

LETTER TO THE EDITOR

# Massive galaxy formation caught in action at $z \sim 5$ with JWST

Shuowen Jin (靳硕文)<sup>1,2,\*</sup>, Nikolaj B. Sillassen<sup>1,2</sup>, Georgios E. Magdis<sup>1,2,3</sup>, Aswin P. Vijayan<sup>1,2</sup>, Gabriel B. Brammer<sup>1,3</sup>, Vasily Kokorev<sup>4</sup>, John R. Weaver<sup>5</sup>, Raphael Gobat<sup>6</sup>, Clara Giménez-Arteaga<sup>1,3</sup>, Francesco Valentino<sup>1,3</sup>, Malte Brinch<sup>1,2</sup>, Carlos Gómez-Guijarro<sup>7</sup>, Marko Shuntov<sup>1,3,8</sup>, Sune Toft<sup>1,3</sup>, Thomas R. Greve<sup>1,2</sup>, and David Blaquez Sese<sup>1,2</sup>

<sup>1</sup> Cosmic Dawn Center (DAWN), Copenhagen, Denmark  
e-mail: [shuowen.jin@gmail.com](mailto:shuowen.jin@gmail.com)

<sup>2</sup> DTU Space, Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark  
e-mail: [shuji@space.dtu.dk](mailto:shuji@space.dtu.dk)

<sup>3</sup> Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark

<sup>4</sup> Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands

<sup>5</sup> Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

<sup>6</sup> Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

<sup>7</sup> Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France

<sup>8</sup> Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France

Received 19 December 2022 / Accepted 18 January 2023

## ABSTRACT

We report the discovery of a compact group of galaxies, CGG-z5, at  $z \sim 5.2$  in the EGS field covered by the JWST/CEERS survey. CGG-z5 was selected as the highest overdensity of galaxies at  $z > 2$  in recent JWST public surveys and it consists of six candidate members lying within a projected area of  $1.5'' \times 3''$  ( $10 \times 20 \text{ kpc}^2$ ). All group members are HST/F435W and HST/F606W dropouts while securely detected in the JWST/NIRCam bands, yielding a narrow range of robust photometric redshifts  $5.0 < z < 5.3$ . The most massive galaxy in the group has a stellar mass  $\log(M_*/M_\odot) \approx 9.8$ , while the rest are low-mass satellites ( $\log(M_*/M_\odot) \approx 8.4\text{--}9.2$ ). While several group members were already detected in the HST and IRAC bands, the low stellar masses and the compactness of the structure required the sensitivity and resolution of JWST for its identification. To assess the nature and evolutionary path of CGG-z5, we searched for similar compact structures in the EAGLE simulations and followed their evolution with time. We find that all the identified structures merge into a single galaxy by  $z = 3$  and form a massive galaxy ( $\log(M_*/M_\odot) > 11$ ) at  $z \sim 1$ . This implies that CGG-z5 could be a “proto-massive galaxy” captured during a short-lived phase of massive galaxy formation.

**Key words.** Galaxy: formation – Galaxy: evolution – galaxies: high-redshift – infrared: galaxies – galaxies: groups: individual: CGG-z5

## 1. Introduction

Understanding the formation of massive galaxies ( $M_* > 10^{11} M_\odot$ ) is a major goal in modern astrophysics. In  $\Lambda$ CDM cosmology, massive galaxies are formed via hierarchical clustering, where massive dark matter haloes are formed through the continuous merging of low-mass subhaloes (e.g., [Springel et al. 2005, 2008](#); [Boylan-Kolchin et al. 2009](#); [Klypin et al. 2011](#); [Somerville & Davé 2015](#)). The assembly of dark matter haloes, as one of the most important processes in the formation of structure in the Universe, can be traced by merger graphs or trees, based on analytical approaches and simulations (e.g., [Press & Schechter 1974](#); [Somerville & Primack 1999](#); [Zhang et al. 2008](#); [Neistein & Dekel 2008](#); [Roper et al. 2020](#)). Advanced cosmological simulations have displayed great details on this process, showing that the most rapid growth of nearby massive galaxies occurred at  $z > 2$  via the merging of low-mass satellites with  $M_* \lesssim 10^{10} M_\odot$  as well as in situ gas accretion (e.g., [Hopkins et al. 2009](#); [Oser et al. 2010](#); [Benson et al. 2012](#); [Hirschmann et al. 2012](#)).

However, the observational confirmation of these merging systems before coalescence is still lacking, especially for low-

mass systems at high redshift. This is not only due to the extreme faintness of low-mass galaxies at high redshift, but also their optical lines and stellar continuum are redshifted to infrared wavelengths that are not available with the *Hubble* Space Telescope (HST) and ground-based facilities. Now the situation has changed with the successful launch of the *James Webb* Space Telescope (JWST). With the unprecedented sensitivity, spatial resolution, and long wavelength coverage, we are right on the cusp of discovering more distant structures down to a significantly low-mass limit (e.g., [Morishita et al. 2022](#)).

In this Letter we report a low-mass and compact galaxy group CGG-z5 at  $z \sim 5.2$  revealed by JWST, a candidate of a merging system forming a massive galaxy. We adopt cosmology  $H_0 = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.27$ , and  $\Lambda_0 = 0.73$  as well as a Chabrier initial mass function ([Chabrier 2003](#)).

## 2. Data, measurements, and selection

### 2.1. JWST+HST data processing and measurements

We use the reduced and calibrated JWST+HST image products, and photometric and photo- $z$  catalogs that are publicly

\* Marie Curie Fellow.

available<sup>1</sup>. The JWST data reduction and calibration follow the same pipeline applied for data in Naidu et al. (2022), Bradley et al. (2022), Barrufet et al. (2022), Fujimoto et al. (2022), and Bezanson et al. (2022). A detailed description of the pipeline will be provided in Brammer et al. (in prep.). Briefly, we retrieved the pipeline-calibrated level-2 NIRCcam products from the Mikulski Archive for Space Telescopes (MAST), then further calibrated the data and processed them as mosaics using the GRIZLI package (Brammer & Matharu 2021). The GRIZLI pipeline masks various artifacts and models wisps, calibrates for a flat-field, subtracts large-scale sky background, and aligns the fully calibrated images to stars from the *Gaia* DR3 catalog (Gaia Collaboration 2023). The NIRCcam photometric zero-point correction was applied and the derived photometric zero-points are consistent with those derived by other JWST Early Release Science (ERS) teams (Boyer et al. 2022; Nardiello et al. 2022). The NIRCcam images of short wavelength channels were processed to a final pixel scale of 0.02", and 0.04" per pixel for long wavelength channels. Sources were first extracted in the longest wavelength image using Source Extraction and Photometry (SEP, Barbary 2016), and photometry was measured within an aperture of 0.5" diameter on the position from the first extraction. The measurement was performed in all JWST images with aperture correction. The HST images in these fields were also processed with GRIZLI and the photometry was measured in an identical way. The photometric redshifts were computed by fitting all available JWST+HST photometry with EAZY (Brammer et al. 2008). The catalog in the Cosmic Evolution Early Release Science Survey (CEERS) field (Finkelstein et al. 2022) consists of 14 band photometry of which seven are from JWST (F115W, F150W, F200W, F277W, F356W, F410W, and F444W), and the remaining seven are from HST (F105W, F125W, F140W, F160W, F435W, F606W, and F814W). Spectroscopic redshifts are available for 748 sources in the CEERS catalog, which were collected from literature surveys of DEEP2/3 (Cooper et al. 2011; Newman et al. 2013), Masters et al. (2017), MOSDEF (Shivaei et al. 2018), and SDSS DR15 (Aguado et al. 2019). A comparison of the photo- $z$  estimates to the spectroscopic redshifts yield negligible systematics and a dispersion of  $\Delta z/(1+z) = 2.68\%$ .

For validation, the candidates have been examined from an independent photometric extraction by Weaver et al. (in prep.) using the GRIZLI mosaics that are point spread function PSF-matched to F160W. Sources were identified from a noise-equalized stacking of F277W, F356W, and F444W using SEP. Photometry was extracted using SEP in 0.7" diameter circular apertures on all available HST and JWST bands. Following Whitaker et al. (2011), noise properties were measured in >1000 empty apertures. Photometry and associated uncertainties were then corrected to the total flux based on kron radii measured on the F444W mosaic. The two versions of photometry are found to be consistent, and we continue to use the public catalogs from Brammer et al. in following analysis.

## 2.2. Selection

CGG-z5, a compact galaxy group centered at RA 214.812 and Dec 52.8368, was selected as the highest galaxy overdensity at  $z > 2$  in the JWST catalogs produced by Brammer et al. Using GALCLUSTER (Sillassen et al. 2022), we mapped the distance of the fifth nearest neighbor  $\Sigma_{5th}$  in redshift bins at  $z =$

2–9 in JWST fields of 208 arcmin<sup>2</sup>, including CEERS-EGS, GLASS, SMACS-0723, Sunrise arc, J1235, MACS-0647, Windhorst NEP TDF, NEP 2MASS star, and Stephan's quintet. In order to filter out largely uncertain photometric redshifts,  $z_{phot}$ , in each redshift bin,  $z_{bin}$ , only sources with a 16th percentile of  $z_{phot} - z_{bin} > 8\% \times (1 + z_{bin})$  and  $z_{84th} - z_{bin} < 8\% \times (1 + z_{bin})$  were preselected for overdensity analysis – with 8% being the  $3\sigma$  uncertainty of  $1 + z$  from comparison to spec- $z$  (see Sect. 2.1). CGG-z5 is found with the highest overdensity in  $\Sigma_{5th}$  which is  $6.6\sigma$  larger than the mean galaxy density at  $5 < z < 6$  in the CEERS field. As shown in Fig. 1, it hosts a compact group of galaxies, where six candidate members with  $5.0 < z_{phot} < 5.4$  are assembled in a  $1.5'' \times 3''$  area ( $\sim 10 \times 20$  kpc<sup>2</sup> at  $z = 5.2$ ). We note that the six sources were first identified in F444W and are not blended in the image (Fig. 1). The shorter wavelength photometry was measured within a fixed aperture on the position of F444W sources, thus the photometry measurements are free of blending issues and reliable for this sample. To double-check this, we adopted an alternative overdensity analysis algorithm from Brinch et al. (2022) and reconfirmed the high galaxy overdensity in CGG-z5.

We note that EAZY is a nonparametric code that fits SEDs using a linear combination of 12 preselected flexible stellar population synthesis (FSPS) templates. It is efficient at deriving photo- $z$  for large samples, while further detailed modeling is required for deriving physical parameters of individual sources. To this end, we ran BAGPIPES (Carnall et al. 2018) SED fitting with JWST+HST photometry. Following Heintz et al. (2022), we assumed a constant star formation history (SFH) and the attenuation curve of Salim et al. (2018). We also adopted radiation fields in the range of  $-3 < \log U < -1$ , a metallicity grid from 0 to solar metallicity ( $Z_{\odot}$ ), an  $A_V$  grid from 0 to 2, and an age grid from 1 Myr to 2 Gyr. We found excellent agreement in  $z_{phot}$  between BAGPIPES and EAZY (Fig. A.1). We thus reran BAGPIPES adopting the photometric redshifts as derived by EAZY. The results are listed in Table 1. We note that the stellar masses from BAGPIPES are  $\sim 1.5$  lower compared to those inferred with EAZY for which the difference is caused by the different sets of templates and algorithms. We adopted this 0.18 dex difference between the two codes as a systematic uncertainty in our stellar masses.

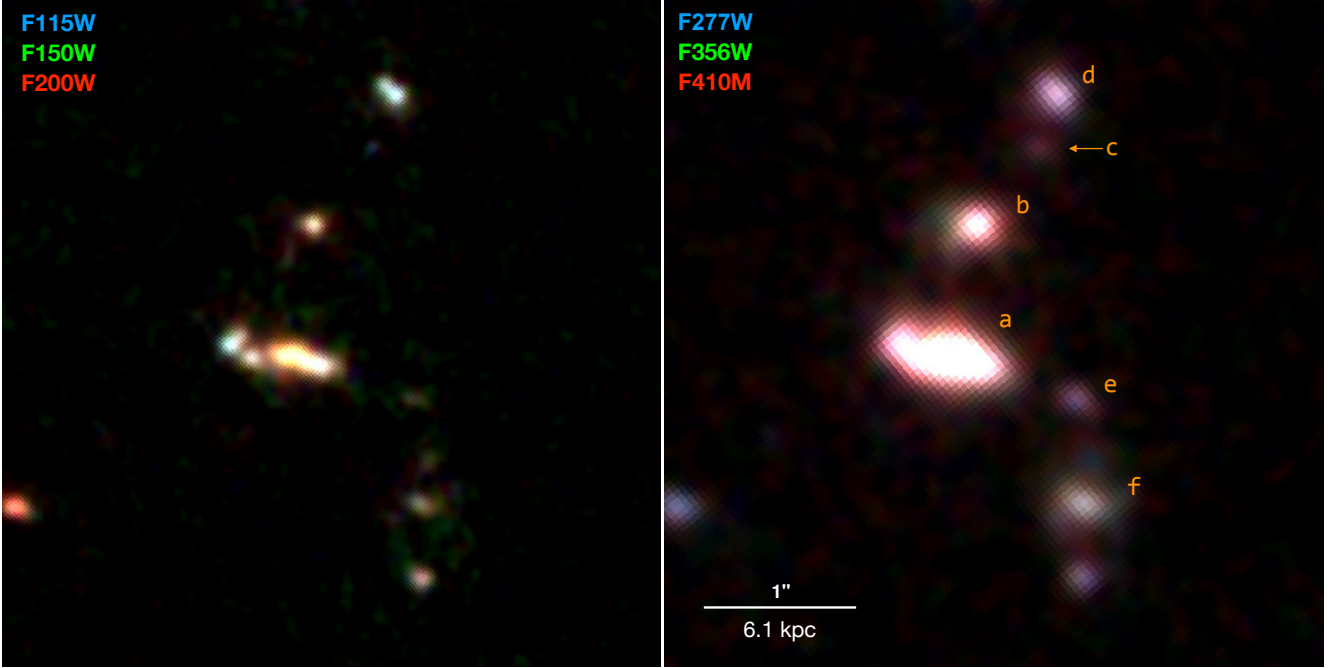
## 3. Results

### 3.1. A compact galaxy group at $z \sim 5.2$

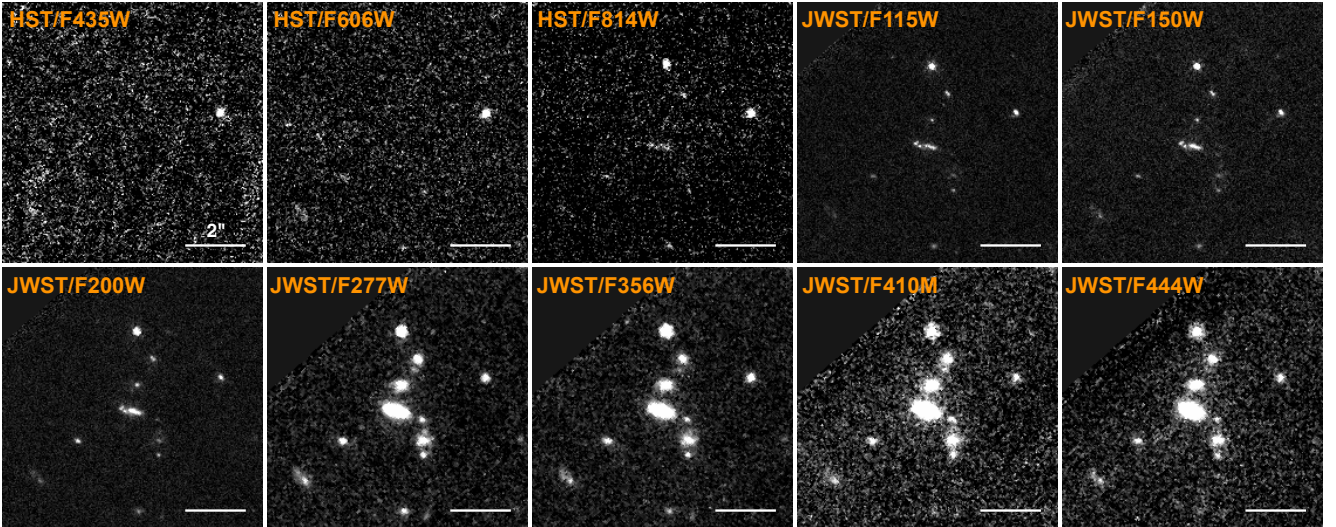
We present the JWST NIRCcam color images of CGG-z5 in Fig. 1, and show HST and JWST images in Fig. 2. The six candidate members all drop out in HST F435W and F606W data, they start to emerge in F814W, and are all well detected in NIRCcam bands with similar colors. These properties are consistent with that of Lyman break galaxies (LBGs) at  $z \sim 5$ .

In Fig. 3, we show the best-fitted SEDs from EAZY. The 14 band photometry of JWST+HST are well fitted with a single peak at  $z \sim 5.2$  in the redshift probability distribution function PDF( $z$ ). All of the six galaxies agree well at  $z \approx 5.2$  within a  $1\sigma$  uncertainty (Sect. 2.2). We note that the excesses in F277W and F410M exhibited in the six SEDs support the  $z \sim 5.2$  solutions, as the F277W band brackets multiple strong lines [OII]+H $\gamma$ +H $\beta$ + [OIII] and F410M flux is boosted by strong H $\alpha$ + [NII]. The best-fitted templates are all LBG-like with low dust attenuation, which is consistent with the nondetection in MIPS, Herschel, SCUBA2, and radio images. For a sanity check, we compared our results to two alternative versions of photo- $z$ ,

<sup>1</sup> <https://s3.amazonaws.com/grizli-v2/JwstMosaics/v4/index.html>



**Fig. 1.** JWST RGB color images of CGG-z5 at  $z \sim 5.2$ . *Left:* color image composed by NIRCам F115W (blue), F150W (green), and F200W (red) data. *Right:* color image composed by NIRCам F277W (blue), F356W (green), and F410M (red) data. Candidate members are labeled with text in orange. The two color images are matched in the World Coordinate System, and the RGB frames are composed in linear scale with identical limits.



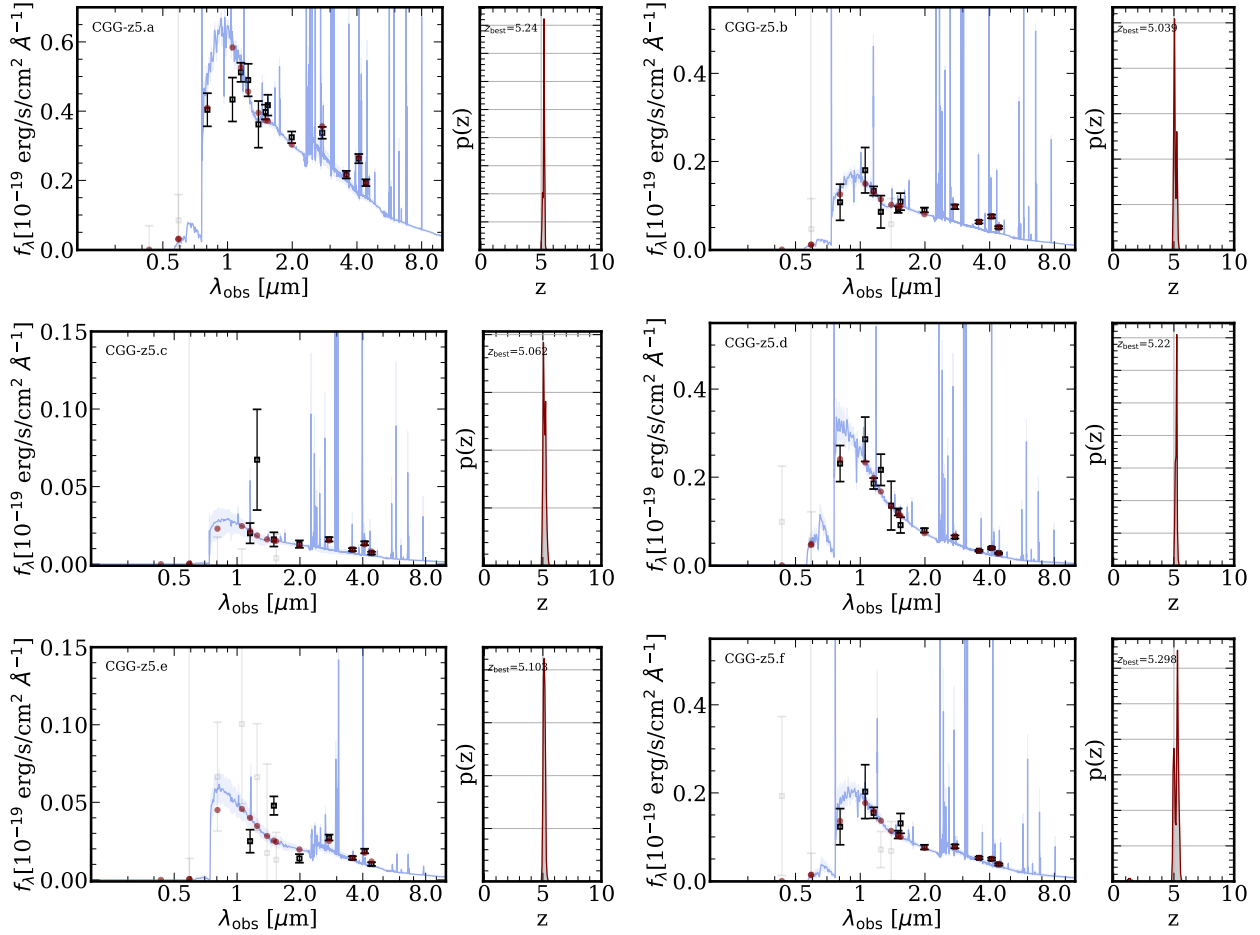
**Fig. 2.** HST and JWST cutouts in  $8.5'' \times 8.5''$  size on CGG-z5. Telescopes and filters are shown in text in each cutout. The white bar in the bottom right marks a  $2''$  scale (12 kpc). All images are shown in linear scale with identical limits.

one computed from an independent photometric extraction by Weaver et al. and the other one from BAGPIPES (see Sect. 2.2). We found that the three versions of photo- $z$  are in excellent agreement (Fig. A.1). These together suggest that CGG-z5 is a compact group of galaxies at  $z \sim 5.2$ .

We list the results of EAZY and BAGPIPES SED fitting results in Table 1. The most massive member CGG-z5.a has a stellar mass  $M_* \sim 10^{9.8} M_\odot$  and a star formation rate (SFR)  $\sim 50 M_\odot \text{ yr}^{-1}$ , placed  $\sim 2\times$  above, yet still within the scatter of the main sequence at  $z \sim 5$  (Schreiber et al. 2015). The satellites around CGG-z5.a are less massive down to  $\log(M_*/M_\odot) \approx 8.4$  with low SFRs  $\approx 2\text{--}12 M_\odot \text{ yr}^{-1}$ . The stellar mass ratios between

the center and satellite galaxies are larger than 4:1, suggesting minor mergers. Interestingly, CGG-z5.a is resolved into three components in the high resolution NIRCам images (Fig. 1, left), and CGG-z5.d is resolved into two, suggesting ongoing mergers, star-forming clumps, or extra low-mass satellites.

CGG-z5 has some unique features with respect to previously identified compact structures. For example, compared to the compact group HPC1001 at  $z = 3.61$  (Sillassen et al. 2022, and in prep.), CGG-z5 has a higher projected number density by a factor of  $\sim 3$  and is less massive by an order of magnitude in stellar mass. AzTEC5, a starbursting group at  $z \sim 3.6$  hosting four candidate members (Gómez-Guijarro et al. 2018),



**Fig. 3.** SEDs of candidate members in CGG-z5. JWST+HST photometry is shown via data points with error bars (black:  $S/N > 2$ ; light gray:  $S/N < 2$ ), which are best fitted by galaxy templates in blue. The uncertainty of the best-fit template is shown in light blue, and the best template flux in filters is shown with filled circles. On the right side of each panel, we show PDF( $z$ ) from EAZY with the best-fitted  $z_{\text{phot}}$  labeled.

**Table 1.** Candidate members in CGG-z5.

Name	$z_{\text{phot}}$	$\log(M_*/M_{\odot})$	SFR [ $M_{\odot} \text{ yr}^{-1}$ ]	$A_V$	$Z$ [ $Z_{\odot}$ ]	Age [Gyr]
CGG-z5.a	$5.24^{+0.03}_{-0.10}$	$9.81^{+0.04}_{-0.04}$	$52^{+8}_{-6}$	$0.42^{+0.06}_{-0.05}$	$0.93^{+0.04}_{-0.05}$	$0.16^{+0.04}_{-0.03}$
CGG-z5.b	$5.04^{+0.20}_{-0.04}$	$9.0^{+0.07}_{-0.05}$	$12^{+2}_{-2}$	$0.48^{+0.06}_{-0.06}$	$0.44^{+0.13}_{-0.12}$	$0.09^{+0.03}_{-0.02}$
CGG-z5.c	$5.06^{+0.20}_{-0.04}$	$8.36^{+0.35}_{-0.17}$	$2^{+1}_{-1}$	$0.48^{+0.11}_{-0.12}$	$0.32^{+0.30}_{-0.21}$	$0.14^{+0.27}_{-0.08}$
CGG-z5.d	$5.22^{+0.03}_{-0.11}$	$8.71^{+0.09}_{-0.07}$	$6^{+1}_{-1}$	$0.20^{+0.04}_{-0.04}$	$0.73^{+0.12}_{-0.14}$	$0.07^{+0.02}_{-0.02}$
CGG-z5.e	$5.10^{+0.10}_{-0.07}$	$8.52^{+0.22}_{-0.13}$	$3^{+1}_{-1}$	$0.39^{+0.08}_{-0.10}$	$0.25^{+0.17}_{-0.09}$	$0.14^{+0.18}_{-0.06}$
CGG-z5.f	$5.30^{+0.08}_{-0.32}$	$9.20^{+0.11}_{-0.09}$	$8^{+2}_{-1}$	$0.27^{+0.06}_{-0.06}$	$0.57^{+0.22}_{-0.20}$	$0.25^{+0.13}_{-0.09}$

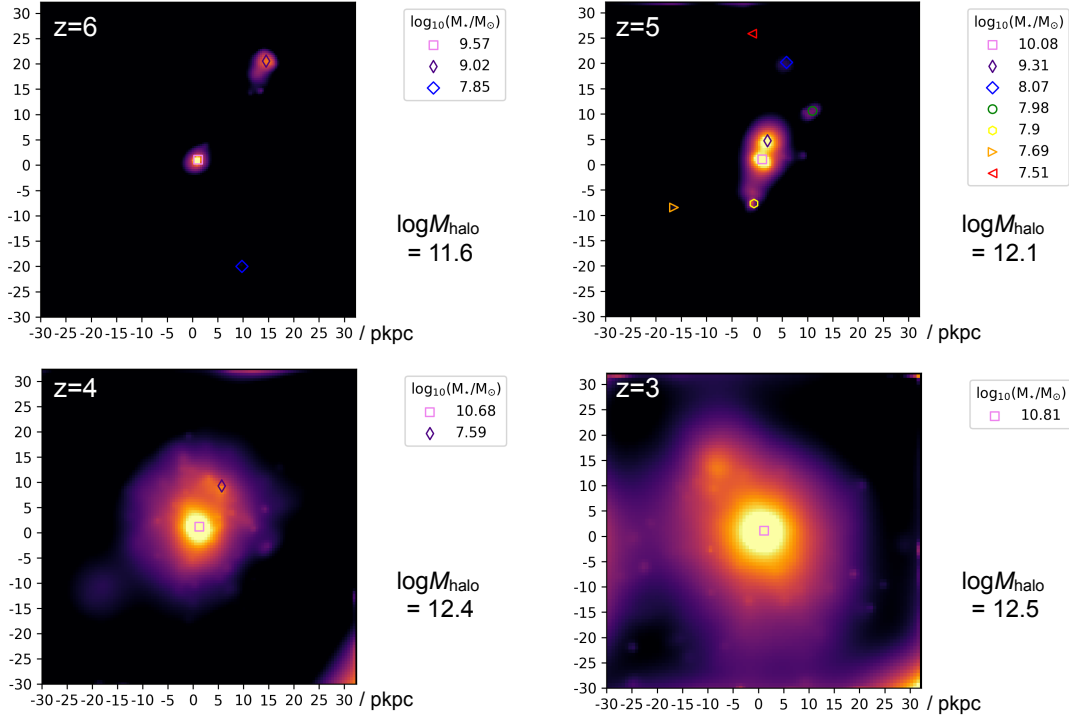
**Notes.** The  $z_{\text{phot}}$  results are from EAZY SED fitting with JWST+HST photometry, and the remaining parameters were computed by BAGPIPES at the EAZY  $z_{\text{phot}}$  assuming a constant SFH and a Salim attenuation curve. Errors are quoted from the 16th and 84th quartiles without including a systematic uncertainty of  $\sim 0.18$  dex in stellar mass.

has a total stellar mass and a SFR that are  $3\times$  and  $20\times$  higher compared to that of CGG-z5, respectively. Another interesting case is the  $z = 2.94$  SDSS J165202.64+172852.3, which was recently observed by the NIRSPEC IFU (Wylezalek et al. 2022). This structure hosts a red quasar and several companion galaxies in a scale of 10–15 kpc. However, CGG-z5 is at a higher redshift, hosts two times more galaxies (accounting for the resolved substructures of CGG-z5.a and d, see Fig. 1, left), and the center galaxy is less massive than the quasar by 1.4–2.4 orders of magnitude. Recently, a lensed protocluster at  $z \sim 7.9$  was discov-

ered with JWST in the Abell 2744 field (Morishita et al. 2022), where the most overdense region contains six galaxies within a projected distance of 60 kpc. It is at much higher redshift, but still less compact than CGG-z5 by a factor of 20 in the projected comoving number density.

### 3.2. Detectability

One could contemplate why this structure was previously undetected despite the breadth of the existing ancillary data. To



**Fig. 4.** Stellar mass map of a “CGG-z5”-like group in the EAGLE simulations. Stellar mass of individual galaxies are shown in the legend with halo mass  $M_{\text{halo}} = (M_{200}/M_{\odot})$ . The simulations suggest that CGG-z5 will form a single massive galaxy by  $z = 3$ .

answer this question, we examined CGG-z5 in the HST images and the public catalogs (Skelton et al. 2014; Stefanon et al. 2017). As shown in Fig. 2, the whole structure is totally invisible at wavelengths shorter than  $0.8 \mu\text{m}$  (indicative of the presence of a Lyman break), two members of the structure are marginally detected in F814W, and three (CGG-z5.a, b, and d) are detected in the HST F160W image. These three sources are included in the 3D-HST catalog (Skelton et al. 2014), however with largely uncertain photo- $z$  estimates spanning from  $z = 3.2$ – $5.5$ . In the EGS-CANDELS catalog from Stefanon et al. (2017), four sources (CGG-z5.a, b, d, and f) are found in this group but again with largely dispersed  $z_{\text{phot}} = 1.2$ – $5.0$ . At the same time, while CGG-z5 is detected in the IRAC images, it is also severely hampered by blending due to its compactness with respect to the IRAC beam.

In contrast, the state-of-the-art JWST imaging detects and resolves six members at seven wavelengths ( $1.15$ – $4.44 \mu\text{m}$ ), including three low-mass members previously undetected in HST images. JWST photometry significantly improved the constraints on the rest-frame UV slope and Balmer realm, while the nondetection in HST/F435W and F606W anchored the Lyman break. These combined remarkably reduced the photo- $z$  uncertainties. Moreover, the F277W and F410M bands perfectly bracket the features of strong emission lines at  $z \sim 5$ , and the implicit flux excesses in the two bands further narrowed the redshift range. In Fig. A.2, we compared the PDF( $z$ ) with and without JWST data, demonstrating that the identification of such faint structures, just one billion years after the Big Bang, can only be achieved with JWST.

#### 4. Discussion

Admittedly, the nature of such a compact and relatively low-mass structure at  $z \sim 5$  is unclear, with a range of possible evolutionary stages and paths: (1) a merging system where

the low-mass satellites will merge into the central member of the structure, eventually forming a single, massive galaxy; (2) a forming cluster core, which keeps growing via merger and cold gas accretion (Dekel et al. 2013; Daddi et al. 2021, 2022) and eventually forms a massive protocluster by  $z = 1$ ; and (3) a filamentary structure of galaxies in the line of sight. While spectroscopic follow-up observations (e.g., Keck/DEIMOS and JWST/NIRSpec) are instrumental to disentangle between these scenarios, we can resort to hydrodynamical simulations in the meantime.

In order to better understand the nature and evolution of CGG-z5, we searched for similar structures in the EAGLE simulations (Crain et al. 2015; Schaye et al. 2015), a suite of smooth particle hydrodynamic (SPH) simulations. The free parameters of the model were calibrated to reproduce the stellar mass function, stellar mass – black hole mass relation, and the galaxy disk sizes at  $z \sim 0.1$ . The model was also shown to reproduce many low-high redshift observational results which were not used in the calibration (e.g., Furlong et al. 2015; Trayford et al. 2015; Katsianis et al. 2017). For this work we used the fiducial reference volume, a periodic box of 100 comoving megaparsecs per side. To find similar structures, we first searched for galaxies more massive than  $M_* > 10^{9.5} M_{\odot}$  at  $z \sim 5$ . From the sample, we then selected those with more than three nearby galaxies within a 30 proper kiloparsec radius. Only nearby galaxies with stellar mass greater than  $10^7 M_{\odot}$  (all the galaxies have more than 100 stars and gas particles combined) were considered as companions. With this selection criteria, we found 14 such structures in the EAGLE reference volume<sup>2</sup>. We traced their evolution with time by following the most bound star particles belonging to the different structures toward lower redshift, while the progenitors

<sup>2</sup> The 3D stellar maps of the 14 EAGLE structures are publicly available here: <https://drive.google.com/drive/folders/18b4e7E1jbNgwOgYMLIjfgsIOQurgua5m>

were traced using the most bound dark matter particles. As an example, in Fig. 4 we show the one with closest resemblance to CGG-z5; it contains seven members and the two most massive members have stellar masses similar to those found in CGG-z5.

As shown in Fig. 4, the simulated structure has a halo mass of  $\log(M_{\text{halo}}/M_{\odot}) = 12.1$  at  $z = 5$ , which is consistent with the estimate of  $0.9 \times 10^{12} M_{\odot}$  for CGG-z5 from the stellar-mass-to-halo-mass relation in Shuntov et al. (2022). Intriguingly, most of these members merge into the central galaxy by  $z = 4$ , and form a single galaxy with  $\log(M_{*}/M_{\odot}) \sim 10.7$ . The stellar mass will keep growing to  $\log(M_{*}/M_{\odot}) \sim 10.8$  at  $z = 3$  and  $\log(M_{*}/M_{\odot}) \sim 11.5$  at  $z = 1$ . The other 13 simulated structures share the same fate, merging into a single, central, massive galaxy by  $z = 4 - 3$  with  $\log(M_{*}/M_{\odot}) \gtrsim 11$  at  $z \sim 1$ . It is thus tempting to speculate that the CGG-z5 system will also follow a similar evolutionary path, and we captured the process of satellites merging into a single massive galaxy.

It is also interesting to notice that the merging timescale in the simulations is  $\sim 400$  Myrs, which is in line with the fact that similar compact and overdense structures (i.e., with more than five members and overdensity  $\Sigma_{5\text{th}} > 5\sigma$ ) are not identified at lower  $z$  nor at higher  $z$  in the JWST catalogs. Looking at Fig. 4, this can be explained if similar structures have merged into a single galaxy or group with fewer members by  $z \sim 4$  and if, at  $z > 5$ , JWST observations are not deep enough to detect satellite galaxies of lower  $M_{*}$ , and therefore not meet our adopted  $\Sigma_{5\text{th}}$  overdensity criteria. This again supports the idea that CGG-z5 is in an early and short-lived phase of massive galaxy formation before coalescence, a so-called proto-massive galaxy.

We should stress that similar overdense structures, usually associated with submillimeter galaxies (SMGs) or quasars (e.g., AzTEC5, Gómez-Guijarro et al. 2018; SDSS J165202.64+172852.3, Wylezalek et al. 2022), have been found at lower redshifts ( $z \sim 3-5$ ) hosting more massive galaxies ( $\log(M_{*}) \geq 11$ ) and with higher SFRs with respect to CGG-z5, however. According to the evolutionary picture described above, these structures at  $z \sim 5$  should host an even higher overdensity of galaxies compared to CGG-z5 and should be detectable with JWST. Given the low number density of these massive structures, it is not surprising that we fail to detect any of those in our volume limited observations. Indeed, based on the number density of SMGs in the AS2UDS sample  $4 \text{ deg}^{-2}$  at  $z > 4$  (Stach et al. 2019; Gullberg et al. 2016; Riechers et al. 2020), the expected number of  $z > 4$  SMGs in the volume covered by the JWST observations in this study is  $< 1$ . The upcoming COSMOS-Web JWST survey, covering  $0.54 \text{ deg}^2$  (Casey et al. 2022), is expected to capture such structures.

An interesting question that arises is whether the merging groups, such as CGG-z5, are the main avenue for the formation of massive galaxies at  $z \sim 2-3$ . We can attempt to gain some initial insights by comparing the space density of overdense structures at  $z \sim 5$ , with the number density of massive galaxies at  $z \sim 3$ . Adopting a redshift bin of  $4.5 < z < 5.5$ , we find that the space density of galaxies with  $\log(M_{*}/M_{\odot}) > 9.8$  and  $\geq 3$  companions within a radius of 30 kpc is  $8 \times 10^{-2} \text{ cMpc}^{-3}$ . On the other hand, in our JWST catalogs, we have identified twenty galaxies at  $2.9 < z < 3.1$  with  $\log(M_{*}/M_{\odot}) > 10.8$ , corresponding to a space density of  $0.12 \text{ cMpc}^{-3}$ . While hampered by uncertainties (e.g., cosmic variance) and selection biases, the rough agreement between the two estimates makes merging groups appealing as the main progenitors of massive galaxies at cosmic noon.

*Acknowledgements.* The Cosmic Dawn Center (DAWN) is funded by the Danish National Research Foundation under grant No. 140. SJ acknowledges the

financial support from the European Union's Horizon research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101060888. GEM acknowledges financial support from the Villum Young Investigator grant 37440 and 13160. APV and TRG acknowledge support from the Carlsberg Foundation (grant no CF20-0534). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyeres-le-Chatel. We also wish to acknowledge the following open source software packages used in the analysis: Astropy (Astropy Collaboration 2013, 2018, 2022), Matplotlib (Hunter 2007), Numpy (Harris et al. 2020), Scipy (Virtanen et al. 2020), TOPCAT (Taylor 2005).

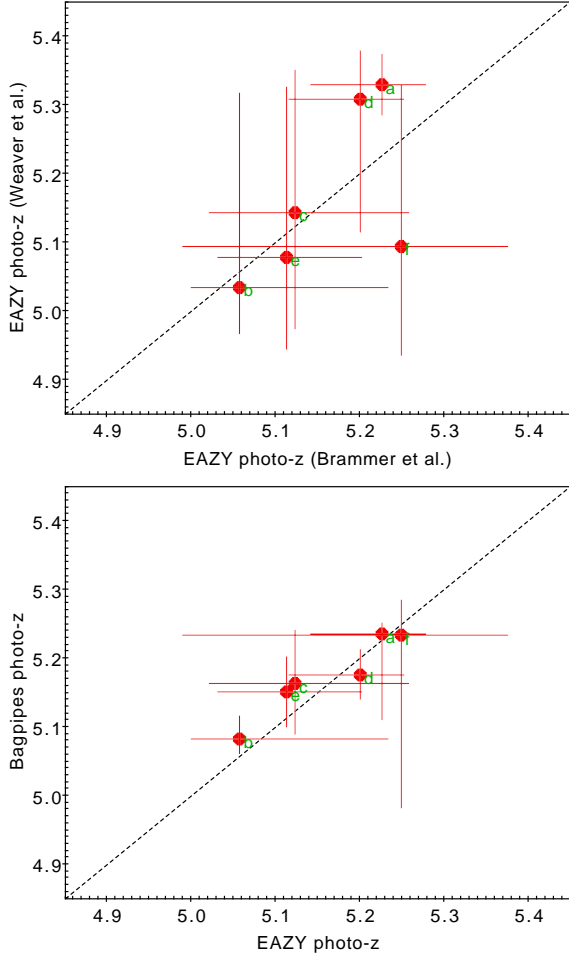
## References

- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, *ApJS*, 240, 23  
 Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33  
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123  
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, *ApJ*, 935, 167  
 Barbary, K. 2016, <https://doi.org/10.5281/zenodo.804967>  
 Barrufet, L., Oesch, P. A., Weibel, A., et al. 2022, *MNRAS*, submitted [arXiv:2207.14733]  
 Benson, A. J., Borgani, S., De Lucia, G., Boylan-Kolchin, M., & Monaco, P. 2012, *MNRAS*, 419, 3590  
 Bezanson, R., Labbe, I., Whitaker, K.E., et al. 2022, *ApJ*, submitted [arXiv:2212.04026]  
 Boyer, M. L., Anderson, J., Gennaro, M., et al. 2022, *Res. Notes Am. Astron. Soc.*, 6, 191  
 Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, *MNRAS*, 398, 1150  
 Bradley, L. D., Coe, D., Brammer, G., et al. 2022, *ApJ*, submitted [arXiv:2210.01777]  
 Brammer, G., & Matharu, J. 2021, <https://doi.org/10.5281/zenodo.5012699>  
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503  
 Brinch, M., Greve, T. R., Weaver, J. R., et al. 2022, *ApJ*, submitted [arXiv:2210.17334]  
 Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, *MNRAS*, 480, 4379  
 Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2022, *ApJ*, submitted [arXiv:2211.07865]  
 Chabrier, G. 2003, *PASP*, 115, 763  
 Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, *ApJS*, 193, 14  
 Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, 450, 1937  
 Daddi, E., Valentino, F., Rich, R. M., et al. 2021, *A&A*, 649, A78  
 Daddi, E., Rich, R. M., Valentino, F., et al. 2022, *ApJ*, 926, L21  
 Dekel, A., Zolotov, A., Tweed, D., et al. 2013, *MNRAS*, 435, 999  
 Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2022, *ApJ*, submitted [arXiv:2211.05792]  
 Fujimoto, S., Finkelstein, S. L., Burgarella, D., et al. 2022, *ApJ*, submitted [arXiv:2211.03896]  
 Furlong, M., Bower, R. G., Theuns, T., et al. 2015, *MNRAS*, 450, 4486  
 Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202243940>  
 Gómez-Guijarro, C., Toft, S., Karim, A., et al. 2018, *ApJ*, 856, 121  
 Gullberg, B., Lehnert, M. D., De Breuck, C., et al. 2016, *A&A*, 591, A73  
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357  
 Heintz, K. E., Giménez-Arteaga, C., Fujimoto, S., et al. 2022, *AAS J.*, submitted [arXiv:2212.06877]  
 Hirschmann, M., Naab, T., Somerville, R. S., Burkert, A., & Oser, L. 2012, *MNRAS*, 419, 3200  
 Hopkins, P. F., Somerville, R. S., Cox, T. J., et al. 2009, *MNRAS*, 397, 802  
 Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90  
 Katsianis, A., Blanc, G., Lagos, C. P., et al. 2017, *MNRAS*, 472, 919  
 Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102  
 Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, *ApJ*, 841, 111  
 Morishita, T., Roberts-Borsani, G., Treu, T., et al. 2022, *ApJ*, submitted [arXiv:2211.09097]  
 Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, *ApJ*, 940, L14  
 Nardiello, D., Bedin, L. R., Burgasser, A., et al. 2022, *MNRAS*, 517, 484  
 Neistein, E., & Dekel, A. 2008, *MNRAS*, 383, 615  
 Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5

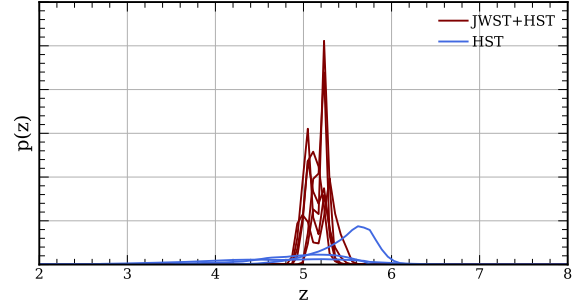
- Oser, L., Ostriker, J. P., Naab, T., Johansson, P. H., & Burkert, A. 2010, *ApJ*, **725**, 2312
- Press, W. H., & Schechter, P. 1974, *ApJ*, **187**, 425
- Riechers, D. A., Hodge, J. A., Pavesi, R., et al. 2020, *ApJ*, **895**, 81
- Roper, W. J., Thomas, P. A., & Srisawat, C. 2020, *MNRAS*, **494**, 4509
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, **859**, 11
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, **575**, A74
- Shivaei, I., Reddy, N. A., Siana, B., et al. 2018, *ApJ*, **855**, 42
- Shuntov, M., McCracken, H. J., Gavazzi, R., et al. 2022, *A&A*, **664**, A61
- Sillassen, N. B., Jin, S., Magdis, G. E., et al. 2022, *A&A*, **665**, L7
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, **214**, 24
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Somerville, R. S., & Primack, J. R. 1999, *MNRAS*, **310**, 1087
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, **435**, 629
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, *MNRAS*, **391**, 1685
- Stach, S. M., Dudzevičiūtė, U., Smail, I., et al. 2019, *MNRAS*, **487**, 4648
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, *ApJS*, **229**, 32
- Taylor, M. B. 2005, in *Astronomical Data Analysis Software and Systems XIV*, eds. P. Shopbell, M. Britton, & R. Ebert, *ASP Conf. Ser.*, **347**, 29
- Trayford, J. W., Theuns, T., Bower, R. G., et al. 2015, *MNRAS*, **452**, 2879
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, **735**, 86
- Wylezalek, D., Vayner, A., Rupke, D. S. N., et al. 2022, *ApJ*, **940**, L7
- Zhang, J., Fakhouri, O., & Ma, C.-P. 2008, *MNRAS*, **389**, 1521

## Appendix A: Comparison of photometric redshifts

We present the comparison of photometric redshifts computed from different catalogs and SED codes in Fig. A.1. In Fig. A.2, we compare the photo- $z$  PDF( $z$ ) with and without JWST data.



**Fig. A.1.** Redshift comparison. *Upper:* Comparison of EAZY photo- $z$  computed from two versions of photometry: Weaver et al. and Brammer et al. *Bottom:* BAGPIPES vs. EAZY photo- $z$ , which were all computed using photometry from Brammer et al. Group members are labeled with green text.



**Fig. A.2.** Comparison of PDF( $z$ ) with and without JWST, normalized to the integral. JWST+HST PDF( $z$ ) are shown for the six candidate members, while HST PDF( $z$ ) are only shown for the three HST/F160W-detected sources.