens, A., Panter, B., Jimenez, R., & Dunlop, J. 2004, Natur, 428, 625
Hill, A. R., Muzzin, A., Franx, M., et al. 2017, ApJ, 837, 147
Hodge, J., Karim, A., Smail, I., et al. 2013, ApJ, 768, 91
Huang, J.-S., Zheng, X., Rigopoulou, D., et al. 2011, ApJL, 742, L13
Huang, L., & Bijaoui, A. 1990, ExA, 1, 311
Hunter, J. D. 2007, CSE, 9, 90
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644
Kajisawa, M., Ichikawa, T., Tanaka, I., et al. 2011, PASJ, 63, 379
Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2
Kashikawa, N., Nagao, T., Toshikawa, J., et al. 2012, ApJ, 761, 85
Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Kennicutt, R. C., Jr., Hao, C.-N., Calzetti, D., et al. 2009, ApJ, 703, 1672
Koekemoer, A. M., Faber, S., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kron, R. G. 1980, ApJS, 43, 305
Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, ApJS, 154, 166
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Lea, S. M., & Kellar, L. A. 1989, AJ, 97, 1238
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Lybanon, M., Lea, S. M., & Himes, S. M. 1994, in Proc. 12th International Conference on Pattern Recognition, Vol. 1 (New York: IEEE), 347
Maccarone, M. C. 1996, VA, 40, 469
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132
Malhotra, S., & Rhoads, J. E. 2002, ApJL, 565, L71
Mancini, C., Matute, I., Cimatti, A., et al. 2009, A&A, 500, 705
Marchesini, D., Muzzin, A., Stefanon, M., et al. 2014, ApJ, 794, 65
Marchesini, D., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 725, 1277
Mármol-Queraltó, E., McLure, R. J., Cullen, F., et al. 2016, MNRAS, 460, 3587
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, ApJ, 842, 21
McCarthy, P. J. 2004, ARA&A, 42, 477
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Michałowski, M., Hjorth, J., & Watson, D. 2010, A&A, 514, A67
Mobasher, B., Dickinson, M., Ferguson, H. C., et al. 2005, ApJ, 635, 832
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Nayyeri, H., Mobasher, B., Hemmati, S., et al. 2014, ApJ, 794, 68
Noeske, K., Weiner, B., Faber, S., et al. 2007, ApJL, 660, L43
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, ApJ, 855, 105
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&ARv, 25, 2
Papovich, C., Moustakas, L. A., Dickinson, M., et al. 2006, ApJ, 640, 92
Penner, K., Pope, A., Chapin, E. L., et al. 2011, MNRAS, 410, 2749
Pentericci, L., Grazian, A., Fontana, A., et al. 2007, A&A, 471, 433
Perera, T. A., Chapin, E. L., Austermann, J. E., et al. 2008, MNRAS, 391, 1227
Pérez-González, P., Egami, E., Rex, M., et al. 2010, A&A, 518, L15
Pérez-González, P. G., Cava, A., Barro, G., et al. 2013, ApJ, 762, 46
Pérez-González, P. G., Rieke, G. H., Egami, E., et al. 2005, ApJ, 630, 82
Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, ApJ, 675, 234
Petrosian, V. 1976, ApJL, 209, L1
Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, A&A, 518, L2
Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81
Pope, A., Borys, C., Scott, D., et al. 2005, MNRAS, 358, 149
Rawle, T. D., Altieri, B., Egami, E., et al. 2016, MNRAS, 459, 1626
Reed, D., Gardner, J., Quinn, T., et al. 2003, MNRAS, 346, 565
Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, Natur, 496, 329
Rieke, G., Young, E., Engelbracht, C., et al. 2004, ApJS, 154, 25
Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, ApJ, 692, 556
Robitaille, T., & Bressert, E. 2012, APLpy: Astronomical Plotting Library in Python, Astrophysics Source Code Library, ascl:1208.017
Rodighiero, G., Cimatti, A., Franceschini, A., et al. 2007, A&A, 470, 21
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJL, 739, L40
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Schaerer, D., de Barros, S., & Sklias, P. 2013, A&A, 549, A4
Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2013, ApJ, 777, 67
Schreiber, C., Elbaz, D., Pannella, M., et al. 2016, A&A, 589, A35
Schreiber, C., Labbé, I., Glazebrook, K., et al. 2018, A&A, 611, A22
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Shimasaku, K., Kashikawa, N., Doi, M., et al. 2006, PASJ, 58, 313
Simpson, J. M., Smail, I., Wang, W.-H., et al. 2017, ApJL, 844, L10
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Smit, R., Bouwens, R. J., Franx, M., et al. 2015, ApJ, 801, 122
Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, ApJ, 784, 58
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, ApJL, 787, L36
Staguhn, J. G., Kovács, A., Arendt, R. G., et al. 2014, ApJ, 790, 77
Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2017, ApJ, 843, 36
Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, ApJ, 803, 11
Stefanon, M., Marchesini, D., Rudnick, G. H., Brammer, G. B., & Whitaker, K. E. 2013, ApJ, 768, 92
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, ApJ, 592, 728
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJL, 783, L14
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2015, ApJL, 808, L29
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Tody, D. 1993, in ASP Conf. Ser. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes (San Francisco, CA: ASP), 173
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJL, 677, L5
Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, Natur, 495, 344
Wang, T., Elbaz, D., Alexander, D. M., et al. 2017, A&A, 601, A63
Wang, T., Elbaz, D., Schreiber, C., et al. 2016, ApJ, 816, 84
Weiß, A., Kovács, A., Coppin, K., et al. 2009, ApJ, 707, 1201
Werner, M., Roellig, T., Low, F., et al. 2004, ApJS, 154, 1
Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012, ApJ, 745, 179
Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, ApJ, 735, 86
Wiklind, T., Dickinson, M., Ferguson, H. C., et al. 2008, ApJ, 676, 781
Williams, R. J., Quadri, R. F., Franx, M., et al. 2010, ApJ, 713, 738
Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011, ApJ, 738, 106
Wuyts, S., Labbé, I., Franx, M., et al. 2007, ApJ, 655, 51
Xue, Y., Luo, B., Brandt, W., et al. 2011, ApJS, 195, 10