Blue rest-frame UV-optical colors in z ~ 8 galaxies from GREATS: very young stellar populations at 650 Myr of cosmic time

Stefanon, M.; Bouwens, R.J.; Labbé, I.; Illingworth, G.D.; Oesch, P.A.; Dokkum, P. van; Gonzalez, V.

The Astrophysical Journal, 927:48 (18pp), 2022 March 1 https://doi.org/10.3847/1538-4357/ac3de7

Blue Rest-frame UV-optical Colors in $z \sim 8$ Galaxies from GREATS: Very Young Stellar Populations at ∼650 Myr of Cosmic Time

Mauro Stefanon[1], Rychard J. Bouwens[1], Ivo Labbé[2], Garth D. Illingworth[3], Pascal A. Oesch[4,5], Pieter van Dokkum[6], and Valentino Gonzalez[7,8]
[1] Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands; stefanon@strw.leidenuniv.nl
[2] Centre for Astrophysics and SuperComputing, Swinburne, University of Technology, Hawthorn, Victoria, 3122, Australia
[3] UCO/Lick Observatory, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[4] Departement d'Astronomie, Université de Genéve, 51 Ch. des Maillettes, CH-1290 Versoix, Switzerland
[5] Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Jagtvej 128, København N, DK-2200, Denmark
[6] Astronomy Department, Yale University, 52 Hillhouse Avenue, New Haven, CT 06511, USA
[7] Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago 7591245, Chile
[8] Centro de Astrofísica y Tecnologias Afines (CATA), Camino del Observatorio 1515, Las Condes, Santiago 7591245, Chile
Received 2021 March 7; revised 2021 November 9; accepted 2021 November 26; published 2022 March 3

Abstract

Deep rest-optical observations are required to accurately constrain the stellar populations of $z \sim 8$ galaxies. Due to significant limitations in the availability of such data for statistically complete samples, observational results have been limited to modest numbers of bright or lensed sources. To revolutionize the present characterization of $z \sim 8$ galaxies, we exploit the ultradeep (∼27 mag, 3σ) Spitzer/Infrared Array Camera (IRAC) 3.6 and 4.5 μm data, probing the rest-frame optical at $z \sim 8$, over ∼200 arcmin^2 of the GOODS fields from the recently completed GOODS Re-ionization Era wide-Area Treasury from Spitzer program (GREATS), combined with observations in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS)/Ultra Deep Survey (UDS) and CANDELS/Cosmic Evolution Survey (COSMOS) fields. We stacked $\gtrsim$100 $z \sim 8$ Lyman-break galaxies in four bins of UV luminosity ($M_{\rm UV} \sim -20.7$ to -18.4 mag) and study their $H_{160} - [3.6]$ and [3.6]–[4.5] colors. We find young ages ($\lesssim$100 Myr) for the three faintest stacks, inferred from their blue $H_{160} - [3.6] \sim 0$ mag colors, consistent with a negative Balmer break. Meanwhile, the redder $H_{160} - [3.6]$ color seen in the brightest stack is suggestive of slightly older ages. We explored the existence of a correlation between the UV luminosity and age, and find either no trend or fainter galaxies being younger. The stacked SEDs also exhibit very red [3.6]–[4.5] ∼ 0.5 mag colors, indicative of intense [O III]+Hβ nebular emission and star formation rate (SFR). The correspondingly high specific SFRs, sSFR $\gtrsim$10 Gyr^{-1}, are consistent with recent determinations at similar redshifts and higher luminosities, and support the coevolution between the sSFR and the specific halo mass accretion rate.

Unified Astronomy Thesaurus concepts: High-redshift galaxies (734); Lyman-break galaxies (979); Galaxy ages (576); Galaxy colors (586); Galaxy properties (615)

1. Introduction

The star formation rate (SFR) and the stellar mass ($M_\star$) are two of the most fundamental physical parameters characterizing a galaxy. The SFR measures its recent (few million to few hundred million years) rate of formation of new stars, while the stellar mass retains the cumulative effects of the (possibly varying) SFR over its entire life (i.e., its star formation history, SFH), combined with its dark matter halo merger history. Comparing the evolution of the specific SFR (sSFR $\equiv$ SFR/$M_\star$) to that of the specific halo mass accretion rate (SHMAR; $\equiv \dot{M}_h/M_h$, where M_h is the dark matter halo mass, and the dot represents the time derivative) across cosmic time thus provides a useful diagnostic for the efficiency of the baryonic mass assembly, the hierarchical merging of the dark matter haloes, and the feedback mechanisms regulating the star formation.

Programs such as Great Ovservatories Origins Deep Survey (GOODS; Giavalisco et al. 2004), Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), 3D-Hubble Space Telescope (HST; van Dokkum et al. 2011; Brammer et al. 2012; Skelton et al. 2014; Momcheva et al. 2016), Hubble Frontier Fields (HFF; Lotz et al. 2017), Cluster Lensing and Supernovae survey with Hubble (CLASH; Coe et al. 2013), Reionization Lensing Cluster Survey (RELICS; Coe et al. 2019) and the Hubble Deep, Ultradeep and Extremely deep field (HDF/HUDF/XDF; Williams et al. 1996; Beckwith et al. 2006; Illingworth et al. 2013) have allowed us to gain insights on, among other properties, the evolution of the sSFR up to $z \lesssim 8$ (González et al. 2010; Labbé et al. 2013; Stark et al. 2013; González et al. 2014; Smit et al. 2014; Salmon et al. 2015; Faisst et al. 2016; Davidzon et al. 2018). A growing number of studies have found that the sSFR monotonically increases with increasing redshift (e.g., Duncan et al. 2014; Faisst et al. 2016; Davidzon et al. 2018), with a factor ∼10 evolution in the last ∼10 Gyr of cosmic history (since $z \lesssim 5$–6).

The situation at higher redshifts is more uncertain. Observational evidence for a further factor ∼5–10 increase in the sSFR in the ∼500 Myr between $5 < z < 8$ (e.g., Labbé et al. 2013; Smit et al. 2014; Salmon et al. 2015; Faisst et al. 2016) contrasts with a quasi-steady value reported by a number of authors (e.g., Mármol-Queraltó et al. 2016; Santini et al. 2017; Davidzon et al. 2018). Furthermore, similarly discrepant results also exist for the evolution of the stellar-to-halo mass ratio.

Observational results have ranged from a constant $M_\star/M_h$ (see, e.g., Durkalec et al. 2015; Stefanon et al. 2017; Harikane et al. 2018a) to an $M_\star/M_h$ ratio evolving with redshift (e.g., Finkelstein et al. 2015a; Harikane et al. 2016). Models show similar variation between an evolving $M_\star/M_h$ ratio (e.g., Behroozi et al. 2013, 2019) and modest evolution (e.g., Tacchella et al. 2018).

These tensions could arise at least in part from systematic uncertainties in age estimates. While some recent work seems to suggest the existence of an age bimodality already at $z \gtrsim 7$, with brighter galaxies characterized by more pronounced Balmer breaks (e.g., Jiang et al. 2016; Castellano et al. 2017), there is also a recent indication of Balmer breaks in lower-mass galaxies as well (e.g., Zheng et al. 2012; Hoag et al. 2018; Hashimoto et al. 2018; Roberts-Borsani et al. 2020). These Balmer breaks perhaps result from more complex physical processes in place already in the first few hundred million years of cosmic history. Nevertheless, evidence indicative of young ages is seen from an increasing number of studies that point toward very strong emission lines at early cosmic times, with typical equivalent widths (EWs) for the most abundant transitions (e.g., $H\alpha$, $H\beta$, [O II], [O III], N II) in excess of several hundred angstrom (e.g., Shim et al. 2011; Labbé et al. 2013; Stark et al. 2013; Smit et al. 2014, 2015; Faisst et al. 2016; Mármol-Queraltó et al. 2016; Reddy et al. 2018; De Barros et al. 2019; Tran et al. 2020; Endsley et al. 2021).

Given the lack of spectra for substantial samples of sources at $z \gtrsim 5$–6, current age estimates rely on broadband photometry bracketing the Balmer/4000 Å break (Bruzual 1983; Hamilton 1985; Balogh et al. 1999; Kauffmann et al. 2003), requiring measurements at these redshifts in the Spitzer/Infrared Array Camera (IRAC) bands. However, the relatively shallow coverage available from IRAC has not allowed us so far to probe the rest-frame optical of individual sources for significant samples of galaxies at $z \gtrsim 6$–7.

Some new Spitzer/IRAC observations allow us to revisit these issues for an HST sample of high-redshift galaxies at $z \sim 8$. In particular, we study the physical properties of a sample of star-forming galaxies at $z \sim 8$ identified as Lyman-break galaxies (LBGs) over four of the CANDELS fields (GOODS-N/S, UDS, and COSMOS). The new data set that enables this study is the recently completed GOODS Re-ionization Era wide-Area Treasury from Spitzer (GREATS) program (PI: Labbé - Stefanon et al. 2021a), which provides coverage in the 3.6 and 4.5 μm bands over $\sim$200 arcmin^2 distributed over the GOODS-N and GOODS-S fields, to ultradeep limits of $\sim$27.1, 26.7 mag (AB, 5σ). This depth enables simultaneous detection in the 3.6 μm and 4.5 μm bands for an unprecedented $\sim$40% of the $z \sim 8$ LBG sample in this study.

Our main results are based on stacked spectral energy distributions (SEDs) we construct from the parent sample of $z \sim 8$ LBGs, which we slice into bins of UV luminosity and UV slope. We considered galaxies at redshift $z \sim 8$ for two main reasons: (1) $z \sim 8$ corresponds to the epoch of instantaneous re-ionization (Planck Collaboration et al. 2016b), thus probing the physical conditions of star formation in the heart of the re-ionization era, and (2) at $7.2 \lesssim z \lesssim 9.0$, the [O III] $\lambda\lambda$4959,5007 and $H\beta$ nebular lines lie within the 4.5 μm band coverage, while the 3.6 μm band is free from strong nebular emission ([O II]λ3727 emission enters the 3.6 μm band with a smaller contribution to the broadband photometry; see Section 5). This provides a solid reference for the measurements of the line intensities and, together with the flux in the H band, for bracketing the ages of the stellar populations.

This paper is organized as follows. In Section 2 we introduce the sample adopted in the present work, which is based on Bouwens et al. (2015). In Section 3 we describe the procedure we followed to generate the stacked SEDs. We present the stacked SEDs in Section 4, and discuss our results in Section 5. Finally, Section 6 presents a summary and our conclusions.

Throughout this paper, we adopt $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70\,\mathrm{km\,s^{-1}}$ Mpc^{-1}, consistent with the most recent estimates from Planck (Planck Collaboration et al. 2016a). Magnitudes are given in the AB system (Oke & Gunn 1983), while $M_\star$ and SFR refer to the Salpeter (1955) initial mass function (IMF). For brevity, we denote the HST F435W, F606W, F775W, F850LP, F105W, F125W, F140W, and F160W as B_{435}, V_{606} and i_{775}, z_{850}, Y_{105}, J_{125}, JH_{140}, and H_{160}.

2. Data Set and Sample Selection

2.1. The Lyman-break Selection

We adopted the sample of Y dropouts assembled by Bouwens et al. (2015) identified over the CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) GOODS-N, GOODS-S (Giavalisco et al. 2004), EGS (Davis et al. 2007), UDS (Lawrence et al. 2007), and COSMOS (Scoville et al. 2007) fields, the ERS field (Windhorst et al. 2011), and the UDF/XDF (Beckwith et al. 2006; Ellis et al. 2013; Illingworth et al. 2013) with the HUDF09-1 and HUFD09-2 parallels (Bouwens et al. 2011).[9] The data over the CANDELS UDS and COSMOS fields, although shallower, provide wide area coverage critical for sampling $L \gtrsim L^*$ galaxies. Table 1 summarizes the main properties of the adopted data sets.

The COSMOS and UDS fields lack coverage in the HST/WFC3 Y_{105} band, key for a robust selection of $z \sim 8$ sources. We therefore complemented the HST photometry in these fields with measurements from the deep ground-based near-IR (NIR) mosaics of the UltraVISTA DR3 (McCracken et al. 2012) and VIDEO (Jarvis et al. 2013) programs, respectively. Additionally, these two fields benefit from deep public Subaru/SuprimeCam imaging in the z band. Overall, these data have 5σ depths in the Y band ranging from 26.7–27.5 mag (GOODS fields) to $\sim$26.0 mag (UDS and COSMOS), $\sim$26.8–27.8 in the J and H bands, and about 27.5 mag in the optical bands.

Photometry on the ground-based data was performed with MOPHONGO (Labbé et al. 2006, 2010a, 2010b, 2013, 2015), which uses the brightness profile of each source from a high-resolution image (a combination of J_{125}, JH_{140}, and H_{160} in our case) to remove the corresponding neighboring objects before performing aperture photometry. For the aperture photometry, we adopted an aperture of 1.″2 diameter, and corrected to total magnitudes using the brightness profile of each source on the low-resolution image and the point-spread function (PSF) reconstructed at the specific locations of each source.

The candidate $z \sim 8$ LBGs were initially selected as Y-band dropouts, with the additional constraint of non-detection at 2σ in each of the bands bluer than Y. Specifically, the following color criteria were adopted (see Bouwens et al. 2015 for

[9] We excluded CANDELS/EGS because of the lack of deep data in the Y band, which makes the selection of candidate $z \sim 8$ LBGs more uncertain.

Table 1
Observational Data Used for the SMF Estimates

Field Name		Area (arcmin^2)	H_{160}[a] 5σ (mag)	IRAC Data[b]	3.6 μm[c] 5σ (mag)	4.5 μm[c] 5σ (mag)
XDF		4.7	29.4	GREATS	∼27.2	∼26.7
HUDF09-1		4.7	28.3	GREATS	∼26.3	∼25.8
HUDF09-2		4.7	28.7	GREATS	∼27.0	25.5–26.0
ERS		40.5	27.4	GREATS	26.2–27.0	25.6–26.7
CANDELS	GOODS-N Deep	62.9	27.5	GREATS	27.0–27.3	26.5–26.8
	GOODS-N Wide	60.9	26.7	GREATS	26.3–27.2	25.8–26.8
	GOODS-S Deep	64.5	27.5	GREATS	∼27.3	26.6–26.9
	GOODS-S Wide	34.2	26.8	GREATS	26.5–27.2	26.2–26.7
	COSMOS	151.9	26.8	SEDS+SCANDELS	26.4–26.7	26.0–26.3
	UDS	151.2	26.8	SEDS+SCANDELS	25.4–26.3	25.0–25.9
Totals:		580.2				

Notes.
[a] 5σ limit from Bouwens et al. (2015), computed from the median of measured uncertainties of sources.
[b] GREATS: Stefanon et al. (2021a); SEDS: Ashby et al. (2013); SCANDELS: Ashby et al. (2015).
[c] Nominal 5σ limit for point sources from the SENS-PET exposure time calculator, based on the exposure time maps. Due to inhomogeneities in the coverage, a range of values is quoted when the depth varies by more than ∼0.2 mag across the field. Because of the combined effects of broad Spitzer/IRAC point-spread function (PSF) and significant exposure times, source blending may reduce the actual depth (see discussion in Labbé et al. 2015).

details):

$$(Y_{105} - J_{125} > 0.45) \wedge (J_{125} - H_{160} < 0.5) \wedge (Y_{105} - J_{125}) > 0.75(J_{125} - H_{160}) + 0.525 \quad (1)$$

where $\wedge$ denotes the logical AND operator. These criteria resulted in an initial sample of 185 sources.

2.2. The Final Sample

To gain information on the rest-frame optical properties of the bona fide $z \sim 8$ sources, we incorporated photometry in the Spitzer/IRAC (Fazio et al. 2004) 3.6 and 4.5 μm bands from the SEDS (Ashby et al. 2013) and S-CANDELS (Ashby et al. 2015) programs for the COSMOS and UDS fields, and from the GOODS Re-ionization Era wide-Area Treasury from Spitzer (GREATS; Stefanon et al. 2021a) for sources in the GOODS-N/S fields. GREATS (PI: I. Labbé) is a recently completed Spitzer legacy program, which brings near-homogeneous ∼200–250 hr depth (corresponding to $5\sigma \sim 26.8$–27.1 mag) in Spitzer/IRAC 3.6 and 4.5 μm for 200 arcmin^2 over the two GOODS fields. Flux densities in the IRAC bands were derived using the same MOPHONGO tool, adopting an aperture of $1\farcs8$ diameter.

We computed photometric redshifts and rest-frame UV_{1600} luminosities with EAZY (Brammer et al. 2008), complementing its default template set with three 2.5 Gyr passively evolving, $Z_\odot$ (Bruzual & Charlot (2003) templates with extinction $A_V = 2.0$, 5.0, and 8.0 mag (Calzetti et al. 2000). Recent studies of high-redshift galaxies have revealed that strong nebular emission is ubiquitous (e.g., $H\alpha$, $H\beta$, [O II], [O III], and N II; Shim et al. 2011; Labbé et al. 2013; Stark et al. 2013; Smit et al. 2014, 2015; Faisst et al. 2016; Mármol-Queraltó et al. 2016; Reddy et al. 2018; De Barros et al. 2019; Tran et al. 2020; Endsley et al. 2021). At $z \sim 8$, the Spitzer/IRAC 3.6 μm band is potentially contaminated by [O II], the 4.5 μm band by $H\beta$ and [O III], and the 5.8 μm band by $H\alpha$ and N II. Therefore, to minimize potential systematic biases in our $z \sim 8$ sample selection, and in the average properties of the sample, we excluded the IRAC bands when computing the photometric redshifts. For sources to be included in our sample, we then required that their photometric redshift be $7.3 \leqslant z_{\rm phot} \leqslant 8.7$. We selected this range to allow for modest redshift errors in localizing the $[\mathrm{O\,III}]_{4959,5007}$ lines to the 4.5 μm band. Finally, we visually inspected the neighbor-cleaned cutouts of each source in the IRAC 3.6 μm and 4.5 μm bands and excluded from the sample those with residuals from nearby MOPHONGO-subtracted IRAC sources overlapping with the nominal position of the source. We also excluded those where the background showed gradients, since both issues could introduce systematic biases during the stacking process.

After applying the above criteria, the final sample of candidate $z \sim 8$ LBGs included 102 sources. Their distribution in absolute magnitude is presented in Figure 1. The heterogeneity of our carefully derived data sets allows us to probe a broad interval in UV luminosities, ranging from $\gtrsim L^*$ to $\sim 0.06L^*$. Remarkably, ∼40% of the sources are detected in the 3.6 μm band with a significance in excess of 2σ.

2.3. Impact of LBG Criteria on Sample Selection

We evaluated the impact that the LBG criteria have on the selection of a representative sample of galaxies at $z \sim 8$ through Monte Carlo simulations, which we discuss in detail in Appendix B. Briefly, we generated synthetic flux densities from a set of SED templates uniformly probing the colors of dusty and passive galaxies in the LBG plane. Specifically, we considered star-forming galaxies with $0.0 \leqslant A_V \leqslant 3.0$ mag, and passively evolving systems that formed in a single burst at $z = 20$. Finally, we applied the LBG color criteria (Equation (1)) after scattering the flux densities according to the noise estimates in each field. Our simulations indicate that the LBG criteria are not significantly biased against dusty sources when $A_V < 1.5$ mag. This supports a scenario where the observed lack of sources with UV slopes $\beta > -1.6$ (see Section 4.3) is more likely to be intrinsic, indicating that the colors from our stacks are not significantly altered. Furthermore, the simulations show that we are able to include ∼70% of the passive sources at $z \sim 8$, resulting in marginal

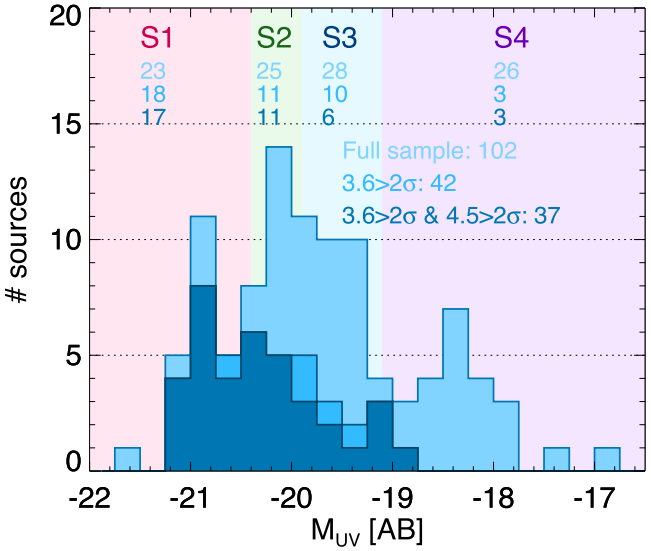

Figure 1. Distribution of the rest-frame UV absolute magnitudes of the sample of $z \sim 8$ candidate LBGs adopted for this work. The four colored background regions correspond to the four bins in absolute magnitude we adopted in our stacking, as labeled at the top. S1 includes galaxies with $M_{\rm UV} < -20.4$ mag, S2 galaxies with $-20.4 < M_{\rm UV} < -19.9$, S3 galaxies with $-19.9 < M_{\rm UV} < -19.1$, while S4 includes galaxies $M_{\rm UV} > -19.1$ mag. This binning scheme was chosen to evenly distribute the number of sources across the four bins and enable the study of trends with luminosity. The different shades of blue correspond to different cuts in IRAC detection significance, as indicated by the legend, where we also quote each sample size. The break-down of these sizes in each $M_{\rm UV}$ bin is presented by the three numbers in the column below the stack labels.

contributions to the colors of our stacks from this class of sources.

The above conclusions rely on the hypothesis of a uniform distribution of red sources as a function of dust extinction or age. Nonetheless, an increasing number of studies are revealing extremely red sources at $z > 3$–4 (see, e.g., Yan et al. 2004; Huang et al. 2011; Caputi et al. 2012, 2015; Stefanon et al. 2015; Wang et al. 2016, 2019; Alcalde Pampliega et al. 2019; Williams et al. 2019; Fudamoto et al. 2021; Gruppioni et al. 2020; Romano et al. 2020; Talia et al. 2021). These sources remain generally undetected even at the NIR wavelengths adopted for the construction of high-redshift galaxy samples. Nevertheless, such samples are still very limited in size, especially at $z > 6$. It is thus difficult to estimate the impact of such sources on the present results, but it would appear to be limited given the rarity of such sources (see also, e.g., Bouwens et al. 2020).

Samples detected at rest-frame optical wavelengths can mitigate potential selection biases against dusty and evolved sources. While the GREATS IRAC mosaics probe the rest-frame optical at $z \sim 8$, the broader PSF of IRAC compared to that of HST significantly increases the difficulty of selections made at such wavelengths. Thanks to the unprecedented sensitivity and resolution of JWST, forthcoming programs like Public Release IMaging for Extragalactic Research (PRIMER; Dunlop et al. 2021) and Ultra-deep NIRCam and NIRSpec Observations Before the Epoch of Reionization (UNCOVER; Labbe et al. 2021) will provide a more direct and comprehensive evaluation of the efficiency of LBG criteria in selecting galaxy populations at high redshifts.

3. Stacking Procedure

Our main results are based on stacking analysis, allowing us to probe the average physical properties of LBGs at $z \sim 8$, while self-consistently including in our analysis all of those sources with limited detection significance in one or more bands.

In our analysis we segregate sources in bins of luminosity (irrespective of UV slope) and of UV slope (irrespective of luminosity). However, in both cases, we applied the same stacking procedure described below.

For the analysis in bins of luminosity, we chose the thresholds in UV magnitudes as a trade-off between maximizing the number of sources in each bin, improving the signal-to-noise ratio (S/N) in the IRAC 3.6 and 4.5 μm bands, and enabling inspection of potential correlations of the results with UV luminosity. Hereafter we will refer to each bin as S1, S2, S3, and S4 in order of decreasing UV luminosity. Specifically, S1 includes galaxies more luminous than $M_{\rm UV} = -20.4$ mag, S2 galaxies with $-20.4 < M_{\rm UV} < -19.9$, S3 galaxies with $-19.9 < M_{\rm UV} < -19.1$, while S4 includes galaxies fainter than $M_{\rm UV} = -19.1$ mag. These luminosity bins were chosen as a trade-off between maximizing the number of sources in each bin and having a sufficient number of bins to quantify possible dependencies on luminosity. The UV magnitude distribution of sources in the four bins is presented in Figure 1.

For the analysis in bins of UV slope, we divided the sample into five bins, centered at UV slopes $\beta = -2.5$, -2.3, -2.1, -1.9, and -1.7, with a width of $\Delta\beta = 0.2$. Since the stacking procedures were similar for both samples, and the luminosity-binned stacks are the primary focus of this paper, the discussion that follows is for the luminosity-selected stacks.

To further mitigate the dependence on the brightness of individual sources within each bin, before deriving the actual stacked fluxes, we normalized the flux density in each band and for each source with the corresponding inverse-variance weighted average of the J_{125} and H_{160} flux densities. The final values were obtained by taking the median of the measurements. Uncertainties were computed bootstrapping the sample 1000 times, drawing with replacement the same number of objects in the considered bin, and taking the standard deviation of the measurements to be the flux uncertainty after perturbing the fluxes according to their photometric uncertainty (see also, e.g., González et al. 2011; Stefanon et al. 2017). This procedure also allows us to naturally account for the intrinsic spread of colors across individual sources in each bin.

We adopted different stacking procedures for the HST and for the IRAC bands, due to the different S/Ns available for the photometric measurements. The higher S/N available for the HST photometry limits the measurement scatter around the true value. This allowed stacking of the HST bands to be performed directly with the photometric measurements. The UDS and COSMOS fields lack coverage in several HST bands compared to the two GOODS fields. Therefore, for each stack, we only derived flux densities for those bands where >90% of sources had coverage in the four fields. Nonetheless, there is a set of bands common to all stacks: V_{606}, J_{125}, and H_{160}. S2, S3, and S4 have measurements in the i_{775} and z_{850}, while S3 and S4 also have Y_{105}. The lack of uniform coverage in the optical bands does not directly affect our analysis, because at $z \sim 8$ we do not expect to register any significant signal in the optical HST filters. The added bands do add robustness to our $z \sim 8$

selections in that they provide stronger discrimination against dusty lower-redshift sources. JH_{140} covers the rest-frame UV at $z \sim 8$. However, its broad wavelength transmission significantly overlaps ($\gtrsim$50%) with both J_{125} and H_{160}, making its use only of marginal value. The Y band is valuable for constraining the redshift as the Lyman break crosses it over the redshift range $z \sim 6.4$–8.7. Nonetheless, the lack of Y-band coverage for all stacks is not a limitation since it is not used to determine physical parameters. The Y-band flux density is dependent on the exact redshift and can not be a reliable determinant of intrinsic fluxes and slopes. Regardless of filter coverage, all available bands are used to determine redshifts.

For the IRAC stacks, the lower S/N of the IRAC bands, instead, could introduce a larger scatter into the stacked measurements if the HST procedure was used. This potentially could affect the median estimates. Given this, stacking in the IRAC bands was performed on the actual data in cutouts centered at the nominal position of each source and cleaned of contamination from neighbors using the procedure of Labbé et al. (2006, 2010a, 2010b, 2013, 2015). The cutouts were aligned to a common location using two approaches that depended on the measured S/N of the individual object in the specific band. For S/N $\geqslant 5$, the location of the source was determined by fitting a Gaussian profile, re-centering the cutout to the new position, and repeating this operation a second time. For S/N < 5, instead, the cutouts were aligned to correspond to the nominal location of the source on the H_{160} mosaic, assuming no significant offset between the position on the H_{160} and that in the IRAC band. To ascertain the robustness of our results against the statistical estimator, we also repeated our stacking analysis using an inverse-variance weighted average, whose main results are presented in Appendix D. Reassuringly, we do not find any significant difference between the two methods.

Stacked IRAC flux densities were measured in circular apertures of $3\farcs6$ diameter and aperture-corrected using the median PSF obtained by combining the PSFs at the location of each source. Aperture correction factors were $\sim$1.4, 1.4, 1.7, and 1.8 for the 3.6 μm, 4.5 μm, 5.8 μm, and 8.0 μm bands, respectively. Flux uncertainties in the IRAC bands were computed by bootstrapping the sample 1000 times, similarly to what was done for the HST bands.

4. Results

4.1. Stacked SEDs at z $\sim$ 8

Figure 2 presents the final image stamps in the four IRAC bands for the four different bins in absolute magnitudes. To provide a better visual context for the data involved in our analysis, we also show median-stacked images in the HST bands, and remind the reader that the median flux densities in the HST bands were estimated from the HST photometry (see Section 3). The stacked photometry of the four SEDs is presented in Table 2, and it is displayed in Figure 3, while in Table 3 we list the main color estimates. Thanks to the increased depth of the GREATS data, the 3.6 μm and 4.5 μm stacks result in a clear detection in the three most luminous bins, with an $\sim 4\sigma$–7σ significance, and a $\gtrsim 2\sigma$ detection in the 4.5 μm band of S4. Remarkably, the exceptional depth of the GREATS mosaics enables a $\sim 1.4\sigma$ measurement in the 3.6 μm band of S4. However, to provide more robust results, in this work we consider as non-detections those flux densities

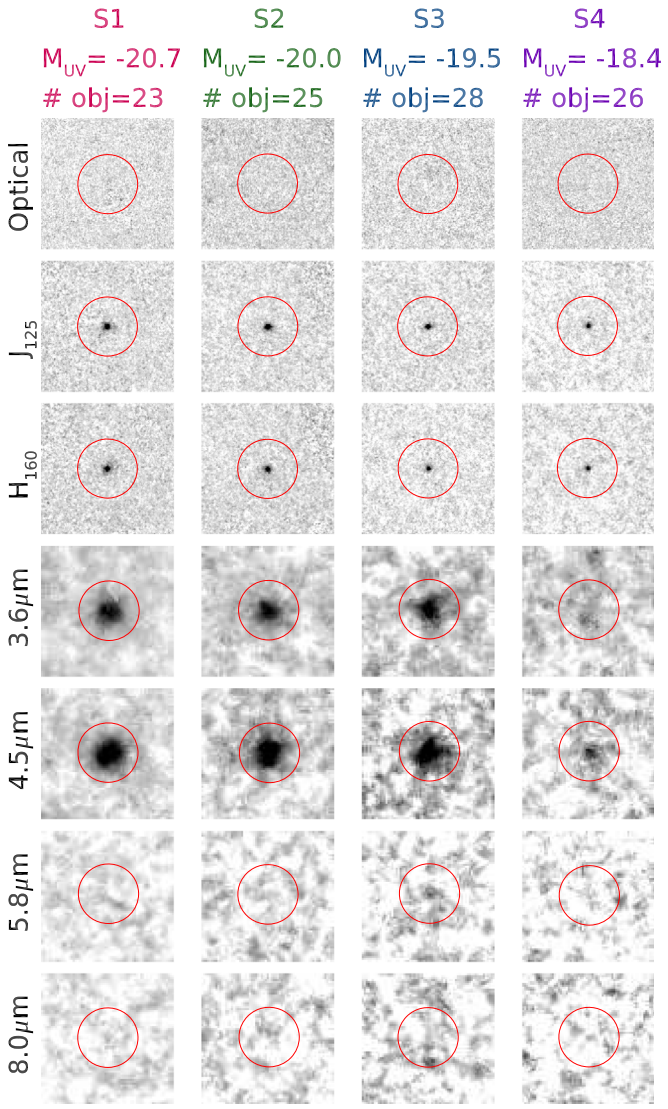

Figure 2. Image stamps of the IRAC and HST median stacks. Each cutout is $\sim 8\farcs0 \times 8\farcs0$ in size. The red circle marks the $3\farcs6$ diameter aperture adopted for the photometry of the IRAC stacks. HST stacks are presented to provide a better visual context of the data involved in our study, as the median flux densities in the HST bands were estimated from the photometry of individual sources. Each row refers to a different band, as labeled on the left side; in particular, the HST optical stack combines all data available in the B_{435}, V_{606}, i_{775}, and z_{850} bands. The columns refer to a stack in a specific luminosity bin, as labeled at the top. The increased coverage offered by the GREATS program enables detections in the 3.6 and 4.5 μm bands to faint limits ($M_{UV} = -19.5$ mag). Remarkably, we find a $\gtrsim 2\sigma$ detection for the flux density in the 4.5 μm band for the faintest luminosity bin (S4, $M_{UV} = -18.4$ mag) and a $\sim 1.4\sigma$ measurement in the 3.6 μm band. The 5.8 and 8.0 μm band stacks, instead, are characterized by non-detections ($<2\sigma$) irrespective of their UV luminosity.

with $<2\sigma$ significance. The stacks in the 5.8 μm and 8.0 μm bands do not show any significant detection in any luminosity bin, as expected.

Table 4 summarizes the main properties of the four stacks. Absolute magnitudes $M_{UV,1600}$ were computed with EAZY, fixing the redshift of each stack to the median redshift of the LBGs entering each group. As a sanity check, we also ran EAZY leaving the photometric redshift as a free parameter. The likelihood distributions of the photometric redshifts ($p(z)$) are reproduced as insets in Figure 3. The $p(z)$ present a single peak centered around $z \sim 8$ increasing our confidence on the nature of galaxies that contribute to the stacks; the broader $p(z)$ of S1

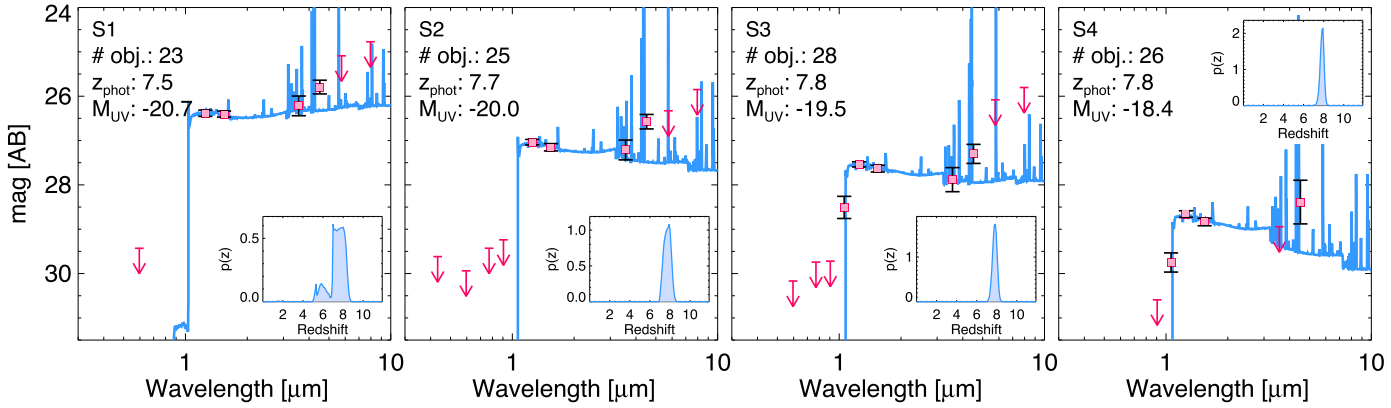

Figure 3. Median-stacked SEDs resulting from our analysis. Each panel refers to a stack at a different luminosity, as indicated by the label in the top-left corner of each panel. In each panel, the filled red squares with error bars mark the stacked photometry, while the red arrows represent 2σ upper limits. The blue curves correspond to the best-fitting EAZY template. The insets present the redshift probability distribution computed by EAZY. The top-left corner of each panel also shows the number of objects entering the specific luminosity bin, the median redshift, and the $M_{\rm UV}$ computed by EAzY. All SEDs show red [3.6]–[4.5] >0 colors; strikingly the SEDs of the three less-luminous stacks show blue $H_{160}-[3.6]<0$ colors, suggesting young stellar populations.

Table 2
Flux Densities for the Four Stacked Luminosity-selected SEDs

Filter	S1 (nJy)	S2 (nJy)	S3 (nJy)	S4 (nJy)
B_{435}		0.4 ± 2.6		
V_{606}	-0.3 ± 3.1	0.0 ± 1.9	-2.1 ± 1.6	-0.3 ± 0.5
i_{775}		-2.3 ± 3.1	-0.8 ± 2.3	-0.2 ± 0.7
z_{850}		-1.0 ± 3.7	-1.7 ± 2.3	0.1 ± 1.1
Y_{105}			14.3 ± 3.3	4.6 ± 0.9
J_{125}	101.4 ± 6.5	55.7 ± 3.1	35.1 ± 1.9	12.5 ± 0.8
H_{160}	99.8 ± 6.3	50.0 ± 3.8	32.1 ± 2.3	10.6 ± 0.8
IRAC 3.6 μm	118.2 ± 24.2	47.3 ± 9.7	25.5 ± 6.4	6.7 ± 4.8
IRAC 4.5 μm	174.7 ± 24.9	85.1 ± 12.8	43.6 ± 8.6	16.0 ± 7.3
IRAC 5.8 μm	207.1 ± 167.3	43.2 ± 53.4	126.2 ± 67.1	3.1 ± 61.2
IRAC 8.0 μm	-8.6 ± 223.9	-14.8 ± 83.0	139.3 ± 86.6	-48.7 ± 68.0

Note. Measurements for the ground-based and Spitzer/IRAC bands are $1\farcs2$ aperture fluxes from MOPHONGO corrected to total using the PSF and luminosity profile information; HST/WFC3-band flux densities are measured in $0\farcs6$ apertures and converted to total using the PSF growth curves. Measurements involving <90% of sources have been left blank. We omit JH_{140} because its coverage is available for <90% of sources for any stack.

Table 3
Main Colors of the Stacked SEDs

Label	$J_{125}-H_{160}$ (mag)	$H_{160}-[3.6]$ (mag)	[3.6]–[4.5] (mag)
S1	-0.02 ± 0.10	0.18 ± 0.23	0.42 ± 0.27
S2	-0.12 ± 0.10	-0.06 ± 0.24	0.64 ± 0.28
S3	-0.09 ± 0.10	-0.25 ± 0.28	0.58 ± 0.35
S4	-0.18 ± 0.11	<-0.11[a]	>0.56[a]
Average	-0.10 ± 0.05	-0.03 ± 0.14	0.55 ± 0.17

Note.
[a] 2σ limit.

and S2 is likely the result of the lack of Y_{105} stacked data for those bins, as discussed in Section 3.

The available filters allow, in principle, for a determination of the UV-continuum slopes β. We computed UV slopes by fitting a power law to the best-fitting SED template from EAZY, in a similar fashion to the procedures of Finkelstein et al. (2012), Duncan et al. (2014), and Bhatawdekar & Conselice (2021). In particular, Duncan et al. (2014) showed that this method is only marginally affected by bias against slopes bluer than the lower β allowed by the templates. The result is that our SEDs have $\beta \lesssim -2$, broadly consistent with other recent determinations at $z \sim 7$–8 (e.g., Bouwens et al. 2010; Finkelstein et al. 2010; McLure et al. 2011; Bouwens et al. 2014), with a qualitative trend to bluer slopes at fainter luminosities. However, the current data make accurate determinations of the UV slopes for the stacked SEDs challenging. Indeed, the Lyman break $\lambda 1215$ Å enters the J_{125} band at $z \sim 8.1$, decreasing the flux observed in this band, and therefore mimicking redder β values. A simple toy model consisting of 1000 SEDs with $f_\lambda \propto \lambda^\beta$ uniformly distributed between $z = 7.6$ and $z = 8.6$ (i.e., for roughly half of the sample, the flux in the J_{125} band is affected by the Lyman break, while for the remaining half, it is not) generates β values systematically redder by $\Delta\beta \sim 0.2$, for $-3.0 < \beta < -1.5$. Furthermore, recent observations have shown that Lyα photons may still escape the IGM even at these high redshifts (see, e.g., Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016). This can potentially increase the flux in the J_{125} band leading to bluer UV slopes. Using the same toy model from above, for EW$_0$ of Lyα EW$_0 \sim 20$–40 Å (e.g., Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Stark et al. 2017), the UV

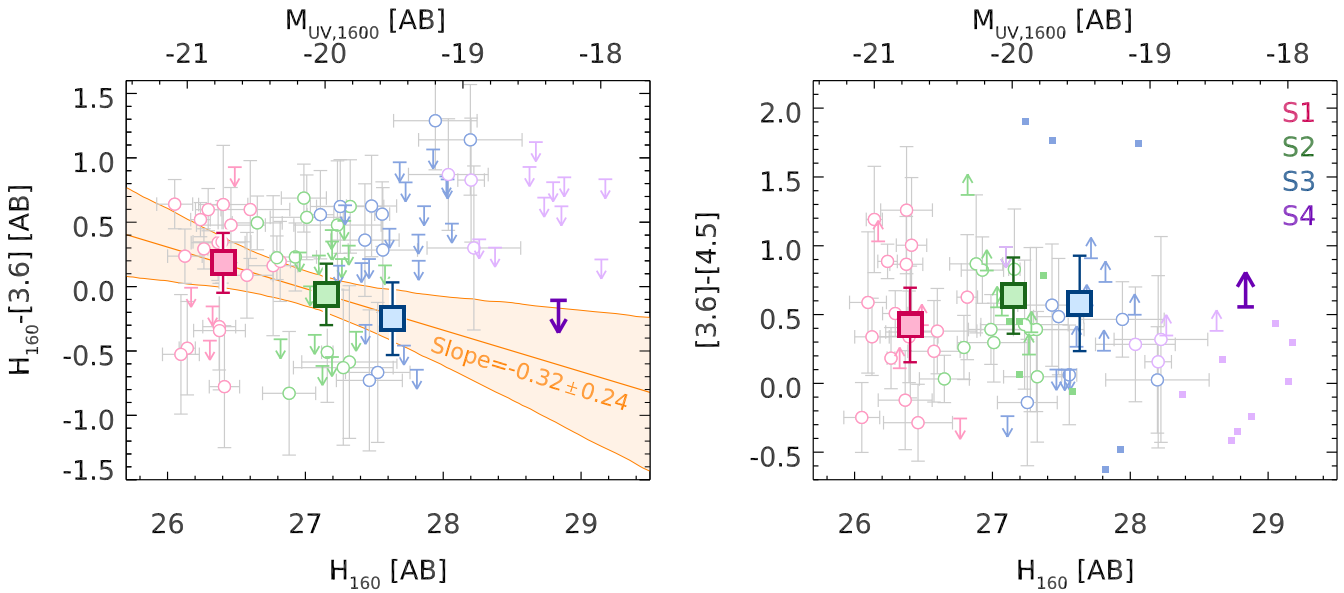

Figure 4. Color–magnitude diagrams for the stacked (colored squares with error bars) and full sample (colored circles with error bars for measurements $>2\sigma$; arrows indicate 2σ upper or lower limits; we do not show the error bars when S/N <2 in each band). Colors refer to specific absolute magnitude bin adopted for the stack: red, green, blue, and purple for $M_{\rm UV} = -20.9$, -20.2, -19.5, and -18.4 mag, respectively (S1, S2, S3, and S4, respectively). The top axis marks UV absolute magnitudes obtained converting the observed H_{160} flux assuming a redshift $z = 8$. Stacks S3 and S4 have blue $H_{160} - [3.6] < 0$ mag at the 1σ–2σ level. When all stacks are considered, we find that the $H_{160} - [3.6]$ colors for lower-luminosity galaxies are either systematically bluer than higher-luminosity galaxies or show similar colors. This is consistent with the idea that lower-luminosity galaxies have younger stellar populations. All stacks instead have [3.6]–[4.5] colors consistent with [3.6]–[4.5] ~ 0.5 mag, as expected because of the strong [O III]+Hβ line emission. All key numbers are given in Table 3.

Table 4
Main Physical Properties of the Stacked SEDs

Label	N. gal.[a]	$z_{\rm phot}$[b]	$M_{\rm UV}$ (mag)	[O III]+Hβ[c] (Å)	Hα[d] (Å)	log(age/yr)	A_V (mag)	$\log(M_\star/M_\odot)$	SFR$_{\rm UV}$[e] $M_\odot$yr^{-1}	SFR$_{\rm [O\,III]+H\beta}$[f] $M_\odot$yr^{-1}	sSFR[g] Gyr^{-1}
S1	23	7.76	-20.7 ± 0.1	768 ± 279	532 ± 193	$8.2^{+0.4}_{-0.3}$	$0.3^{+0.2}_{-0.3}$	$9.2^{+0.2}_{-0.2}$	22.2 ± 1.4	35.6 ± 12.3	$14.3^{+10.5}_{-5.9}$
S2	25	7.69	-20.0 ± 0.1	1369 ± 389	949 ± 270	$7.7^{+0.3}_{-1.0}$	$0.1^{+0.3}_{-0.1}$	$8.4^{+0.2}_{-0.5}$	6.7 ± 0.5	19.3 ± 5.2	$26.5^{+59.4}_{-11.1}$
S3	28	7.75	-19.5 ± 0.1	1184 ± 473	821 ± 328	$7.1^{+0.8}_{-1.1}$	$0.0^{+0.2}_{-0.0}$	$7.8^{+0.6}_{-0.1}$	3.4 ± 0.2	8.7 ± 3.5	$60.3^{+21.9}_{-45.2}$
S4	26	7.70	-18.4 ± 0.1	>1117[h]	>774[h]	$6.8^{+1.4}_{-0.8}$	$0.0^{+0.2}_{-0.0}$	$7.1^{+0.9}_{-0.0}$	1.2 ± 0.1	>3.1[h]	$83.5^{+9.4}_{-73.2}$

Notes. Stellar population parameters (age, A_V, SFR$_{\rm UV}$, and sSFR) were computed adopting a constant SFH, $0.2Z_\odot$ metallicity, and a Salpeter (1955) IMF between 0.1 and 100 $M_\odot$. SFR$_{\rm UV}$ and SFR$_{\rm [O\,III]+H\beta}$ incorporate a correction for dust extinction using the values listed in column A_V, assuming a Calzetti et al. (2000) law and the same dust law for both the stellar continuum and the nebular emission.

[a] Number of galaxies in each bin of absolute magnitude.

[b] Median photometric redshift of the sources in each bin.

[c] Rest-frame equivalent width of [O III] and Hβ obtained assuming a flat f_ν SED redward of the Balmer break and the line ratios in Anders & Fritze-v. Alvensleben (2003). These values would be ~30% lower if we assumed that the [O II] line emission only marginally contributed to the measured 3.6 μm band flux density.

[d] Rest-frame equivalent width of Hα from the Hβ estimate assuming the line ratios in Anders & Fritze-v. Alvensleben (2003) and $H\alpha/H\beta = 2.85$ (Hummer & Storey 1987).

[e] SFR computed from the rest-frame UV continuum.

[f] SFR computed from the Hα luminosity inferred from the [O III]+$H\beta$ line EW and assuming line ratios of Anders & Fritze-v. Alvensleben (2003), with $H\alpha/H\beta = 2.85$.

[g] sSFR based on SFR$_{\rm UV}$.

[h] 2σ lower limit inferred assuming the 2σ upper limit in the flux density of the 3.6 μm band.

slope becomes bluer by $\Delta\beta \sim 0.1$–0.4. Given these systematics, we decided not to highlight the UV slopes of our stacked SEDs even though we will occasionally present some results in the context of their $J_{125} - H_{160}$ colors.

4.2. The [3.6]–[4.5] Color

One of the most notable features evidenced by Figures 3 and 4 is a red [3.6]–[4.5] ~ 0.5 mag color, which, at these redshifts, has been interpreted as the effect of nebular $H\beta$ + [O III] emission significantly contributing to the flux density in the 4.5 μm band (e.g., Schaerer & de Barros 2009; Labbé et al. 2013). Figure 4 shows that this color is approximately unchanged across 2.5 magnitudes in UV luminosity, with the hint of a bluer color for the most luminous systems. Future higher-S/N observations may well reveal changes in line ratios with luminosity.

Under the hypothesis that the measured red IRAC color is mostly the result of [O III] and Hβ nebular emission contributing to the flux in the 4.5 μm band, we can estimate the rest-frame EW$_0$ of [O III]+$H\beta$. For this, we assumed the line ratios of Anders & Fritze-v. Alvensleben (2003) for subsolar metallicity ($Z = 0.2Z_\odot$), and adopted a flat f_ν,

calibrating the f_ν continuum flux to that of the 3.6 μm band. At $z>7$, $[\mathrm{O\,II}]_{3727}$ enters the 3.6 μm band. The likely strong emission by rest-frame optical lines implied by the red IRAC colors suggests that the contribution of the [O II] line emission to the 3.6 μm band flux density might be non-marginal. Adopting the line ratios of Anders & Fritze-v. Alvensleben (2003), we iteratively removed the contribution of the [O II] emission from the 3.6 μm band flux density. We found EW([O II]) $\sim$ 180–370 Å, corresponding to $\sim$0.1–0.3 mag contribution to the 3.6 μm band flux density.

The resulting [O III]+$H\beta$ line strengths, listed in Table 4, are consistent with EW_0 ([O III] + $H\beta$) $\sim$ 1000Å found at these redshifts for galaxies of comparable luminosity (Labbé et al. 2013; Smit et al. 2014, 2015; De Barros et al. 2019; Endsley et al. 2021). The largest EW_0 ([O III] + $H\beta$) $\sim$ 1200–1300Å are found for S2 and S3 ($L_{\mathrm{UV}} \sim 0.2$–$0.4L^*$) and are consistent with the measurements for the most extreme sources reported by Endsley et al. (2021) for $L_{\mathrm{UV}} \sim L^*$ systems at $z \sim 7$.

Although our measurements suffer from large uncertainties, the values corresponding to S1 and S2 are consistent with the extrapolation of estimates from the $M_\star$-EW_0([O III]+$H\beta$) found at $z \sim 3.3$ by Reddy et al. (2018), in turn supporting the reduced rate of evolution with redshift of the EW_0([O III]+$H\beta$) observed at $z \gtrsim 3$ (see, e.g., Khostovan et al. 2016 and their Figure 7).

We estimate that the [O III]+$H\beta$ EW would be $\sim$30% smaller if the contribution of [O II] to the 3.6 μm flux density is negligible. Although our data can not rule out this possibility, the above discussion would not change in its main conclusions, and our results would still point to exceptionally strong emission from [O III]+$H\beta$.

4.3. The H_{160} − [3.6] Color

At redshift $z \sim 8$, the 3.6 μm band covers the rest-frame wavelengths just redward of the Balmer/4000 Å break, and the H_{160} −[3.6] color can then provide a measurement of the strength of the break and hence an estimate of the age of the stellar populations.

The H_{160} − [3.6] colors are presented in the left panel of Figure 4, as a function of the H_{160}-band magnitude (where H_{160} can be taken as a proxy for M_{UV} at $z \sim 8$), for the individual objects in the sample and from stacking. Our stacking results show that while H_{160} −[3.6] $\sim$ +0.2 mag for the most luminous stack, it becomes negative, i.e., H_{160} −[3.6] $\lesssim$ 0 mag for S2, S3, and S4, indicative of very young stellar populations (more discussion about this is presented in Section 5.2). Furthermore, our stacking results suggest a trend of redder H −[3.6] with increasing luminosity, consistent with the picture of more luminous (i.e., likely more massive) systems having more evolved stellar populations. Fitting the H_{160} −[3.6] stack results versus H_{160} magnitudes to a line, we find:

$$H_{160} - [3.6] = -0.01 \pm 0.14 - (0.32 \pm 0.24)(H_{160} - 27) \quad (2)$$

where the adopted $H_{160} = 27$ mag pivot point approximately corresponds to the median of the H_{160} flux density for the entire sample.

The trend is significant at $\sim$80% confidence, and it is qualitatively consistent with results at lower redshifts (e.g., González et al. 2012), where more luminous galaxies show larger Balmer breaks than lower-luminosity sources.

A number of studies have shown that the UV slope (β) correlates with the UV luminosity of LBGs at $z>4$ (e.g., Bouwens et al. 2012, 2014; Finkelstein et al. 2012; Duncan et al. 2014; Bhatawdekar & Conselice 2021). In Figure 5 we explore in more detail the dependence of the H −[3.6] color on β, for the β stacks and for individual sources. As noted earlier, the β-binned stacks were generated similarly to the luminosity-binned stacks (the $\beta \sim -1.7$ bin is not shown here).

The distribution of measurements in Figure 5 does not show any significant accumulation at the blue end, suggesting that our β estimates have same validity. Our individual measurements indicate a prevalence of $\beta < -2$, qualitatively consistent with the blue values of the rest-frame UV found for S2, S3, and S4.

The stacked H_{160} −[3.6] colors present a clear trend, becoming bluer for sources with bluer UV slopes. A simple fit results in the following relationship:

$$H - [3.6] = (0.23 \pm 0.08) + (1.78 \pm 0.36)(\beta + 2.2) \quad (3)$$

valid for $-2.6 \leqslant \beta \leqslant -1.9$. In the same figure, we also present the linear relation found by Oesch et al. (2013) between β and the J_{125} −[4.5] color, from a sample of $z \sim 4$ LBGs in the GOODS fields. We converted the J_{125} −[4.5] color into an H_{160} −[3.6] color assuming a flat f_ν continuum redward of the Balmer break, and a power law with slope β for the rest-frame UV. Our linear relation (Equation (3)) is a little steeper than that of Oesch et al. (2013), indicative of potential evolution toward redder H_{160} −[3.6] colors between $z \sim 8$ and $z \sim 4$, qualitatively consistent with an aging of the stellar populations with cosmic time. However, a closer inspection suggests that the steeper slope we observe at $z \sim 8$ is also likely driven by the very blue H_{160} −[3.6] $\sim$ −0.5 mag for $\beta \sim -2.5$, as the measurements for $\beta \geqslant -2.3$ are consistent at $\gtrsim 1\sigma$ with the $z \sim 4$ relation.

Accounting for a $\sim$0.2 mag contribution from [O II] line emission to the 3.6 μm band, as we derived in Section 4.2, would bring our stacked measurements for the H −[3.6] color as a function of β into even better agreement with those of Oesch et al. (2013) for $\beta > -2.3$, if we assume that the $z \sim 4$ relationship is only marginally affected by nebular line emission. Nonetheless, at $z \sim 4$, Faisst et al. (2016) found EW($H\alpha$) $\sim$ 300 Å, resulting in $\lesssim$0.1 mag contribution from [O II], supporting our initial assumptions.

5. Discussion

5.1. Stellar Population Parameters

We estimated stellar masses ($M_\star$), stellar population ages, and dust extinction (A_V) running FAST (Kriek et al. 2009) on SED templates based on the Bruzual & Charlot (2003) set, with a Salpeter (1955) IMF between 0.1 and 100 $M_\odot$, constant SFH, metallicity $Z = 0.2Z_\odot$, and a Calzetti et al. (2000) dust-extinction law. Prior to our use of these templates for the SED fitting, the templates were processed through CLOUDY version 17.02 (Ferland et al. 2017), assuming a spherical constant-density nebula with $n(H) = 100\,\mathrm{cm}^{-3}$, the same gas metallicity as the stellar component, an ionization parameter $\log U = -2.5$, consistent with recent results at similar redshifts (e.g., Stark et al. 2017; De Barros et al. 2019), and a negligible escape fraction.

To reduce a potential bias toward extremely young ages, the fitting for S1, S2, and S3 was performed on the photometry cleaned of the emission line contribution, using the information on the EW we obtained in Section 4.2. For consistency,

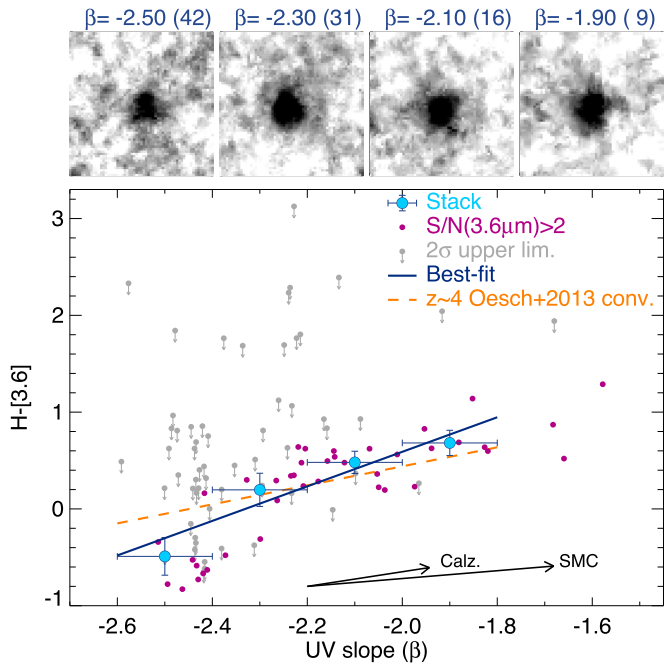

Figure 5. Top panels: image stamps (∼7.″2 per side) of the 3.6 μm stacks in the four bins of UV slope as indicated by the label at the top of each stamp. The number of sources in each bin is shown in parentheses. Bottom panel: the small filled circles and arrows correspond to the individual measurements of the $H-[3.6]$ color and 2σ upper limits, respectively, as a function of the rest-frame UV slope β for the sample adopted for this work. The filled blue circles mark the color measured from the stacking of sources in bins of UV slope as indicated by the horizontal error bars. The best-fitting linear relation for the stacked β values is shown by the solid blue line, while the dashed orange line represents the correlation between $J_{125}-[4.5]$ and β found by Oesch et al. (2013) at $z\sim4$ and converted to $H_{160}-[3.6]$ assuming a flat f_ν at rest-optical. The vectors at the bottom-right corner indicate the effect of a $\Delta A_V=0.1$ mag dust extinction for an SMC law (Pei 1992) and a $\Delta A_V=0.2$ mag for a Calzetti et al. (2000) model.

emission lines were also removed from the SED templates. Physical parameters for S4 were determined by adopting the pristine photometry with templates including emission lines, given the 1.4σ detection in the 3.6 μm band.

We computed SFRs following the relation between rest-frame UV luminosity and the SFR of Madau & Dickinson (2014), which is based on a constant SFH, consistent with the SFH we set in FAST. Moreover, to maintain consistency with the configuration used in estimating $M_\star$, we adopted parameters corresponding to a $Z=0.2Z_\odot$ metallicity (after log-linear interpolating the values tabulated in Madau & Dickinson 2014).

Using the results on the $\mathrm{EW_0}$ of [O III]+$H\beta$ we derived in Section 4.2, we also estimated the SFRs from the nebular emission, which we denote with $\mathrm{SFR_{[O\,III]+H\beta}}$. We converted the measured equivalent width of [O III]+$H\beta$ into $H\alpha$ line intensity assuming Case B recombination, the line ratios of Anders & Fritze-v. Alvensleben (2003) for $Z=0.2Z_\odot$ (specifically $[\mathrm{O\,III}]\lambda4959/\mathrm{H}\beta=1.6$ and $[\mathrm{O\,III}]\lambda5007/\mathrm{H}\beta=4.7$), and $\mathrm{H}\alpha/\mathrm{H}\beta=2.85$ (Hummer & Storey 1987). The $\mathrm{SFR_{[O\,III]+H\beta}}$ were computed from the $H\alpha$ luminosities using the Kennicutt & Evans (2012) relation.

Finally, all SFR estimates were corrected for dust extinction using the A_V values previously estimated from FAST and assuming a Calzetti et al. (2000) dust-extinction law for both the stellar continuum and the nebular emission, and converted from a Kroupa (2001) to a Salpeter (1955) IMF by multiplying the results by a factor of 1.7.

The estimated stellar population parameters for the four stacks are listed in Table 4.

5.2. Age and $M_\star/L_{UV}$ Ratio of the Stacked SEDs

The SED fitting results show that stacks have mean stellar masses per source in the range $M_\star\sim10^{7.1}\,M_\odot$ to $10^{9.2}\,M_\odot$, implying mass-to-light ratios $M_\star/L_{\mathrm{UV}}\sim0.003M_\odot/L_\odot$ to $\sim0.04M_\odot/L_\odot$, consistent with values found at similar redshifts for $L\sim1.4$–$3.6L^*$ sources (e.g., Stefanon et al. 2019). Interestingly, the recovered ages range between ∼8 and ∼300 Myr.

To provide us a better sense of the uncertainties on the estimated ages, in Figure 6 we compare the observed color [3.6]–[4.5] versus $H_{160}-[3.6]$ to the synthetic colors extracted from the library of SED templates adopted to run FAST (see Section 5.1).

The blue $H_{160}-[3.6]\lesssim0$ mag of S2, S3, and S4 indicate ages younger than ∼100 Myr, while the redder color for S1 suggests ages ∼100–300 Myr. These young ages are also supported by the extreme red [3.6]–[4.5] ∼ 0.5 mag colors, indicative of strong line intensities.

It is well established (e.g., Renzini & Buzzoni 1986) that the broadband SED of a young, metal-rich system can mimic an older, metal-poor one, and vice versa. Specifically, in our case, metallicities lower than $Z=0.2Z_\odot$ could imply stellar ages older than ∼100 Myr. To explore this scenario, in Figure 6 we also present synthetic tracks for $Z=0.02Z_\odot$ and $\log U=-2.5$. The lower metal abundances result in decreased intensity of the optical emission lines. The corresponding bluer [3.6]–[4.5] colors are inconsistent with our measurements. Addition of $\gtrsim$0.4 mag dust extinction would bring the track closer to our measurements, but with ages $\lesssim$100 Myr. Finally, because the ionization parameter can affect the intensity of emission lines, we also consider the effects of an exceptionally intense $\log U=-1.0$ ionizing field. The corresponding track is consistent with our estimates for very young ages (less than a few million years), while the addition of dust extinction would result in ages younger than ∼100 Myr.

The observed trend toward redder $H_{160}-[3.6]$ colors at higher luminosities suggests that we could be witnessing the development of a more evolved stellar population at just ∼650 Myr of cosmic time, indicative that the ages of the stellar populations could depend on mass already at such early epochs. The degree of such evolution is hard to establish though, since bright young stars could be outshining any underlying older stellar populations (e.g., Papovich et al. 2001). Recent work suggests the existence of evolved stellar populations in $L\lesssim1$–$0.4L^*$ individual galaxies identified at $z\sim8$–10 (e.g., Hashimoto et al. 2018; Hoag et al. 2018; Salmon et al. 2018; Roberts-Borsani et al. 2020; Strait et al. 2020). Indeed, our data also show $H_{160}-[3.6]>0$ mag for some of the sources, supporting the existence of a more evolved stellar population even for $L<L^*$. Our stacking results, however, indicate that these more evolved populations do not constitute the bulk of the galaxy population at these redshifts, likely implying large duty cycles of star formation. Better quantifying the duty cycle would require the direct detection of a higher fraction of sources in the 3.6 μm band. For this, we await JWST.

5.3. Star Formation Rate

Application of the procedures outlined in Section 5.1 results in $\mathrm{SFR_{UV}}\sim1$–$22\,M_\odot/\mathrm{yr}$ and $\mathrm{SFR_{[O\,III]+H\beta}}\sim3$–$36\,M_\odot/\mathrm{yr}$. The two estimates are compared for the stacks individually in Figure 7.

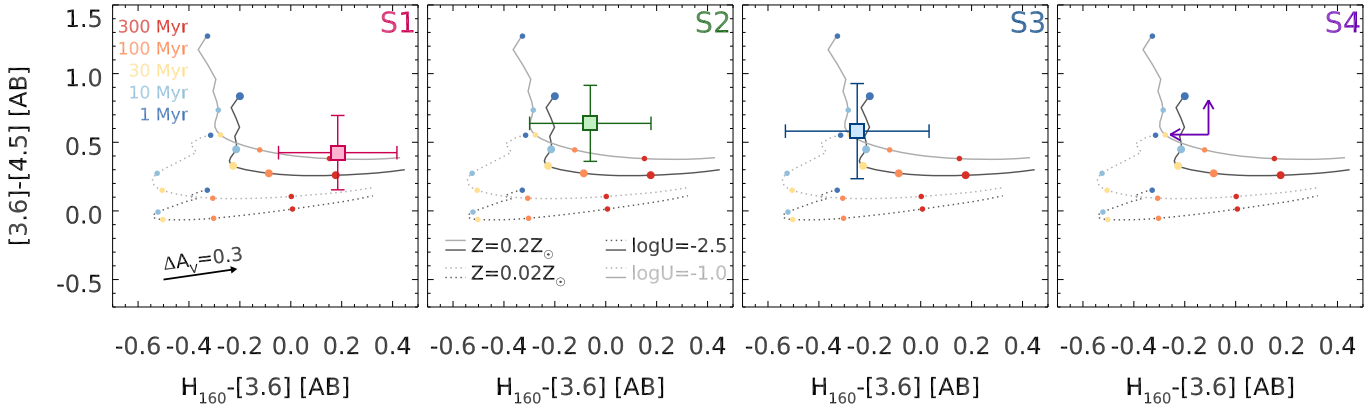

Figure 6. Comparison of the stacked results to models using the [3.6]–[4.5] vs. $H_{160}-[3.6]$ plane. Each panel refers to one of the stacks, as indicated by the labels at the top-right corners. In each panel we also present tracks from the synthetic SEDs with a constant SFH: solid curves mark tracks for a $Z = 0.2Z_{\odot}$ metallicity (adopted as reference value in our study), while the dotted curves explore the colors for $Z = 0.02Z_{\odot}$. We also distinguish between two different ionization parameters, with the black curves marking $\log U = -2.5$, while the gray curves correspond to $\log U = -1$. The colored filled circles mark ages on approximately a logarithmic scale, as indicated by the legend at the top-left corner in the leftmost panel, while the black arrow shows the effect of increasing the extinction by 0.3 mag. Other plotting conventions are the same as in Figure 4. Ages younger than ∼100 Myr are preferred for all but the most luminous stack. Older ages from lower metallicities are ruled out by the implied bluer [3.6]–[4.5] colors resulting from lower EW [O III]$_{4959,5007}$ emission.

While for S1 the two measurements are consistent at $\gtrsim 1\sigma$, the $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta}$ for S2 and S3 are ∼3–5× higher than the $\mathrm{SFR}_{\mathrm{UV}}$. A similar result ($\mathrm{SFR}_{H\alpha}/\mathrm{SFR}_{\mathrm{UV}} \sim 6$) was found by Shim et al. (2011) from analyzing the observed blue IRAC [3.6]–[4.5] color for a spectroscopically confirmed sample of LBGs at $3.8 < z < 5.0$ over the GOODS fields (see also Shivaei et al. 2015; Smit et al. 2015). These studies analyzed different mechanisms to explain such discrepancy, such as dust attenuation, top-heavy IMFs, or increasing ionization parameter, even though no dominant process has been identified so far. In the following paragraphs, we discuss three potential factors that could explain or at least in part contribute to the systematic difference in the SFR estimates from the UV continuum and $H\alpha$ emission: SFH, dust extinction, and line ratios evolving with time. However, the relative contribution of each of these three scenarios is difficult to ascertain, and roughly comparable contributions may also constitute a plausible scenario.

Based on our results in the previous section, the stacked SEDs are likely characterized by very young ages ($\lesssim$30–100 Myr). The UV stellar continuum is mostly generated by O and B stars with typical lifetimes of ∼100–200 Myr (Kennicutt & Evans 2012). A constant SFR equal to the estimated $\mathrm{SFR}_{\mathrm{UV}}$ implies that those stellar masses could be generated in just ∼15–50 Myr, i.e., we might still be witnessing the first burst of star formation, started at $z \sim 9$. However, because the $H\alpha$ emission will be the result of more recent star formation ($\lesssim$10 Myr; e.g., Kennicutt & Evans 2012), the higher $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta}$ values we inferred may indicate an SFR increasing with time. This arises because the stacking analysis, by construction, lessens the variation between individual measurements.

Systematic differences in SFR measurements could be introduced by the differential nature of dust extinction with wavelength. Assuming a Calzetti et al. (2000) extinction curve and assuming that the extinction curve of the nebular emission is the same for the stellar continuum ($E(B-V)_{\mathrm{neb}} = E(B-V)_{\mathrm{cont}}$; e.g., Erb et al. 2006; Reddy et al. 2012; but see, e.g., Calzetti et al. 2000; Steidel et al. 2014; Reddy et al. 2020 for different results), we find that an extinction of $A_V \sim 0.4$–0.5 mag is required in order to match the two SFR estimates. This extinction would result in UV slopes redder by $\Delta\beta \sim 0.7$, or $\beta \sim -2$ (when assuming the bluest intrinsic slope value to be $\beta = -2.6$, as we measure for the youngest template), qualitatively only marginally consistent with the upper limit of the UV slopes of our stacks.

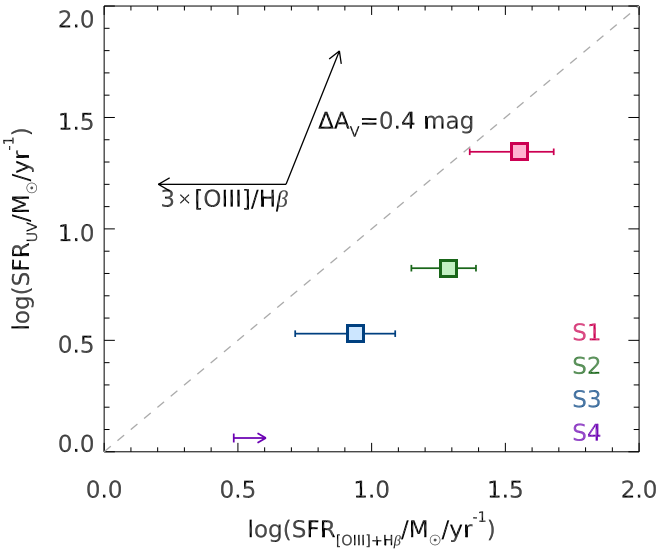

Figure 7. Comparison between $\mathrm{SFR}_{\mathrm{UV}}$ and $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta}$. The former is computed on the basis of the UV luminosities of galaxies, and the latter is computed converting the [O III]+Hβ luminosity measured from the IRAC photometry into $H\alpha$ luminosity assuming the line ratios of Anders & Fritze-v. Alvensleben (2003) and $H\alpha/H\beta = 2.85$. The $H\alpha$ luminosity is finally converted into SFR using the relation of Kennicutt & Evans (2012). The colored points with error bars or lower limits refer to the four stacks as indicated by the legend. The gray dashed line marks the identity, $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta} = \mathrm{SFR}_{\mathrm{UV}}$. The vectors at the top-left corner show the impact of increasing the [O III]/Hβ ratio by three times, and correcting for a Calzetti et al. (2000) dust extinction of $\Delta A_V = 0.4$ mag.

A third possibility for explaining the systematic offset between the $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta}$ and $\mathrm{SFR}_{\mathrm{UV}}$ is if the nebular line ratios are evolving with cosmic time. Indeed, an increase in the ionization parameter (e.g., harder spectrum) has been suggested by recent work (e.g., Maseda et al. 2020). This can result in an increase of the [O III]/Hβ ratio (see, e.g., discussions in Brinchmann et al. 2008 and Steidel et al. 2014).

The [O III]λ5007/H$\beta = 4.7$ adopted in our estimates (Anders & Fritze-v. Alvensleben 2003) is consistent with recent determinations

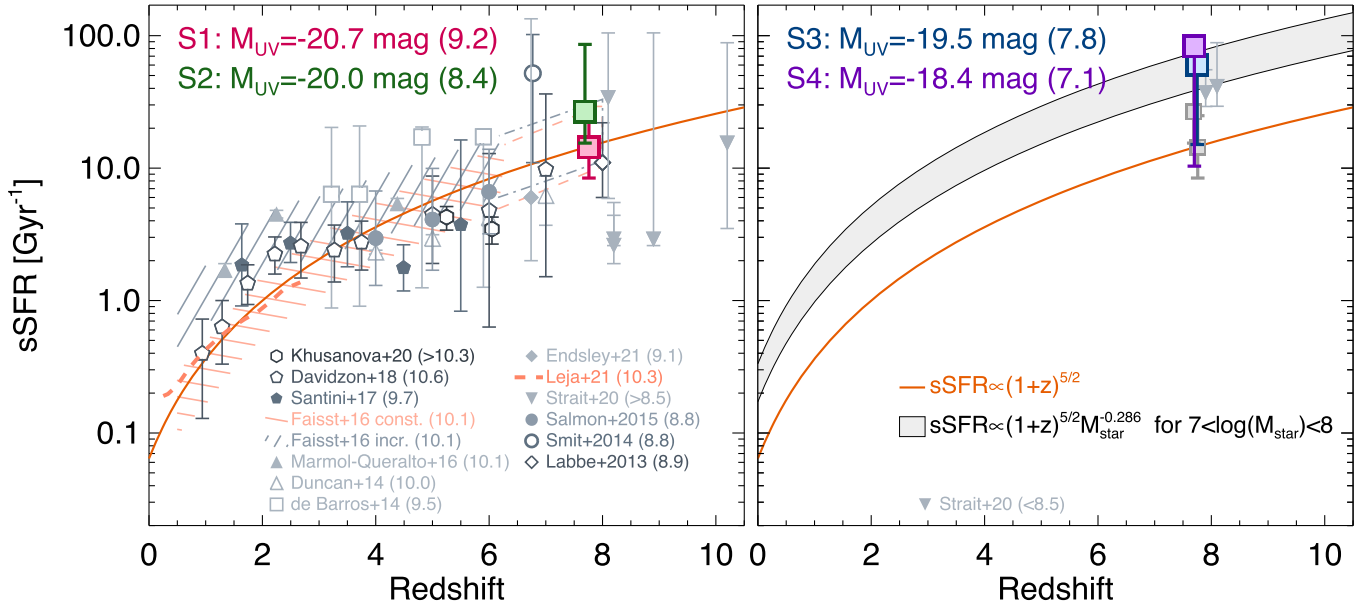

Figure 8. Evolution of the sSFR with redshift. The measurements from the stacking analysis are presented in the panels according to their $M_{\rm UV}$, indicated at the top (together with our "S" label). For reference, in the right panel we also reproduce our measurements for the two more luminous bins (gray squares). In each panel we also plot recent estimates of the sSFR at high redshifts from the literature, as indicated by the legends at the bottom, segregated according to their stellar mass, quoted in parentheses in log scale. The solid orange curve corresponds to the evolution of the sSFR from the toy model of Dekel et al. (2013), of the form sSFR $\propto (1+z)^{5/2}$, as expected from cold gas inflow that follows the hierarchical merging of the dark matter halos. The filled gray region marks the Dekel et al. (2013) model, renormalized to $M_\star = 10^7$–$10^8\,M_\odot$ using the Reddy et al. (2018) stellar mass scaling for $H\alpha$ and assuming the Dekel et al. (2013) relationship to be representative of $M_\star = 10^{9.5}\,M_\odot$ galaxies. The overall good match of the curves to the observations suggests there is no strong evolution in the star formation efficiency of galaxies with cosmic time.

from spectroscopic measurements of LBGs at $z\sim2$–4 (e.g., Dickey et al. 2016; Holden et al. 2016). Forcing the relation $\mathrm{SFR}_{[\mathrm{O\,III}]+H\beta} = \mathrm{SFR}_{\rm UV}$, and keeping $[\mathrm{O\,III}]\lambda5007/[\mathrm{O\,II}] = 3$, the red [3.6]–[4.5] color implies $[\mathrm{O\,III}]\lambda5007/\mathrm{H}\beta = 17\pm8$ and 13 ± 8, for S2 and S3, respectively. This is approximately three times the ratio we initially adopted (see Figure 7).

Even though estimates of the $[\mathrm{O\,III}]\lambda5007/\mathrm{H}\beta$ ratio are still very uncertain at high redshifts, a number of studies suggest the existence of a trend of increasing $[\mathrm{O\,III}]/\mathrm{H}\beta$ with redshift (e.g., Brinchmann et al. 2008; Steidel et al. 2014; Kewley et al. 2015; Faisst et al. 2016; Harikane et al. 2018b). In particular, Faisst et al. (2016) found that at $z\gtrsim5$, the $[\mathrm{O\,III}]/\mathrm{H}\beta$ ratio could be larger, $[\mathrm{O\,III}]/\mathrm{H}\,\beta\gtrsim3$, with values potentially as high as $[\mathrm{O\,III}]/\mathrm{H}\,\beta\sim15$ and with a lower limit of $[\mathrm{O\,III}]/\mathrm{H}\,\beta\sim2$. On the other hand, using a spectroscopically confirmed sample of Lyα emitters at $z\sim5.7$, 6.6, and 7.0, Harikane et al. (2018b) found that the ratio $[\mathrm{O\,III}]/\mathrm{H}\beta$ depends on the Lyα EW, but it is roughly independent from redshift (see their Figure 15). Specifically, at $z\sim5.7$, which provided the widest coverage in terms of $\mathrm{EW}_{\rm Ly\alpha}$, they measured a maximum value of $[\mathrm{O\,III}]/\mathrm{H}\,\beta\sim10$, for $\mathrm{EW}_{\rm Ly\alpha}\sim90$ Å, consistent with our estimates for S2 and S3. However, their measurement at $z\sim7$ results in an upper limit ($[\mathrm{O\,III}]/\mathrm{H}\,\beta<2.8$) somewhat in tension with our results.

5.4. Specific Star Formation Rate and Halo Mass Growth

Combining our results from the previous sections, we can now estimate the sSFR at $z\sim8$. The resulting values are listed in Table 4 and shown in Figure 8, segregated according to the UV luminosity (which can roughly be taken as a stellar mass proxy). These values are consistent with the young ages we previously recovered, as for a constant SFH, the sSFR$\sim1/$ age $\sim$6–160 Gyr^{-1}. It is perhaps noteworthy that the sSFRs we derive at different stellar masses are consistent with each other. While recent observations indicate that the sSFR marginally depends on stellar mass to $z\sim6$ for $\log(M_\star/M_\odot)\lesssim10$ (e.g., Fumagalli et al. 2012; Whitaker et al. 2014; Salmon et al. 2015; Schreiber et al. 2015; Reddy et al. 2018), current models still provide contrasting predictions for a mass-dependent sSFR (see, e.g., Behroozi et al. 2013) or for a mass-independent sSFR (see, e.g., Sparre et al. 2015; Ceverino et al. 2018). The agreement between measurements at different stellar masses could be interpreted as the existence of a main sequence of star-forming galaxies across 1.5 orders of magnitude already at $\sim$650 Myr of cosmic time.

In Figure 8 we also compare our sSFR estimates to recent determinations at high redshifts from Labbé et al. (2013), Smit et al. (2014), Duncan et al. (2014), Salmon et al. (2015), Mármol-Queraltó et al. (2016), Santini et al. (2017), Davidzon et al. (2018), Khusanova et al. (2021), Strait et al. (2020), and Endsley et al. (2021; median of the sample and 68% range), together with the measurements at $0\lesssim z\lesssim6$ of Faisst et al. (2016), and at $0.2\lesssim z\lesssim2.7$ of Leja et al. (2021), segregated by stellar mass and/or $M_{\rm UV}$. Most of theses measurements correspond to stellar masses $>10^9\,M_\odot$ and show large systematic differences for $z>4$, spreading over $\sim$1 order of magnitude, with sSFR ranging from ~3 Gyr^{-1} (see also Endsley et al. 2021) to ~20 Gyr^{-1}.

Our estimates are consistent with those of $z\sim7$–8 LBGs from broadband photometry (Labbé et al. 2013; Smit et al. 2014; Jiang et al. 2016; Castellano et al. 2017) and with the extrapolation that Faisst et al. (2016) propose to $z\sim8$. They indicate an increasing sSFR with redshift out to $z\sim8$, irrespective of the stellar mass bin.

The solid black curve in Figure 8 marks the evolution of the sSFR from the toy model of Dekel et al. (2013). In this model, the specific dark matter halo accretion rate (SHMAR) evolves as $\propto(1+z)^{5/2}$. It is interesting to note that the exponent 5/2

was obtained analytically from simple theoretical considerations on the Extended Press-Schechter formalism describing the evolution of the dark matter halo assembly (see also Neistein & Dekel 2008; Weinmann et al. 2011; Genel et al. 2014). The SHMAR was then converted into sSFR assuming that the baryonic accretion involved exclusively cold gas. The resulting form is simply a re-normalization of the SHMAR, which maintains the dependence of $\propto (1+z)^{5/2}$. Similarly, an sSFR increasing with redshifts out to $z\sim 8$ was also predicted by Tacchella et al. (2013) assuming the SFH closely follows the halo assembly time, by Ceverino et al. (2018), with which our measurements are consistent to within a factor of two, and by Park et al. (2019), who assumed a non-evolving star formation efficiency.

The similarity between the evolution of the sSFR and of the SHMAR at $z\gtrsim 3$ suggests that at high redshifts, galaxy formation could be dominated by the assembly of cold gas, driven by the hierarchical formation of the dark matter halos, with marginal dependence on an evolving star formation efficiency. This result is qualitatively consistent with a marginal evolution with redshift of the $M_\star/M_h$ ratio recently observed at $z>4$ (Stefanon et al. 2017, 2021b) and with such models providing a natural explanation for the evolution in the UV LF at $z\gtrsim 4$ (Tacchella et al. 2013, 2018; Bouwens et al. 2015, 2021; Mason et al. 2015; Harikane et al. 2018a; Park et al. 2019).

Because feedback mechanisms could influence the star formation differently depending on the stellar and halo mass, in the right panel of Figure 8, we also present (gray filled region) a speculative curve for the sSFR with a dependence on stellar mass. Specifically, we assumed that the Dekel et al. (2013) relationship well represents galaxies with a stellar mass of $M_\star = 10^{9.5} M_\odot$. This was then rescaled for $M_\star = 10^7$–$10^8\,M_\odot$ with the $M_\star^{-0.286}$ dependence of $H\alpha$ luminosity of Reddy et al. (2018; but see, e.g., Ceverino et al. 2018, who found a marginal dependence of the sSFR on stellar mass).

6. Summary and Conclusions

Through the analysis of the deepest available Spitzer IRAC and Hubble data for a significant sample of star-forming galaxies at $z\sim 8$, we have been able to gain striking new insights into the SFH of galaxies just 650 Myr after the Big Bang. These insights are derived on the basis of deep stacked SEDs of $z\sim 8$ LBGs identified over the CANDELS/GOODS-N/S, ERS, XDF, CANDELS/UDS, and CANDELS/COSMOS fields. These fields are characterized by deep coverage with HST in the HST/ACS V_{606} and I_{814}, HST/WFC3 Y_{105}, J_{125}, JH_{140}, and H_{160} bands, and by Spitzer/IRAC in the 3.6 μm and 4.5 μm bands. In particular, the GOODS fields benefit from ultradeep IRAC 3.6 μm and 4.5 μm mosaics, combining all of the observations from the recently completed GREATS (Stefanon et al. 2021a). This program brings near-homogeneous ∼200–250 hr depth (corresponding to $5\sigma\sim 26.8$–27.1 mag) in Spitzer/IRAC 3.6 and 4.5 μm in 200 arcmin^2 over the two GOODS fields.

The full sample was segregated into four UV luminosity bins (S1, S2, S3, and S4, in order of decreasing luminosity). Stacking in the HST bands was performed directly on the flux measurements, while for the IRAC bands, we carried out aperture photometry on the stacked image cutouts, after each source was cleaned of neighboring sources. We excluded those cutouts with neighboring residuals that remained from the MOPHONGO image processing that overlapped with the nominal location of a source. To account for the overall shape of each SED, stacking was performed after normalizing the fluxes and cutouts of each source to the weighted mean of the measurements in the J_{125} and H_{160} bands.

Our main observational results can be summarized as follows:

1. The stacked SEDs are characterized by $H_{160}-$ [3.6] colors ranging between ~ -0.3 mag and $\sim +0.2$ mag. The blue $H_{160}-[3.6]\lesssim 0$ mag colors we measure for the three faintest stacks imply very young stellar ages ($\lesssim 10^8$ yr).
2. We explored the dependence of the $H_{160}-[3.6]$ color on the rest-frame UV luminosity. We find that our measurements are consistent with either no trend or for fainter galaxies to have younger ages (Figure 4). This suggests that the stellar population ages of galaxies may be mass-dependent ∼650 Myr after the Big Bang.
3. The stacks are characterized by red [3.6]–[4.5] $\gtrsim 0.5$ mag colors. Given the plausible assumption that the red [3.6]–[4.5] color results from the [O III] and $H\beta$ emission lines entering the 4.5 μm band at $z\gtrsim 7.5$, these colors correspond to EW$_0$([O III]+$H\beta$) ∼ 800–1300Å (Table 4).

The above results lead to the following conclusions:

1. The SFRs we infer from the measured [O III]+$H\beta$ EWs are factors ∼3 higher than the SFR derived from the UV luminosity. Possible explanations include either an increasing SFR with time or an [O IIIλ5007]/$H\beta$ ratio evolving with time as a result of an evolving ionizing radiation field (Figure 7). A dust extinction of $A_V\sim 0.4$ mag is less effective as an explanation, given that it would imply UV slopes β redder by $\Delta\beta\sim 0.7$, only marginally consistent with our current and other earlier observations. An older stellar population also is not effective as an explanation, given the clear discrepancy with the measured blue $H_{160}-[3.6]\lesssim 0$ mag colors.
2. The high SFRs inferred from the UV light imply sSFRs of $\gtrsim 10$ Gyr^{-1}. When these new sSFR values at $z\sim 8$ are compared to the sSFRs at lower redshifts, they suggest a very similar evolution in the sSFRs of galaxies and the specific halo mass accretion rate, and, ultimately, a marginally evolving $M_\star/M_{\rm h}$ ratio for $z\gtrsim 3$. However, the sSFR for the two fainter luminosity bins are also consistent with a scenario where the sSFR weakly depends on stellar mass (Figure 8).

The results presented in this paper are based on broadband photometry probing rest-frame wavelengths up to ∼0.6 μm. More accurate estimates of the stellar continuum redward of the Balmer break will benefit by probing the continuum SED of galaxies at >6 μm. Upcoming facilities like JWST will be pivotal in placing these results on a more robust observational footing. Specifically, the increased sensitivity at redder wavelengths will allow us to probe the contribution of lower-mass evolved stars that dominate the stellar mass. Observations with higher spectral resolution will, on the other hand, allow us to further assess the ionization state of these systems and their SFHs.

The authors would like to thank the referee for the careful reading and for the positive and constructive comments that helped improve the quality of the paper. M.S. and R.J.B. acknowledge support from TOP grant TOP1.16.057. P.A.O. acknowledges support from the Swiss National Science Foundation through the SNSF Professorship grant 190079 "Galaxy

Build-up at Cosmic Dawn." The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. We also acknowledge the support of NASA grants HSTAR-13252, HST-GO-13872, HST-GO-13792, and NWO grant 600.065.140.11N211 (vrij competitie). G.D.I. acknowledges support for GREATS under RSA No. 1525754. This paper utilizes observations obtained with the NASA/ESA Hubble Space Telescope, retrieved from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI). STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. This work is based (in part) on observations made with the Spitzer Space Telescope, which was operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

Appendix A
Candidate Passive Galaxies

One significant aim of this paper was to characterize the stellar population ages of galaxies at $z \sim 8$. While a significant focus of this effort is on the mean/median age, and our overall results suggest a very young age in the typical source, it is also interesting to ask whether any of the $z \sim 8$ sources in our selection show evidence for a significantly older age. The purpose of this Appendix is to investigate this issue.

In Figure 9 we show the colors of a passively evolving galaxy formed in a burst at $z = 20$ (e.g., Mawatari et al. 2020) over both the redshift range of interest $z = 7.3$–8.7 (solid line) and $z = 1.1$–10.0. The passively evolving template is computed from the Bruzual & Charlot (2003) code, assuming a $Z = 0.2Z_{\odot}$ metallicity, a Salpeter (1955) IMF, and that it is processed through CLOUDY (see Section 5.1). Over the redshift range of interest, these sources should show red $H_{160} - [3.6] > 1.1$ mag colors (see also Roberts-Borsani et al. 2020).

The application of an $H_{160} - [3.6] > 1.1$ mag selection criterion to our sample results in the selection of two sources (ID GNDY-6469616570, R.A. = 12:36:46.965, decl. = 62:16:57.04, and ID GSDY-2431346284, R.A. = 03:32:43.135, decl. = −27:46:28.46). . In both cases, their [3.6]–[4.5] colors are consistent with that predicted by the template. Despite the tantalizing nature of these candidates, both candidates show relatively blue ($J_{125} - H_{160} < 0$ mag) colors, bluer than expected for a passively evolving galaxy at $z \sim 8$ and thus potentially calling into question the robustness.

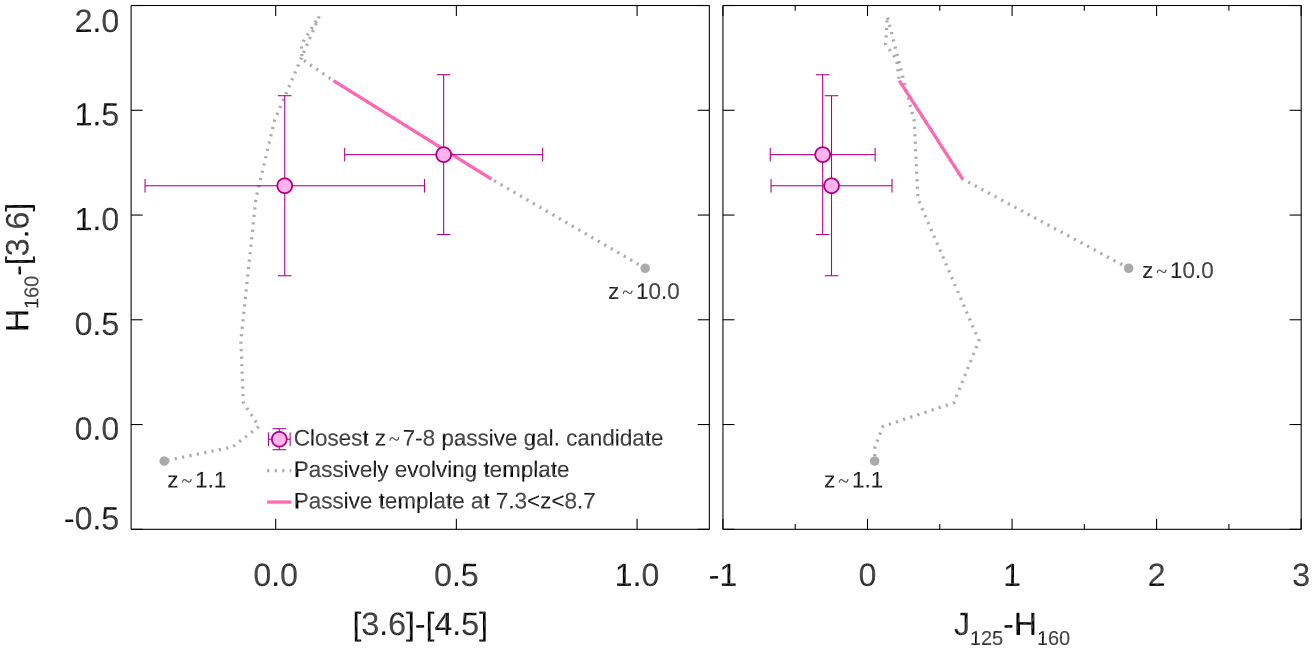

Figure 9. The two panels present synthetic tracks of a passively evolving template formed at $z_{\rm burst} = 20$ in two distinct color–color planes (dotted curve). The pink solid segments correspond to the colors of the template in the redshift range $7.5 < z < 8.5$. Passive systems at $7.5 < z < 8.5$ are expected to have red $H_{160} - [3.6] > 1.1$ mag colors. Also presented are the two sources in our sample that satisfy this criterion (pink filled circles with error bars). While their [3.6]–[4.5] color is still consistent with those of the template, they are characterized by bluer $J_{125} - H_{160} < 0$ mag than expected from the template, casting some doubts on their genuine nature of passively evolving sources at $z \sim 8$.

Appendix B
Impact of Potential Selection Bias

One long-standing concern with LBG selections is the possible exclusion of very red sources, like passively evolving or dust-obscured star-forming galaxies. In the left panel of Figure 10, we present, together with the selected sample of $z\sim8$ galaxies, tracks of dust-attenuated star-forming galaxies and passively evolving systems, extracted from the templates we adopt to estimate the main physical parameters of the stacked SEDs (see Section 5.2). These indicate that our LBG selection would fully include the colors of $A_V=1$ mag star-forming galaxies at $z>7$, while significantly sampling passive galaxies to $z\sim8$. This suggests that our selection should reasonably sample such sources, if they were common at $z\sim8$.

We quantified the impact of the LBG criteria in the selection of dusty galaxies at $z\sim8$ through a Monte Carlo simulation as follows. We computed the synthetic flux densities in the Y_{105}, J_{125}, and H_{160} bands for a set of templates with constant SFH and ages ranging between 10^6 and $10^{8.8}$ yr, after redshifting them to values uniformly distributed between $z=7.3$ and $z=8.7$. We then included the effects of dust extinction assuming a Calzetti et al. (2000) dust law, with values between $A_V=0.0$ and $A_V=3.0$ mag, in steps of 0.2 mag. Each of the resulting templates was then rescaled to have H_{160} flux densities, such as are present in our selection of sources in each field. Finally we randomly scattered the flux density according to a normal distribution whose σ was randomly extracted from the set of observed uncertainties in each field. This process was repeated 100 times for each source in our sample. The central panel of Figure 10 presents the results of applying the LBG criteria to the simulated sources at three specific redshifts. At $z\lesssim8$, the selection efficiency remains around $\gtrsim$75% over the entire range $0<A_V\lesssim1.5$ mag. At $z\sim8.5$, however, the selection efficiency shows a greater dependence on A_V, decreasing rapidly for $A_V\gtrsim0.5$ mag.

Because of photometric scatter, sources with intrinsically redder UV slopes can still satisfy the $J_{125}-H_{160}\leqslant0.5$ mag LBG criterion (see Equation (1)), potentially inflating the selection efficiencies for sources with intrinsic $A_V>0$ mag. We therefore estimated the selection efficiency by also requiring that the measured UV slopes[10] do not differ by more than 30% from the values of the corresponding unperturbed template. The results of this procedure are presented as solid curves in the central panel of Figure 10. For $A_V\sim0.2$ mag ($\beta\sim-2.0$), the efficiency is $\sim$10% lower than that at $A_V=0$ mag, and for $A_V\sim0.5$ mag ($\beta\sim-1.6$), it is $\lesssim$15% lower for all redshifts, indicating a marginal dependence of the selection efficiency on the dust extinction. Qualitatively, the smooth decrease of the density of sources within the LBG color–color box (left panel of Figure 10) toward redder $J_{125}-H_{160}$ colors suggests an intrinsic lack of dusty sources at $z\sim8$. This is quantitatively supported by the analysis we perform in Section 4.3 (see Figure 5), which shows that only $\sim$13% of the sources have UV slopes redder than $\beta=-2$, and only $\sim$4% have $\beta>-1.8$. Our simulation therefore indicates that we are likely missing at

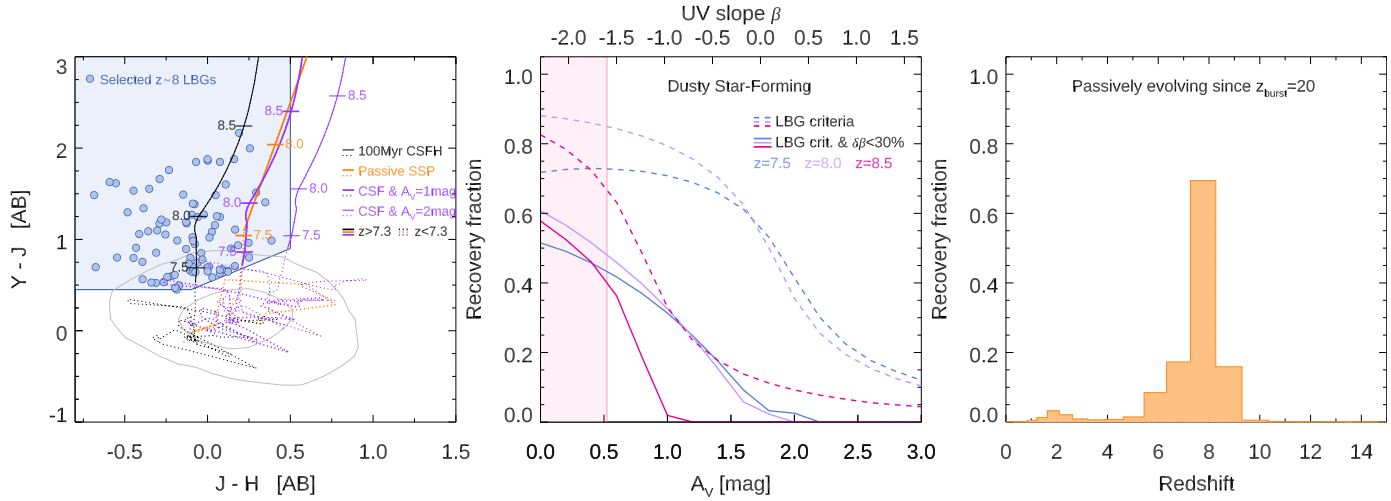

Figure 10. Illustration of the effectiveness of the adopted Lyman-break criteria in sampling the bulk of the galaxy population at $z\sim8$. In the left panel, we present the sample of candidate $z\sim8$ LBG galaxies selected for this study together with the LBG color criteria of Equation (1) (blue filled area). We replaced Y_{105}-band flux density measurements with S/N < 1 with the corresponding 1σ uncertainties. We also present synthetic tracks of a 100 Myr-old template with constant SFH and no dust (black line), after adding $A_V=1$ and 2 mag Calzetti et al. (2000) dust extinction (thick and thin purple lines, respectively) and of a system passively evolving since its formation burst at $z=20$ (yellow). For reference, the gray ellipsoids enclose ~50% and ~80% of all of the sources in the multiwavelength photometric catalog of Whitaker et al. (2019). In the central panel, we present the recovery efficiency of our LBG criteria for star-forming galaxies as a function of dust extinction, inferred from a Monte Carlo simulation (see the text for details). The dashed curves refer to the results of applying the LBG criteria only, while the solid curves also require that the measured UV slope does not differ by more than 30% from that of the unperturbed template. The top axis presents approximate values for the UV slope measured on our templates after adding the dust extinction, parameterized by the A_V values. The pink region marks the range in UV slopes we measure for our sample (discussed in Section 4.3). Finally, the panel on the right presents the efficiency of our LBG criteria in recovering passively evolving systems, inferred from the Monte Carlo simulation described in the main text. The adopted LBG criteria allow for the inclusion of passive and moderately dusty sources at $z\sim8$. In particular no significant ($\lesssim$15%) selection effects are introduced by the LBG criteria on the recovery of dusty star-forming sources for $\beta<-1.6$. Therefore, the smooth decrease of the density of sources within the color–color box toward redder $J_{125}-H_{160}$ colors (equivalent to redder β's) suggests an intrinsic lack of passive and dusty sources at $z\sim8$, increasing our confidence that our $z\sim8$ sample is reasonably representative.

[10] We computed UV slopes by fitting a power law for wavelengths redder than the Lyα at the nominal redshift.

most approximately only one source with $-2 < \beta < -1.8$ and another source with $\beta \sim -1.6$, resulting in negligible systematics in the colors measured from the stacking.

While at lower redshifts the red colors of passively evolved systems escape the Lyman-break selections, the still quite short cosmic time corresponding to $z \sim 8$ (~650 Myr) could suggest that even maximally old stellar populations at these epochs would still emit sufficient rest-frame UV radiation to enable their identification through Lyman-break criteria (assuming negligible dust content). We therefore tested the above hypothesis and the impact of the LBG color–color criteria on the selection of passively evolving sources at $z \sim 8$ through a distinct Monte Carlo simulation as follows.

We extracted simple stellar population templates with ages $< 10^{10}$ yr from the set of Bruzual & Charlot (2003) models, and processed them through CLOUDY (see Section 5.1). We then computed the flux densities in the V_{606}, I_{814}, Y_{150}, J_{125}, and H_{160} bands assuming the galaxies could have formed during a burst at $z_{\rm burst} = 20$. Such an early formation epoch (~180 Myr of cosmic time) still plausibly allows for the creation of galaxies with stellar masses $\sim 10^{8-9}\,M_{\odot}$, comparable to the stellar mass estimates we obtained for our stacks, from initial bursts lasting $\lesssim$100 Myr and SFR ~1–10 $M_{\odot}$ yr^{-1}. Moreover, the existence of significant star formation at $z \sim 15$–30 has been inferred by recent studies on individual galaxies at $z > 6$ selected to show prominent Balmer breaks, indicative of evolved stellar populations (e.g., Hashimoto et al. 2018; Mawatari et al. 2020). Nevertheless, considering even earlier formation epochs (e.g., $z \sim 30$–40) would not significantly impact our conclusions. Indeed, the age of the corresponding stellar population would be just ~60–120 Myr older than that obtained with a $z = 20$ burst, resulting in very similar predicted colors (the largest difference is observed for the $J_{125} - H_{160}$ color, with $\Delta(J_{125} - H_{160}) \lesssim 0.2$ mag, while the $H_{160} - [3.6]$ and [3.6]–[4.5] colors would differ by $\lesssim$0.1 mag). Because of the early formation scenario, we assumed dust extinction to be negligible, consistent with recent results (e.g., Bouwens et al. 2016, 2020; Dunlop et al. 2017; Casey et al. 2021).

Successively, we renormalized the SEDs such that their H_{160} flux densities probed the observed range of values in the H_{160} band, and introduced photometric scatter in a fashion similar to what was done to recover the selection efficiency of the dusty galaxies. Finally, we selected the simulated sources applying the LBG color–color criteria. The results of this procedure are summarized in the right panel of Figure 10. The plot shows that the LBG criteria allow us to include ~70% of passively evolving systems at $z \sim 8$.

The synthetic tracks of passive galaxies (see Figure 9 in Appendix A) indicate these sources at $z \sim 8$ typically have $H_{160} - [3.6] \sim 1.1$–1.5 mag. Applying the $H_{160} - [3.6] > 1.1$ mag criterion to our sample results in the identification of two candidate passive galaxies (see Appendix A). Our simulation then suggests that the intrinsic number of passive sources in the fields considered for our study should be ~3. Assuming both these additional sources were in the same luminosity bin, their contribution to the $H_{160} - [3.6]$ color measured from stacking would be $\Delta(H_{160} - [3.6]) \lesssim 0.06$ mag for any of the four luminosity bins. However, the two sources have $J_{125} - H_{160} < 0$ mag colors (see Appendix A), consistent with a more active star formation phase. We therefore consider those color corrections to be upper limits. We finally stress here that the identification of robust passive galaxy candidates requires a more detailed analysis, which, even though very interesting, we defer to a future study.

In summary, our simulations combined with the observed properties of our sample, indicate that our color measurements are not significantly affected by the LBG selection criteria.

Appendix C
Comparison to Previous SEDs at $z \sim 8$

Given the challengingly low S/N in the IRAC bands for $z \sim 8$ LBGs available so far, only a limited number of studies have been undertaken that investigate the average stellar population properties of $z \sim 8$ LBGs.

Labbé et al. (2013) leveraged deep IRAC 3.6 μm and 4.5 μm imaging from the IUDF program (Labbé et al. 2015) to explore $z \sim 8$ LBGs with a sample of 76 Y −dropouts. Their sample has luminosities that correspond to our S1 and, in part, S2. The Labbé et al. (2013) stacked colors for $H_{160} - [3.6] \sim 0.4$–0.5 mag suggest more evolved stellar populations than our results indicate. Reassuringly, the median of our $H_{160} - [3.6]$ measurements in our work for sources in common between the two studies is $H_{160} - [3.6] \sim 0.3$ mag. Our HST photometry was calibrated to be statistically consistent with the 3D-HST values. Comparison to matching sources in the 3D-HST catalog (Skelton et al. 2014; Momcheva et al. 2016) also showed that the H_{160}-band fluxes of Labbé et al. (2013) are fainter by approximately 0.15 mag. Complicating the comparison, we found that the sources in common between the two samples in our analysis are split into two distinct magnitude bins, and that our sample also includes sources with bluer $H_{160} - [3.6]$ colors than those found by Labbé et al. (2013) for the same range of magnitudes.

Song et al. (2016) discussed $z \sim 8$ median SEDs from stacking the photometry of 77 sources at $7.5 < z_{\rm phot} < 8.5$ from the Finkelstein et al. (2015b) sample over the CANDELS GOODS fields. Their brightest bin ($M_{\rm UV} \sim -20.95$ mag) exhibits a moderately red color $H_{160} - [3.6] \sim 0.3$ mag, consistent with our estimate for S1. However, the lack of detection in the 4.5 μm band, and in the other IRAC bands for lower-luminosity bins, is likely the result of the shallower depth of the IRAC data available at that time (before GREATS), making further comparisons more difficult.

Appendix D
Comparison between Median and Inverse-variance Weighted Average Stacked SEDs

We tested the robustness of the stacked SEDs for the IRAC stacks against the adopted statistical estimator by repeating our analysis with the inverse-variance weighted average instead of the median, adopting the value of the exposure time in the relevant IRAC band for the weight.

The weighted mean stacked cutouts in the four IRAC bands are shown in Figure 11, while the resulting SEDs are presented in Figure 12. The corresponding $H_{160} - [3.6]$ and [3.6]–[4.5] colors are represented in Figure 13, while in Figure 14 we compare the main colors from the inverse-variance weighted average to those from the median analysis. Figure 14 shows that there is no significant difference between the two analyses. The limit shown in the [3.6]–[4.5] versus H_{160} plot for S4 in Figures 13 and 14 results from

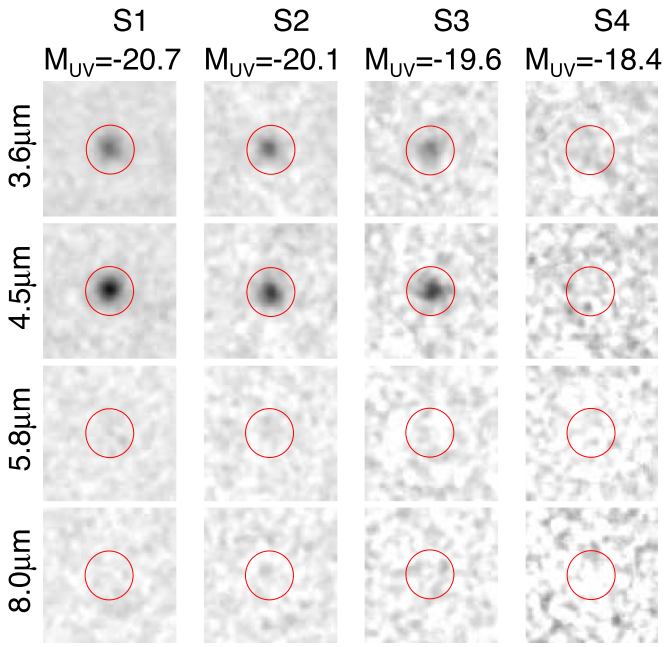

Figure 11. Stacked IRAC cutouts using the inverse-variance weighted average estimator in the four bands and for the four absolute magnitude segregations, as indicated at the top of the figure. Other plotting conventions are the same as in Figure 2.

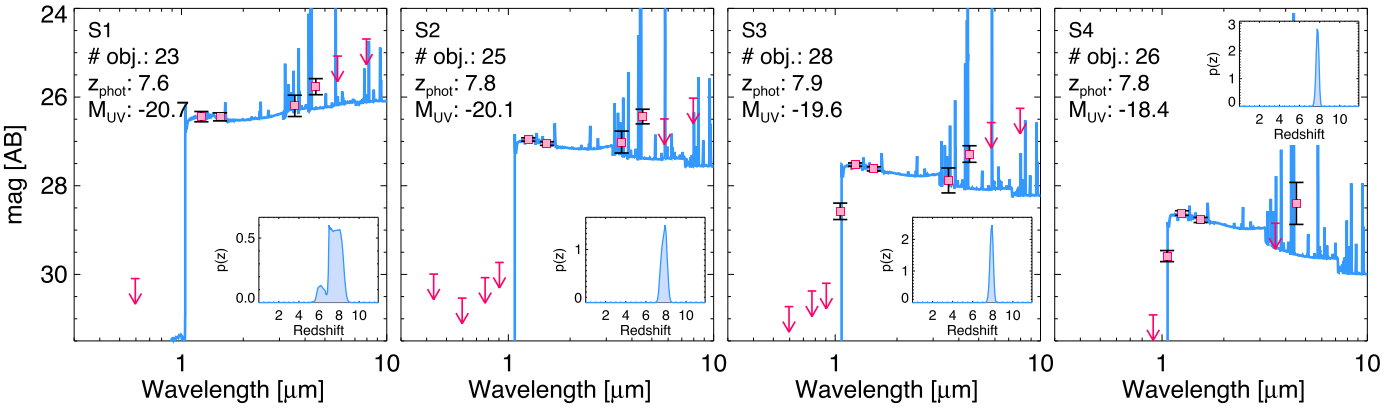

Figure 12. Full stacked SEDs in the four absolute magnitude bins for the inverse-variance weighted average stacking procedure. Other plotting conventions are the same as in Figure 3.

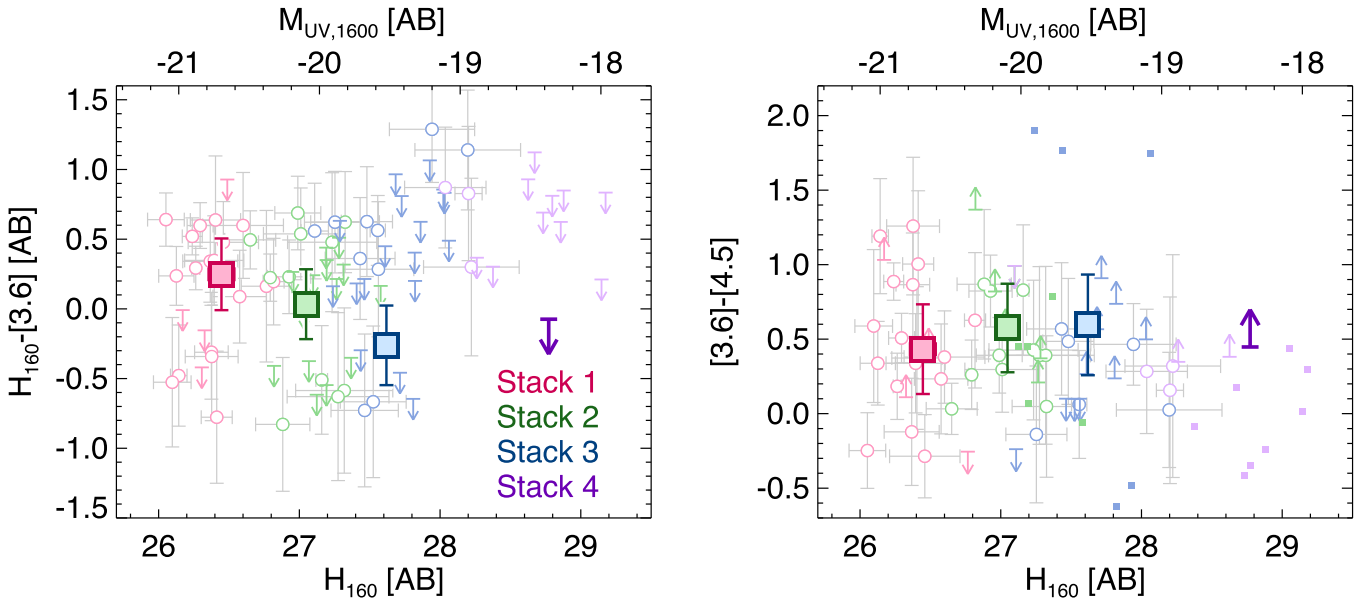

Figure 13. Color–magnitude plots for the inverse-variance weighted average stacks. The two panels shown here adopt the same plotting conventions as in Figure 4.

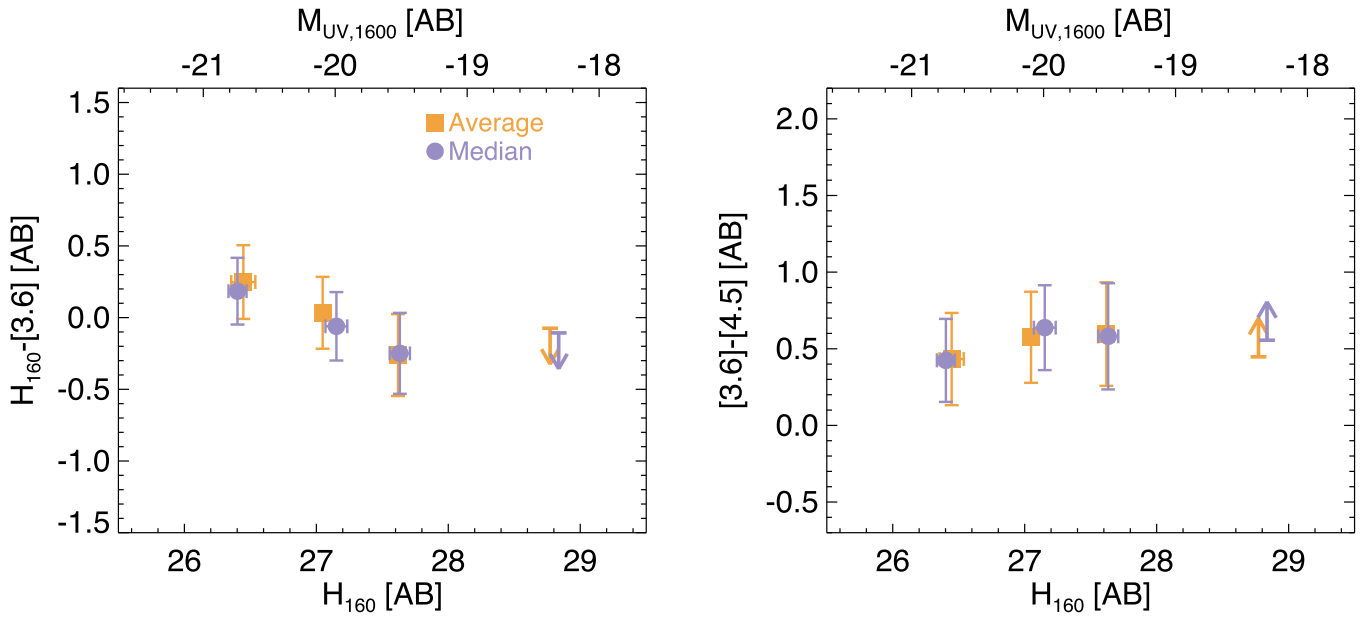

Figure 14. Comparison of the colors from the inverse-variance weighted average (filled yellow squares) to those from the median (filled purple circles). The median is our primary approach in the paper. No systematic difference is observed for any of the four bins.

a $<2\sigma$ detection in the 4.5 μm band when the inverse-variance mean is adopted.

ORCID iDs

Mauro Stefanon https://orcid.org/0000-0001-7768-5309
Rychard J. Bouwens https://orcid.org/0000-0002-4989-2471
Garth D. Illingworth https://orcid.org/0000-0002-8096-2837
Pascal A. Oesch https://orcid.org/0000-0001-5851-6649
Pieter van Dokkum https://orcid.org/0000-0002-8282-9888
Valentino Gonzalez https://orcid.org/0000-0002-3120-0510

References

Alcalde Pampliega, B., Pérez-González, P. G., Barro, G., et al. 2019, ApJ, 876, 135
Anders, P., & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, ApJS, 218, 33
Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Bhatawdekar, R., & Conselice, C. J. 2021, ApJ, 909, 144
Bouwens, R., González-López, J., Aravena, M., et al. 2020, ApJ, 902, 112
Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, ApJ, 833, 72
Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, ApJ, 795, 126
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, ApJL, 708, L69
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, ApJ, 737, 90
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, AJ, 162, 47
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
Brinchmann, J., Pettini, M., & Charlot, S. 2008, MNRAS, 385, 769
Bruzual, A. G. 1983, ApJ, 273, 105
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
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
Casey, C. M., Zavala, J. A., Manning, S. M., et al. 2021, ApJ, 923, 215
Castellano, M., Pentericci, L., Fontana, A., et al. 2017, ApJ, 839, 73
Ceverino, D., Klessen, R. S., & Glover, S. C. O. 2018, MNRAS, 480, 4842
Coe, D., Salmon, B., Bradač, M., et al. 2019, ApJ, 884, 85
Coe, D., Zitrin, A., Carrasco, M., et al. 2013, ApJ, 762, 32
Davidzon, I., Ilbert, O., Faisst, A. L., Sparre, M., & Capak, P. L. 2018, ApJ, 852, 107
Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1
De Barros, S., Oesch, P. A., Labbé, I., et al. 2019, MNRAS, 489, 2355
Dekel, A., Zolotov, A., Tweed, D., et al. 2013, MNRAS, 435, 999
Dickey, C. M., van Dokkum, P. G., Oesch, P. A., et al. 2016, ApJL, 828, L11
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Dunlop, J. S., Abraham, R. G., Ashby, M. L. N., et al. 2021, PRIMER: Public Release IMaging for Extragalactic Research, JWST Proposal. Cycle 1, 1837
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Durkalec, A., Le Fèvre, O., de la Torre, S., et al. 2015, A&A, 576, L7
Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJL, 763, L7
Endsley, R., Stark, D. P., Chevallard, J., & Charlot, S. 2021, MNRAS, 500, 5229
Erb, D. K., Steidel, C. C., Shapley, A. E., et al. 2006, ApJ, 647, 128
Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, ApJ, 821, 122
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Finkelstein, S. L., Papovich, C., Giavalisco, M., et al. 2010, ApJ, 719, 1250
Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, ApJ, 756, 164
Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015b, ApJ, 810, 71
Finkelstein, S. L., Song, M., Behroozi, P., et al. 2015a, ApJ, 814, 95
Fudamoto, Y., Oesch, P. A., Schouws, S., et al. 2021, Natur, 597, 489
Fumagalli, M., Patel, S. G., Franx, M., et al. 2012, ApJL, 757, L22
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
González, V., Bouwens, R., Illingworth, G., et al. 2014, ApJ, 781, 34
González, V., Bouwens, R. J., Labbé, I., et al. 2012, ApJ, 755, 148
González, V., Labbé, I., Bouwens, R. J., et al. 2010, ApJ, 713, 115
González, V., Labbé, I., Bouwens, R. J., et al. 2011, ApJL, 735, L34
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, A&A, 643, A8
Hamilton, D. 1985, ApJ, 297, 371
Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, ApJ, 821, 123
Harikane, Y., Ouchi, M., Ono, Y., et al. 2018a, PASJ, 70, S11
Harikane, Y., Ouchi, M., Shibuya, T., et al. 2018b, ApJ, 859, 84
Hashimoto, T., Laporte, N., Mawatari, K., et al. 2018, Natur, 557, 392
Hoag, A., Bradač, M., Brammer, G., et al. 2018, ApJ, 854, 39
Holden, B. P., Oesch, P. A., González, V. G., et al. 2016, ApJ, 820, 73
Huang, J.-S., Zheng, X. Z., Rigopoulou, D., et al. 2011, ApJL, 742, L13

Hummer, D. G., & Storey, P. J. 1987, MNRAS, 224, 801
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
Jarvis, M. J., Bonfield, D. G., Bruce, V. A., et al. 2013, MNRAS, 428, 1281
Jiang, L., Finlator, K., Cohen, S. H., et al. 2016, ApJ, 816, 16
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kewley, L. J., Zahid, H. J., Geller, M. J., et al. 2015, ApJL, 812, L20
Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2016, MNRAS, 463, 2363
Khusanova, Y., Béthermin, M., Le Fèvre, O., et al. 2021, A&A, 649, A152
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kroupa, P. 2001, MNRAS, 322, 231
Labbe, I., Bezanson, R., Atek, H., et al. 2021, UNCOVER: Ultra-deep NIRCam and NIRSpec Observations Before the Epoch of Reionization, JWST Proposal Cycle 1, 2561
Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, ApJL, 649, L67
Labbé, I., González, V., Bouwens, R. J., et al. 2010a, ApJL, 716, L103
Labbé, I., González, V., Bouwens, R. J., et al. 2010b, ApJL, 708, L26
Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, ApJL, 777, L19
Labbé, I., Oesch, P. A., Illingworth, G. D., et al. 2015, ApJS, 221, 23
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Leja, J., Speagle, J. S., Ting, Y.-S., et al. 2021, arXiv:2110.04314
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Mármol-Queraltó, E., McLure, R. J., Cullen, F., et al. 2016, MNRAS, 460, 3587
Maseda, M. V., Bacon, R., Lam, D., et al. 2020, MNRAS, 493, 5120
Mason, C. A., Trenti, M., & Treu, T. 2015, ApJ, 813, 21
Mawatari, K., Inoue, A. K., Hashimoto, T., et al. 2020, ApJ, 889, 137
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
McLure, R. J., Dunlop, J. S., de Ravel, L., et al. 2011, MNRAS, 418, 2074
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Neistein, E., & Dekel, A. 2008, MNRAS, 388, 1792
Oesch, P. A., Labbé, I., Bouwens, R. J., et al. 2013, ApJ, 772, 136
Oesch, P. A., van Dokkum, P. G., Illingworth, G. D., et al. 2015, ApJL, 804, L30
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, MNRAS, 484, 933
Pei, Y. C. 1992, ApJ, 395, 130
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016a, A&A, 594, A13
Planck Collaboration, Adam, R., Aghanim, N., et al. 2016b, A&A, 596, A108
Reddy, N., Dickinson, M., Elbaz, D., et al. 2012, ApJ, 744, 154
Reddy, N. A., Shapley, A. E., Kriek, M., et al. 2020, ApJ, 902, 123
Reddy, N. A., Shapley, A. E., Sanders, R. L., et al. 2018, ApJ, 869, 92
Renzini, A., & Buzzoni, A. 1986, Spectral Evolution of Galaxies, Astrophys. Space Sci. Libr. Vol. 122 ed. A. Renzini & C. Chiosi, (Dordrecht: Reidel), 195
Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, ApJ, 823, 143
Roberts-Borsani, G. W., Ellis, R. S., & Laporte, N. 2020, MNRAS, 497, 3440
Romano, M., Cassata, P., Morselli, L., et al. 2020, MNRAS, 496, 875
Salmon, B., Coe, D., Bradley, L., et al. 2018, ApJL, 864, L22
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Salpeter, E. E. 1955, ApJ, 121, 161
Santini, P., Fontana, A., Castellano, M., et al. 2017, ApJ, 847, 76
Schaerer, D., & de Barros, S. 2009, A&A, 502, 423
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Shim, H., Chary, R.-R., Dickinson, M., et al. 2011, ApJ, 738, 69
Shivaei, I., Reddy, N. A., Steidel, C. C., & Shapley, A. E. 2015, ApJ, 804, 149
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
Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5
Sparre, M., Hayward, C. C., Springel, V., et al. 2015, MNRAS, 447, 3548
Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, MNRAS, 464, 469
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., Bouwens, R. J., Labbé, I., et al. 2021b, ApJ, 922, 29
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2019, ApJ, 883, 99
Stefanon, M., Labbé, I., Oesch, P. A., et al. 2021a, ApJS, 257, 68
Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, ApJ, 803, 11
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Strait, V., Bradač, M., Coe, D., et al. 2020, ApJ, 888, 124
Tacchella, S., Bose, S., Conroy, C., Eisenstein, D. J., & Johnson, B. D. 2018, ApJ, 868, 92
Tacchella, S., Trenti, M., & Carollo, C. M. 2013, ApJL, 768, L37
Talia, M., Cimatti, A., Giulietti, M., et al. 2021, ApJ, 909, 23
Tran, K.-V. H., Forrest, B., Alcorn, L. Y., et al. 2020, ApJ, 898, 45
van Dokkum, P. G., Brammer, G., Fumagalli, M., et al. 2011, ApJL, 743, L15
Wang, T., Elbaz, D., Schreiber, C., et al. 2016, ApJ, 816, 84
Wang, T., Schreiber, C., Elbaz, D., et al. 2019, Natur, 572, 211
Weinmann, S. M., Neistein, E., & Dekel, A. 2011, MNRAS, 417, 2737
Whitaker, K. E., Ashas, M., Illingworth, G., et al. 2019, ApJS, 244, 16
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Williams, C. C., Labbe, I., Spilker, J., et al. 2019, ApJ, 884, 154
Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Yan, H., Dickinson, M., Eisenhardt, P. R. M., et al. 2004, ApJ, 616, 63
Zheng, W., Postman, M., Zitrin, A., et al. 2012, Natur, 489, 406
Zitrin, A., Labbé, I., Belli, S., et al. 2015, ApJL, 810, L12