The Astrophysical Journal Letters, 946:L13 (35pp), 2023 March 20 https://doi.org/10.3847/2041-8213/acade4

CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST

Steven L. Finkelstein[1], Micaela B. Bagley[1], Henry C. Ferguson[2], Stephen M. Wilkins[3,4], Jeyhan S. Kartaltepe[5], Casey Papovich[6,7], L. Y. Aaron Yung[8,55], Pablo Arrabal Haro[9], Peter Behroozi[10,11], Mark Dickinson[9], Dale D. Kocevski[12], Anton M. Koekemoer[13], Rebecca L. Larson[1,14], Aurélien Le Bail[15], Alexa M. Morales[1], Pablo G. Pérez-González[16], Denis Burgarella[17], Romeel Davé[18,19], Michaela Hirschmann[20,21], Rachel S. Somerville[22], Stijn Wuyts[23], Volker Bromm[1], Caitlin M. Casey[1], Adriano Fontana[24], Seiji Fujimoto[1,25,26,56], Jonathan P. Gardner[27], Mauro Giavalisco[28], Andrea Grazian[29], Norman A. Grogin[2], Nimish P. Hathi[2], Taylor A. Hutchison[8,55], Saurabh W. Jha[30], Shardha Jogee[1], Lisa J. Kewley[31], Allison Kirkpatrick[32], Arianna S. Long[1,56], Jennifer M. Lotz[33], Laura Pentericci[24], Justin D. R. Pierel[2], Nor Pirzkal[2,34], Swara Ravindranath[2], Russell E. Ryan, Jr.[2], Jonathan R. Trump[35], Guang Yang[36,37], Rachana Bhatawdekar[38], Laura Bisigello[29,39], Véronique Buat[17], Antonello Calabrò[24], Marco Castellano[24], Nikko J. Cleri[6,7], M. C. Cooper[40], Darren Croton[41,42], Emanuele Daddi[15], Avishai Dekel[43], David Elbaz[15], Maximilien Franco[1], Eric Gawiser[30], Benne W. Holwerda[44], Marc Huertas-Company[45,46,47], Anne E. Jaskot[48], Gene C. K. Leung[1], Ray A. Lucas[13], Bahram Mobasher[49], Viraj Pandya[50,56], Sandro Tacchella[51,52], Benjamin J. Weiner[53], and Jorge A. Zavala[54]

[1] Department of Astronomy, The University of Texas at Austin, Austin, TX, USA
[2] Space Telescope Science Institute, Baltimore, MD, USA
[3] Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK
[4] Institute of Space Sciences and Astronomy, University of Malta, Msida MSD 2080, Malta
[5] Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA
[6] Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
[7] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
[8] Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA
[9] NSF's National Optical-Infrared Astronomy Research Laboratory, 950 North Cherry Avenue, Tucson, AZ 85719, USA
[10] Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA
[11] Division of Science, National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[12] Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA
[13] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[14] NSF Graduate Fellow
[15] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, F-91191, Gif-sur-Yvette, France
[16] Centro de Astrobiología (CAB), CSIC-INTA, Ctra. de Ajalvir km 4, Torrejón de Ardoz, E-28850, Madrid, Spain
[17] Aix Marseille Univ, CNRS, CNES, LAM Marseille, France
[18] Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK
[19] Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, Bellville, Cape Town 7535, South Africa
[20] Institute of Physics, Laboratory of Galaxy Evolution, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
[21] INAF–Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy
[22] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA
[23] Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK
[24] INAF–Osservatorio Astronomico di Roma, via di Frascati 33, I-00078 Monte Porzio Catone, Italy
[25] Cosmic Dawn Center (DAWN), Jagtvej 128, DK-2200 Copenhagen N, Denmark
[26] Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen Ø, Denmark
[27] Astrophysics Science Division, Goddard Space Flight Center, Code 665, Greenbelt, MD 20771, USA
[28] University of Massachusetts Amherst, 710 North Pleasant Street, Amherst, MA 01003-9305, USA
[29] INAF–Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy
[30] Department of Physics and Astronomy, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA
[31] Center for Astrophysics, Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[32] Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
[33] Gemini Observatory/NSF's National Optical-Infrared Astronomy Research Laboratory, 950 North Cherry Avenue, Tucson, AZ 85719, USA
[34] ESA/AURA, STScI, 3700 San Martin Drive, MD 21218, USA
[35] Department of Physics, University of Connecticut, 196 Auditorium Road, Unit 3046, Storrs, CT 06269, USA
[36] Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands
[37] SRON Netherlands Institute for Space Research, Postbus 800, 9700 AV Groningen, The Netherlands
[38] European Space Agency, ESA/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[39] Dipartimento di Fisica e Astronomia "G.Galilei," Universitá di Padova, Via Marzolo 8, I-35131 Padova, Italy
[40] Department of Physics & Astronomy, University of California, Irvine, 4129 Reines Hall, Irvine, CA 92697, USA
[41] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[42] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[43] Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
[44] Physics & Astronomy Department, University of Louisville, Louisville, KY 40292, USA
[45] Instituto de Astrofísica de Canarias, La Laguna, Tenerife, Spain
[46] Universidad de la Laguna, La Laguna, Tenerife, Spain
[47] Université Paris-Cité, LERMA—Observatoire de Paris, PSL, Paris, France
[48] Department of Astronomy, Williams College, Williamstown, MA 01267, USA
[49] Department of Physics and Astronomy, University of California, 900 University Avenue, Riverside, CA 92521, USA
[50] Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA
[51] Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

[52] Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge, CB3 0HE, UK
[53] MMT/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[54] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
Received 2022 November 10; revised 2022 December 15; accepted 2022 December 22; published 2023 March 27

Abstract

We present an investigation into the first 500 Myr of galaxy evolution from the Cosmic Evolution Early Release Science (CEERS) survey. CEERS, one of 13 JWST ERS programs, targets galaxy formation from $z \sim 0.5$ to >10 using several imaging and spectroscopic modes. We make use of the first epoch of CEERS NIRCam imaging, spanning 35.5 arcmin^2, to search for candidate galaxies at $z > 9$. Following a detailed data reduction process implementing several custom steps to produce high-quality reduced images, we perform multiband photometry across seven NIRCam broad- and medium-band (and six Hubble broadband) filters focusing on robust colors and accurate total fluxes. We measure photometric redshifts and devise a robust set of selection criteria to identify a sample of 26 galaxy candidates at $z \sim 9$–16. These objects are compact with a median half-light radius of $\sim$0.5 kpc. We present an early estimate of the $z \sim 11$ rest-frame ultraviolet (UV) luminosity function, finding that the number density of galaxies at $M_{\rm UV} \sim -20$ appears to evolve very little from $z \sim 9$ to 11. We also find that the abundance (surface density [arcmin^{-2}]) of our candidates exceeds nearly all theoretical predictions. We explore potential implications, including that at $z > 10$, star formation may be dominated by top-heavy initial mass functions, which would result in an increased ratio of UV light per unit halo mass, though a complete lack of dust attenuation and/or changing star formation physics may also play a role. While spectroscopic confirmation of these sources is urgently required, our results suggest that the deeper views to come with JWST should yield prolific samples of ultrahigh-redshift galaxies with which to further explore these conclusions.

Unified Astronomy Thesaurus concepts: Early universe (435); Galaxy formation (595); Galaxy evolution (594); High-redshift galaxies (734)

1. Introduction

The epoch of reionization marks the period when energetic photons (presumably from massive stars in early galaxies; e.g., Stark 2016; Finkelstein et al. 2019; Robertson 2022) ionized the gas in the intergalactic medium (IGM). Understanding when and how this process occurs is crucial to constraining both the earliest phases of galaxy formation (which kick-started this process) and how the evolution of the IGM temperature affects the star formation efficiency in low-mass halos throughout (and after) this transition.

Advances in deep near-infrared (near-IR) imaging with the Hubble Space Telescope (HST) have pushed constraints on galaxy evolution into the first billion years after the Big Bang. Studies of public blank-field surveys including the Hubble Ultra Deep Field (HUDF; Beckwith et al. 2006; Bouwens et al. 2010; Oesch et al. 2010), the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), the Hubble Frontier Fields (HFF; Lotz et al. 2017), and the Brightest of Reionizing Galaxy survey (Trenti et al. 2011) have uncovered thousands of galaxies at $z > 6$ (e.g., Bouwens et al. 2015, 2022a; Finkelstein et al. 2015, 2022a; Ishigaki et al. 2015; McLeod et al. 2016; Morishita et al. 2018; Oesch et al. 2018; Bridge et al. 2019; Rojas-Ruiz et al. 2020; Bagley et al. 2022a). The evolution of the rest-frame ultraviolet (UV) luminosity function has been well studied to $z \sim 8$ (e.g., Bouwens et al. 2015; Finkelstein et al. 2015), with some constraints placed at $z \sim 9$ and 10 (e.g., Oesch et al. 2018; Bouwens et al. 2019; Bagley et al. 2022a; Finkelstein et al. 2022a). However, little was known about the $z > 10$ universe prior to JWST beyond the unexpected discovery of an exceptionally bright $z = 10.957$ galaxy (Oesch et al. 2016; Jiang et al. 2021). This knowledge gap is due to the modest light-gathering power of the 2.4 m HST and the fact that at $z > 10$, galaxies become one-band (F160W; HST/WFC3's reddest filter) detections. Rest-UV emission from galaxies completely redshifts out of HST observability at $z \gtrsim 12.5$.

This has left a major gap in our knowledge of galaxy formation at early times. Do galaxies form stars fairly inefficiently, like our own Milky Way, and build up slowly? Or is star formation in the early universe more rapid due to high gas densities and frequent interactions? Equally exciting and unknown, does the initial mass function (IMF) begin to show signs of evolution? Models predict that top-heavy IMFs should dominate at very low metallicities (e.g., Bromm & Larson 2004), so observations should begin to see such signatures. Answering these questions about the physical processes dominating the earliest star formation requires detailed observations of the earliest galaxies to form in our universe, and JWST was designed to push our cosmic horizons to the highest redshifts. The 7× larger light-gathering power combined with the large field of view and near-IR sensitivity of NIRCam (Rieke et al. 2005) sets the stage for major advances in our ability to study early galaxy formation. Cycle 1 of JWST includes several programs encompassing hundreds of hours that all have the early universe as their primary science goal.

Indeed, in just the days to weeks after the first science data were released, several papers were submitted discussing not only the detection of objects at the expected redshifts of $z \sim 10$–11 (e.g., Adams et al. 2022; Castellano et al. 2022; Labbe et al. 2022; Naidu et al. 2022b; Whitler et al. 2023) but some candidates at $z \sim 12$–17 (e.g., Donnan et al. 2023, hereafter D23; Finkelstein et al. 2022b; Harikane et al. 2023).

[55] NASA Postdoctoral Fellow.
[56] NASA Hubble Fellow.

The existence of galaxies at such early times, and especially at such bright magnitudes for some sources, could potentially challenge early models of galaxy formation (e.g., Boylan-Kolchin 2022; Ferrara et al. 2022; Mason et al. 2023). However, these studies originally relied on very early photometric calibration; subsequent calibration data shifted the photometric zero-points significantly (Boyer et al. 2022). Now that the flux calibration and overall data reduction pipeline has stabilized, it is prudent to take a detailed look at what constraints we can place on this early epoch.

Here we use the first epoch of data from the Cosmic Evolution Early Release Science (CEERS; survey description to come in S. Finkelstein et al. 2022, in preparation) survey. CEERS was designed in part to provide our earliest detailed glimpse into the $z > 10$ universe, and these CEERS data were among the first Cycle 1 science exposures taken, included in the first publicly released data on 2022 July 14. We search these data for $z \gtrsim 9$ galaxy candidates, heretofore difficult (if not impossible) to find with HST. We place an emphasis on building a robust sample via a detailed photometric cataloging process coupled with stringent selection criteria, both backed by simulations. Section 2 describes the observations and data reduction, while Section 3 describes our photometry and photometric redshift measurements, and Section 4 discusses our sample selection procedure. We describe our sample in Section 5 and present a comparison to other early samples in Section 6. In Section 7, we present the $z \sim 11$ UV luminosity function and the cumulative surface density of early galaxies, and we discuss implications on the physics dominating galaxy formation at the earliest times in Section 8. We summarize our results and present our conclusions in Section 9. In this paper, we assume the latest Planck flat ΛCDM cosmology with $H_0 = 67.36$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3153$, and $\Omega_\Lambda = 0.6847$ (Planck Collaboration et al. 2020). All magnitudes are in the absolute bolometric (AB) system (Oke & Gunn 1983).

2. Data

CEERS is one of 13 early release science surveys designed to obtain data covering several scientific themes of astronomy early in Cycle 1, along with testing out multiple instrument modes and providing early public data to the community. CEERS consists of a mosaic of 10 NIRCam pointings in the CANDELS Extended Groth Strip (EGS) field, with six obtained in parallel with prime NIRSpec observations and four in parallel with prime MIRI observations (four of these pointings also include NIRCam wide-field slitless grism spectroscopy). Here we use imaging data from NIRCam obtained during the first epoch of CEERS, during 2022 June 21–22. This consists of short-wavelength (SW) and long-wavelength (LW) images in both NIRCam A and B modules taken over four pointings, labeled NIRCam1, NIRCam2, NIRCam3, and NIRCam6. Each pointing was observed with seven filters: F115W, F150W, and F200W on the SW side and F277W, F356W, F410M, and F444W on the LW side.

2.1. Data Reduction

The NIRCam images used here are those released with the first CEERS public data release (Data Release 0.5), which are fully described in Bagley et al. (2022b). Here we briefly highlight the key features of the data reduction, directing the reader to Bagley et al. (2022b) for more details.

We reduce the raw NIRCam imaging through version 1.7.2 of the JWST Calibration Pipeline, with custom modifications designed to correct for additional features in the data. We use the calibration reference file context[57] pipeline mapping (pmap) 0989. We begin by running stage 1 of the calibration pipeline, which performs detector-level corrections and outputs a count-rate image in units of counts per second, adopting all default parameters for this stage. We then perform custom corrections to flag and remove snowballs from all exposures, subtract off the large-scale wisps in F150W and F200W using wisp templates created by the NIRCam team, and measure and remove $1/f$ noise via a median measured (amplifier by amplifier) along rows and columns. The images are then flat-fielded and flux-calibrated using stage 2 of the calibration pipeline, again adopting the default values, to produce images in units of MJy sr^{-1}. The pmap 0989 reference files include ground flats that have been corrected for illumination gradients measured with in-flight data and improved but still preliminary photometric calibration reference files. We find that the flux calibration does a good job of synchronizing the zero-points across the NIRCam detectors (to within the 2%–5% level; Bagley et al. 2022b) but that an additional absolute flux calibration may be required at the few percent level (see Section 3.6). These flat and photometric calibration reference files will continue to be improved and updated throughout Cycle 1.

We align the images using a custom version of the TweakReg routine, which is designed to register an image to an absolute world coordinate system frame by matching sources detected in each image with those in a reference catalog. Our modified version of the routine uses SExtractor (hereafter SE; Bertin & Arnouts 1996) to measure source centroids in each individual image. We then align each image to a reference catalog constructed from an HST F160W mosaic with astrometry tied to Gaia EDR3 (see Section 2.2 and Koekemoer et al. 2011, for details on the methods used for the mosaic construction). We first align images from the same detector but different dithers to each other, allowing for shifts in x and y and achieving an rms of ∼3–6 mas source^{-1} for this relative alignment. We next align all images to the F160W reference image, allowing xy shifts and rotations in the SW images and an additional scaling factor to account for large-scale distortions in the LW images. The rms of this absolute alignment is ∼12–15 and ∼5–10 mas when comparing WFC3 to NIRCam and NIRCam to NIRCam, respectively. We note that we followed a slightly different procedure for NIRCam3, aligning F277W to HST/WFC3 F160W and then using F277W as the new reference for all other NIRCam filters. This altered procedure was required to address additional offsets registered in one portion of the F160W image in this region (see Bagley et al. 2022b, for details).

Finally, we create mosaics for each pointing in all filters in the following way. We subtract a pedestal value off of each individual image and scale the read-noise variance maps such that they include an estimate of the robustly measured sky variance. The mosaics are then created using the Resample routine in stage 3 of the calibration pipeline, which uses the drizzle algorithm with an inverse variance weighting (Casertano et al. 2000; Fruchter & Hook 2002). We drizzle the images to an output pixel scale of $0\farcs03$ pixel^{-1} and use the

[57] jwst-crds.stsci.edu, pmap 0989 corresponds to NIRCam instrument mapping 0232.

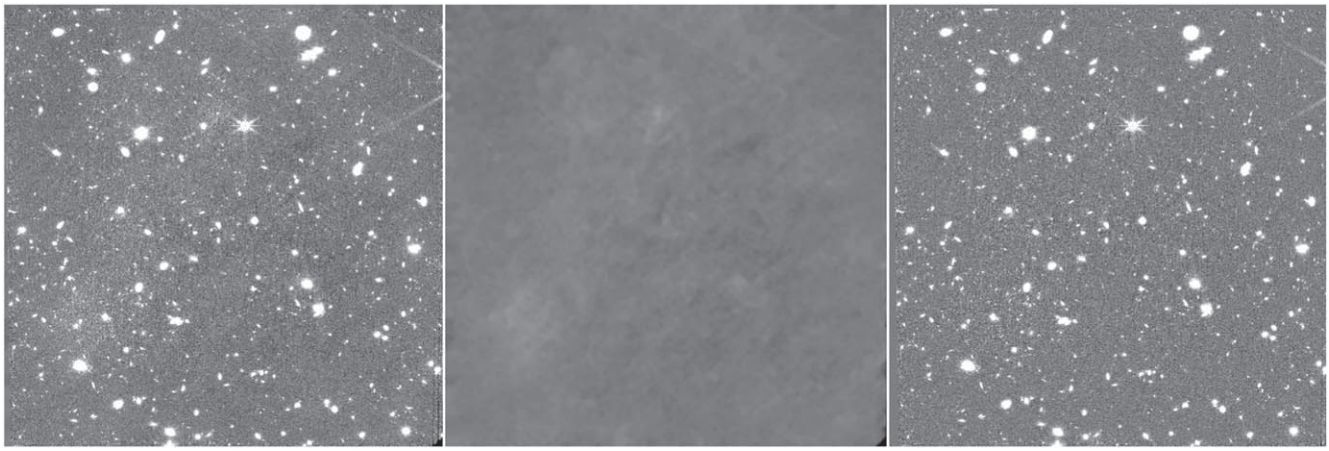

Figure 1. Example of the results of our background-subtraction procedure. The left panel shows a zoom-in on the science image of the ALONG module of our F444W mosaic in the CEERS1 pointing. The middle panel is the derived background, and the right panel is the background-subtracted science image. By progressively masking out objects in smaller tiers, this method is able to capture both small- and large-scale fluctuations. The full details of this process are available in Bagley et al. (2022b).

same tangent point as that of the HST mosaics, such that the images in all filters, NIRCam and HST, are pixel-aligned.

2.2. HST Imaging Data

The EGS field has archival HST imaging from the All-wavelength Extended Groth Strip International Survey (AEGIS; Davis et al. 2007), CANDELS (Grogin et al. 2011; Koekemoer et al. 2011), 3D-HST (Momcheva et al. 2016), and various follow-up programs. The entire CEERS field is covered by F606W, F814W, F125W, (shallow 800 s) F140W, and F160W; portions are covered by F105W. For CEERS, we produced an updated version (v1.9) of the CANDELS EGS mosaics,[58] specifically aligning their astrometry onto Gaia DR3 and on the same 30 mas pixel scale as our NIRCam images. In each of the HST mosaics, we create smaller cutouts to match the footprints of the drizzled NIRCam mosaics in each pointing. In this way, we have pixel-aligned imaging in 12–13 filters per field (NIRCam1 does not include F105W coverage) from $\sim$0.5 to 5 μm.

3. Photometry

In this section, we describe the creation of our HST+JWST catalog. As the focus is on high-redshift galaxies, this catalog is optimized for faint, compact sources. To ensure accurate photometric redshifts, significant attention is paid to calculation of accurate colors and uncertainties.

3.1. Background Subtraction

The procedure adopted to remove background nonuniformities is described in detail by Bagley et al. (2022b). Briefly, it involves several different tiers of source masking aimed at removing large-scale structures—including the wings of bright, extended galaxies—while preserving small-scale structures (including the wings of faint galaxies). We have found this procedure to be more effective than the built-in background-subtraction procedure in SE , and we therefore use these images as inputs to SE (Section 3.3) and disable its background-subtraction step. We show an example of our background subtraction results in Figure 1.

3.2. Point-spread Function Matching

The FWHM of the NIRCam point-spread function (PSF) varies significantly across the wavelengths of the filters used. As the selection of high-redshift galaxies depends on our ability to measure accurate colors, ensuring that a similar fraction of light is measured in all bands is crucial. In this work, we accomplish this by matching the PSFs in our images to the F444W filter, as it is the reddest filter and thus has the largest PSF (FWHM = 0.″161).

Our procedure for PSF matching follows Finkelstein et al. (2022a), which we summarize here. In each filter, we create a preliminary photometric catalog made using SE v2.25.0 (Bertin & Arnouts 1996) and identify potential PSF stars by searching for the stellar locus in a plane of half-light radius versus source magnitude (making custom cuts in both quantities for each filter). We excluded objects with neighbors within 50 pixels with magnitudes brighter than the magnitude of the star in question minus 1. As stars were more difficult to identify in some bands (e.g., F115W and the shallower F410M), we combined stars identified in the SW and LW channel bands to one list for each channel, which typically resulted in 10–20 stars per NIRCam pointing. We then visually inspected each star in each image, removing stars near detector edges, with other defects, or with close neighbors not excluded by the previous cut to generate one star list per filter per pointing.

The PSFs were then generated by stacking stars that passed this inspection. As our observations in all four pointings utilized the same dither pattern and were taken at a similar time, we create one PSF per filter by stacking all stars over all four pointings, increasing the signal-to-noise ratio (S/N) of our PSF. For each star, we extract a 101 $\times$ 101 pixel box, upsample by a factor of 10, measure the centroid, and shift the star to be centered in this upsampled image. We then downsample back to the native resolution, rotate the star by a random position angle (to account for situations when the position angle of the observations was not identical), and normalize the star's peak flux to unity. The final PSF was made by median-combining

[58] https://ceers.github.io/releases.html#hdr1

Table 1
Imaging Data Summary

Camera	Filter	FWHM	PSF Enclosed Flux ($d = 0\farcs2$)	Point-source Limiting Magnitude (5σ)	Zero-point Correction
JWST/NIRCam SW	F115W	$0\farcs066$	0.80	29.2	1.07 ± 0.03
JWST/NIRCam SW	F150W	$0\farcs070$	0.80	29.0	1.05 ± 0.02
JWST/NIRCam SW	F200W	$0\farcs077$	0.76	29.2	1.03 ± 0.03
JWST/NIRCam LW	F277W	$0\farcs123$	0.64	29.2	1.00 ± 0.03
JWST/NIRCam LW	F356W	$0\farcs142$	0.58	29.2	1.01 ± 0.02
JWST/NIRCam LW	F410M	$0\farcs155$	0.56	28.4	1.00 ± 0.02
JWST/NIRCam LW	F444W	$0\farcs161$	0.52	28.6	0.99 ± 0.02
HST/ACS	F606W	$0\farcs118$	0.70	28.6	1.02 ± 0.02
HST/ACS	F814W	$0\farcs124$	0.63	28.3	0.96 ± 0.03
HST/WFC3	F105W	$0\farcs235$	0.35	27.1	0.97 ± 0.04
HST/WFC3	F125W	$0\farcs244$	0.33	27.3	0.95 ± 0.03
HST/WFC3	F140W	$0\farcs247$	0.32	26.7	0.95 ± 0.03
HST/WFC3	F160W	$0\farcs254$	0.30	27.4	0.95 ± 0.03

Note. The PSFs were created by stacking stars across all four pointings. For our photometry, we PSF match all filters with FWHM smaller than the F444W PSF FWHM to the F444W PSF. We note that the HST imaging used is on the same 30 mas pixel scale, which affects the FWHM of the PSF. The limiting magnitude is that measured in a $0\farcs2$ diameter aperture on the unmatched images, corrected to a total based on the PSF flux enclosed in that aperture size, averaged over the four fields. The derived corrections to the photometric zero-points for each filter were derived using the best-fitting EAZY models to ∼900 galaxies with secure spectroscopic redshifts. These corrections are due to a combination of residual differences between our estimated total fluxes and true total fluxes, differences between the model templates and true galaxies, and true photometric zero-point inaccuracies (using the photometric reference files from pmap 0989 for NIRCam). Because these corrections depend specifically on our photometry procedure, they may not be appropriate for other photometric catalogs.

the individual stars. The final PSFs have a centroiding accuracy of ∼0.05–0.1 pixels.

Kernels to match bluer PSFs to F444W were created with the PYPHER Python routine[59] (Boucaud et al. 2016), and these bluer images were then convolved with their respective kernels. We included the HST/ACS F606W and F814W imaging in this PSF-matching process, as their PSF FWHMs are smaller than that of F444W. However, the HST/WFC3-band PSF FWHMs are larger than F444W, so we do not convolve these images. This will necessitate a correction accounting for the lower fraction of flux encompassed in a given aperture in these filters, which we discuss in Section 3.3.1.

We tested our PSF-matching process by measuring curves of growth of the PSF stars in the images. We find that the median enclosed flux at an aperture diameter of $0\farcs3$ was within 5% (and often less) of the F444W value for all filters (while prior to this PSF-matching process, the bluer filter measurements encompassed ∼20% more light at this radius). We provide the median FWHM values in each filter in Table 1.

3.3. Source Extraction

We use SE in two-image mode to measure accurate photometry for each of our four pointings. It requires a detection image to identify sources. We elect to use the inverse variance–weighted sum of the PSF-matched F277W and F356W images as our detection image to better detect faint sources. We do not include F200W in this stack, as the Lyα break enters this filter at $z = 13.4$, and we do not wish to bias our catalog against extreme-redshift galaxies (the blue edge of F277W corresponds to a Lyα break redshift of $z = 18.9$), while the inclusion of F444W could have potentially begun to bias against very blue sources.

Using this detection image, we run SE cycling through the seven NIRCam images and six HST images as the measurement image. The key source detection parameters were initially optimized using the CEERS simulated imaging[60] (Bagley et al. 2022b) and further tweaked by inspecting their performance on the final mosaics. The parameters we used best recovered faint sources while minimizing contamination by spurious objects. These key parameters were DETECT_THRESH = 1.3, DETECT_MINAREA = 5 pixels, and a top-hat convolution kernel with a width of 4 pixels.

We forced SE to skip the background-subtraction step, as this was previously removed (Section 3.1). We use MAP_RMS for the source weighting. As the pipeline-produced ERR images include Poisson noise, they are not appropriate for source detection. We thus convert the weight map associated with the detection image into an effective rms map by taking 1/sqrt(WHT) and assign this to the detection image. For the measurement image, we use the pipeline ERR image.

Following previous work (e.g., Finkelstein et al. 2022a), we measure colors in small elliptical apertures, which has been shown to accurately recover the colors of distant galaxies. In SE, these apertures are defined by two parameters: a Kron factor and a minimum radius. We set these two quantities to (1.1, 1.6). These are the same values found by Finkelstein et al. (2022a) via optimization simulations, and we verified via our own simulations (Section 3.3.1) that the S/N in these apertures was significantly better than larger parameters, and that gains in S/N were negligible for smaller values. We estimate an aperture correction to the total flux for these small apertures by performing a second run of SE on the F444W image with the Kron parameters set to the default "MAG_AUTO" parameters of (2.5, 3.5), deriving an aperture correction as the ratio between the flux in this larger aperture and that in the smaller aperture for each object. The median aperture correction across all four fields was 1.5. This aperture correction was then applied multiplicatively to the fluxes and uncertainties for all filters.

[59] https://pypher.readthedocs.io

[60] https://ceers.github.io/releases.html#sdr3

3.3.1. Residual Aperture Correction

While we use small Kron apertures to derive accurate colors, the aperture correction applied above should yield total fluxes close to the true value. However, several previous studies have noted that the default Kron parameters we use for this aperture correction can miss light in the wings of the PSF (e.g., Bouwens et al. 2015; Finkelstein et al. 2022a), yielding underestimates of the total fluxes at the 5%–20% level.

We estimate these corrections using source-injection simulations, adding 3000 mock sources to our real images in each field. We add sources from $m = 23$ to 27 mag (to ensure a robust photometric measurement), with a lognormal half-light radius distribution peaking at ∼1.5 pixels (∼0.2 kpc at $z = 10$; compact but modestly resolved, comparable to high-redshift sources; see Section 5.4) and a lognormal Sérsic parameter distribution peaking at 1.2. These mock sources were generated with GALFIT (Peng et al. 2002) and added at random positions to the F277W, F356W, and F444W images. We combined the former two to create a detection image and ran SE in the same way as on our real data to generate an F444W catalog (focusing on this one band, as all images were PSF-matched to F444W). Finally, we match sources in the SE catalog to their input values and compare the ratio of input-to-recovered fluxes. We find a median ratio of 1.08, measured in the range $25 < m_{\rm F444W} < 26$. There is a slight trend with magnitude of lower corrections for brighter sources and higher for fainter sources, but only at the 1%–2% level. We thus elect to use a single correction factor of 1.08 to all NIRCam fluxes and uncertainties.

For the HST fluxes, we elect to derive any residual aperture correction from comparison to the Finkelstein et al. (2022a) photometric catalog, which performed similar simulations to derive total fluxes. Matching sources in each of the six HST bands, we find a typical needed correction factor of ∼1.35 (±0.02). These values are roughly consistent with the combination of the correction derived in Finkelstein et al. (2022a) of 1.20 with the NIRCam correction derived here of 1.08. We apply this same 1.35 correction to all HST bands, such that the colors among these bands are not changed. We note that in Section 3.6 below, we test for the presence of any remaining photometric offsets in our catalog and find these to be small ($\lesssim$5%), indicating that our procedure for deriving total fluxes in all 13 HST + JWST bands is robust, especially for this nascent observatory.

3.4. Noise Estimation

While SE does provide an estimate of the noise, it is reliant on the accuracy of the provided error maps. Given the nascent nature of the JWST reduction pipeline, we obtain estimates of our image noise directly from the images themselves. We follow the methods of Finkelstein et al. (2022a), based on previous methodology outlined in Papovich et al. (2016). Our goal is to estimate the noise based on the number of pixels in an aperture. We fit for the noise as a function of aperture size by measuring the fluxes in circular apertures with 30 different diameters, ranging from 0.″1 (3.33 pixels) to 3″ (100 pixels).

When defining these positions, we restrict aperture placement to pixels with real nonzero values in the ERR image and a zero value in the SE segmentation map within the aperture, avoiding real objects. We also require these apertures to be nonoverlapping to avoid correlating our noise estimation. To improve statistics for smaller apertures, we placed apertures in two separate iterations: "small" ($d \leqslant 1.''5$) and "large" ($d > 1.''5$). We were able to place 5000 and 500 nonoverlapping apertures in these two iterations, respectively.

We created a detection image setting the pixels at these positions to unity, with the rest of the image set to zero, and ran SE in two-image mode. We measured fluxes at these positions in all 30 circular apertures with diameters ranging from 1 to 200 pixels. We calculated the 1σ noise in each aperture size by measuring the median absolute deviation of the measured flux values (multiplying by 1.48 to convert to a Gaussian-like standard deviation). Finally, we fit a curve to the noise in a given aperture as a function of the pixels in that aperture, using the equation (Gawiser et al. 2006)

$$\sigma_N = \sigma_1 \alpha N^{\beta}, \quad (1)$$

where σ_N is the noise in an aperture containing N pixels, and σ_1 is the pixel-to-pixel noise measured in each image as the sigma-clipped standard deviation of all nonobject pixels (see Figure 3 in Finkelstein et al. 2022a for an example of this process).

We fit the free parameters α and β with an IDL implementation of EMCEE (see Finkelstein et al. 2019 for details). We used these functional form fits for each filter to calculate the photometric uncertainties for each object using both the number of pixels in its Kron aperture (area = $\pi \times$ A_IMAGE $\times$ B_IMAGE $\times$ KRON_RADIUS^2) and a value for a given circular aperture. These values were scaled by the ratio of the error image value at the central position of a given source to the median error value of the whole map, thereby allowing the noise to be representative of the local noise level.

Finally, to account for variable image noise not captured by the error image value at the central pixel, for each object in our catalog, we also calculate a local noise measurement. This local noise was calculated at 0.″2, 0.″3, 0.″4, and 0.″5 diameter apertures by fitting a Gaussian to the negative side of the flux distribution in the 200 closest apertures from the above process.

3.5. Multiband Catalog

We compose a multiband catalog from the individual-filter catalogs created by SE. As SE cannot parse the world coordinate system in the JWST data model image headers, we use astropy.wcs WCS_PIX2WORLD to derive celestial coordinates from the SE x, y positions. We apply a photometric zero-point to convert the image from MJy sr^{-1} to erg s^{-1} cm^{-2} Hz^{-1} and apply both aperture corrections derived above to all flux and flux error estimates. We correct for Galactic extinction using an $E(B-V)$ of 0.006 for the EGS field and a Cardelli et al. (1989) Milky Way attenuation curve.

Both the MAG_AUTO-derived aperture corrections and simulation-based (and HST catalog–based) residual aperture corrections were applied to all Kron-derived fluxes and uncertainties, maintaining our accurate colors while representing the total fluxes for each source. In our final catalog, we include both these Kron-based fluxes and the fluxes measured in circular apertures with diameters ranging from 0.″05 to 2.″0. While the Kron fluxes will be used for most of the analysis, we do use the circular apertures (with a fiducial diameter of 0.″2, or 6.67 pixels) as a measure of detection significance, as the sizes of the Kron apertures can be affected by the proximity to bright sources. These circular apertures are corrected for Galactic

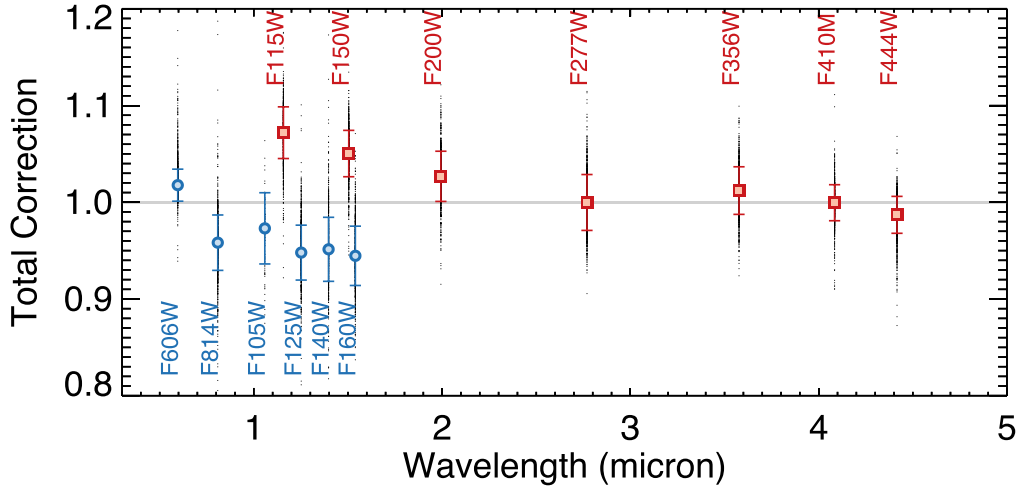

Figure 2. Total corrections applied to our catalog to account for residual systematic offsets in our total flux estimations, mismatches between the spectral templates used for photometric redshift fitting, and any remaining photometric calibration corrections (shown as red squares for NIRCam and blue circles for HST; small black dots are individual galaxies). These multiplicative offsets were derived by comparing the observed fluxes in each filter to the best-fitting EAZY model templates for ∼900 sources with robustly known spectroscopic redshifts in our field. The values are tabulated in Table 1.

attenuation but not corrected to total, as we will use them solely for detection significance.

For both the Kron and circular apertures, we calculate the noise per source following the method in Section 3.4, which is dependent on both the aperture area and the effective rms map value at the position of the source. Finally, we flag any sources that had either a zero or NaN in any error column, replacing their flux error with 10^{12} nJy (several orders of magnitude larger than any real source error).

In Table 1, we include an estimate of the limiting 5σ magnitude for our catalog. To calculate this, we use the noise functions described above to derive the flux density uncertainty in an aperture of diameter $0\farcs2$. We then measure the enclosed flux at this radius from the stacked PSF. We then divide the flux uncertainty by the enclosed fraction of flux to estimate the total noise for a point source. Finally, we multiply this value by 5 and convert to an AB magnitude. Both the enclosed flux values and the limiting magnitudes are listed in Table 1. While the depths were broadly as expected based on the prelaunch exposure-time calculator, the F115W image (with double the exposure time) was expected to be 0.3 mag deeper. The very low background at that wavelength has led to those images being more read noise–dominated than expected; thus, this image has a depth comparable to the bulk of the NIRCam filters. The primary impact is that photometric redshifts will be slightly more uncertain at $m > 29$ than originally planned.

3.6. Photometric Calibration Validation

While the photometric calibration of NIRCam has substantially improved over the first few months since science acquisition began, there is still some uncertainty, and many of the reference files still used in the pipeline are preliminary (to be refined during all of Cycle 1; see Boyer et al. 2022 for more details on the NIRCam photometric calibration used here). It is thus prudent to check the accuracy of the photometric calibration. The JWST pipeline applies a photometric calibration taking into account the area of each pixel (using a *photom* reference file), resulting in image units of MJy sr^{-1}. For the 30 mas scale of our images, the conversion factor from MJy sr^{-1} to erg s^{-1} cm^{-2} Hz^{-1} is 2.1154×10^{-31}.

To check the accuracy of this calibration, we use a sample of objects with published spectroscopic redshifts in this field (compiled by N. Hathi 2022, private communication; including DEEP2, Newman et al. 2013; MOSDEF, Kriek et al. 2015; and 3D-HST, Momcheva et al. 2016, among others). Applying a matching radius of $0\farcs5$, removing duplicates, keeping the two highest-quality redshift flags, and restricting to F160W < 24 mag, we find 988 matches that fall in our CEERS/NIRCam-covered area. These objects span $z = 0$–4, with a median $z_{\rm spec} = 1.1$.

We generate the expected fluxes of these sources in all 13 photometric bands using the EAZY (Brammer et al. 2008) photometric redshift fitting software. While in Section 4 below, we will use EAZY to measure photometric redshifts, here we run EAZY with the redshift fixed to the spectroscopic redshift value. This thus obtains the best-fitting galaxy template to the observed photometry. Assuming that this template set spans the color range of real galaxies (see Section 4.1), a comparison of the observed fluxes in a given filter to those predicted by EAZY can inform us of any systematic discrepancies in the fluxes in our catalog. We measured the median offset for the subset of these sources with a measured S/N of >5 and an EAZY goodness of fit (χ^2) of <20. This ensures that a well-fit model is found to a set of robust photometric measurements. This typically resulted in ∼850–900 sources per filter (84 in F105W, which has significantly less coverage), from which we measured a sigma-clipped median and standard deviation.

We tabulate these derived corrections in Table 1 and plot the median values and full dispersion in Figure 2. For NIRCam, the LW bands are all consistent with unity, which is a significant improvement over the state of the calibration in late summer 2022. For the SW channels, we find a needed correction on the 3%–7% level. For HST, we find that with the exception of F606W, the remaining bands are ∼3%–5% too bright. As HST is extremely well calibrated, this implies that the 35% correction we applied based on the comparison to the Finkelstein et al. (2022a) catalog may have been slightly too large (we note that that study did not complete such a zero-point offset analysis). Nonetheless, we applied these corrections. To test their accuracy, we performed another iteration of this analysis after applying these offsets and found that no

significant residual correction was present. We thus apply these corrections, listed in Table 1, to all fluxes and flux errors in our final photometric catalog. We reiterate that while photometric calibration was our motivation for this test, the resulting corrections are a combination of residual differences between our estimated and true total fluxes, differences between the model templates and true galaxies, and true photometric zero-point inaccuracies.

4. Selection of Redshift >9 Galaxies

Our analysis method produced a photometric catalog that contains our best estimates of the total fluxes in each of 13 filters spanning 0.6–5.0 μm with robust flux uncertainty values. We also include fluxes measured in a range of circular apertures. In this section, we will use this catalog to select galaxies at $z > 8.5$. Below, we will create identifiers for each object inclusive of the CEERS field the object was covered in and the SE number within that field. For example, "Maisie's Galaxy," the previously identified $z \sim 12$ galaxy candidate from Finkelstein et al. (2022b), is referred to here as CEERS2_5429.

4.1. Photometric Redshift Estimation

We measure photometric redshifts with EAZY for our entire photometric catalog, which contains ~40,000 sources across all four fields. EAZY fits nonnegative linear combinations of user-supplied templates to derive probability distribution functions (PDFs) for the redshift based on the quality of fit of the various template combinations to the observed photometry for a given source. The template set we use includes the "tweak_fsps_QSF_12_v3" set of 12 flexible stellar population synthesis (FSPS; Conroy & Gunn 2010) templates recommended by the EAZY documentation.

As shown by Larson et al. (2022b), these recommended templates do not span the blue colors expected for very high redshift galaxies. They developed six additional templates spanning bluer colors by combining stellar population spectra from BPASS (Eldridge & Stanway 2009) with (optional) nebular emission derived with CLOUDY (Ferland et al. 2017). They used models with low metallicities (5% solar), young stellar populations (stellar ages of 10^6, $10^{6.5}$, and 10^7 Myr) inclusive of binary stars, and a high ionization parameter (log $U = -2$). They showed that the inclusion of these templates significantly improved the recovery of photometric redshifts from a mock catalog derived by a semianalytic model (SAM; Yung et al. 2022) due to the better match between galaxy and template colors (see their Figure 4). These templates were also used by Finkelstein et al. (2022b) in their analysis of the $z \sim 12$ Maisie's Galaxy, where they found that the inclusion of these bluer templates significantly improved the goodness of fit between the data and the best-fitting model.

We assume a flat prior in luminosity; while bright galaxies will have a redshift distribution significantly tilted toward lower redshift, the bright end of the high-redshift luminosity function is poorly known; thus, we do not wish to bias against the selection of true, bright distant galaxies. We include a systematic error of 5% of the observed flux values and fit to our measured total flux and flux error values. In our fiducial EAZY run, the redshift can span 0–20. We perform an additional "low-z" run with the maximum redshift set to $z_{\rm max} = 7$ such that we can visualize the best-fitting low-redshift solution.

4.2. Sample Selection

Here we describe the selection criteria we use to identify candidate $z > 8.5$ galaxies. Following our previous work (Finkelstein et al. 2010, 2015, 2022a, 2022b), we utilize a combination of flux detection significance values and quantities derived from the full photometric redshift PDF (denoted $\mathcal{P}[z]$) to select our galaxy sample. As a part of this selection, we make use of the integral of the $\mathcal{P}(z)$ in $\Delta z = 1$ bins centered on integer redshift values. Specifically, we denote the unit redshift where the integral in a $z \pm 0.5$ bin is the maximum compared to all other redshifts as $\mathcal{S}_z$. For example, a source with $\mathcal{S}_z = 10$ would have $\int_{9.5}^{10.5} \mathcal{P}(z)\, dz$ greater than the integrated $\mathcal{P}(z)$ in all other $\Delta z = 1$ bins.

Our primary selection criteria are as follows.

1. An S/N, as measured in 0.″2 diameter apertures in the non-PSF-matched images, of >5.5 in at least two of the F150W, F200W, F277W, F356W, and F444W bands. We required this to be true with both the global and local noise values. This allows the selection of galaxies across a broad range of redshifts and rest-UV colors. We note that we experimented with requiring a higher-significance detection in just one band, but this significantly increased the spurious source fraction.
2. Error map values <1000 in all of the F115W, F150W, F200W, and F277W images. This includes only objects with a measurable (though not necessarily significant) flux in the four bluest filters, necessary for selection of galaxies at $z \sim 9$–13.
3. An S/N of <3 in bands blueward of the Lyα break. While studies occasionally use more stringent S/N cuts, we choose this value to account for the fact that any positive flux in all dropout bands is already accounted for by EAZY and that >1σ random fluctuations can coincide at the positions of real galaxies with non-Gaussian noise, as is present in these images. For this criterion, we include F606W and F814W for all redshifts considered here. We add F115W for $\mathcal{S}_z = 11$–12 and F150W for $\mathcal{S}_z = 13$–17. These redshift values were chosen to ensure that the Lyα break is fully redward of the $\Delta z = 1$ range for a given filter.
4. $\int \mathcal{P}(z > 7) \geqslant 0.7$. This requires less than 30% of the integrated $\mathcal{P}(z)$ to be at $z < 7$.
5. Best-fitting photometric redshift $z_{\rm best} > 8$ (defined as "za" with EAZY; the redshift corresponding to the highest likelihood) with a goodness-of-fit $\chi^2 < 60$.
6. $\mathcal{S}_z \geqslant 9$ (selecting a sample of galaxies at $z \gtrsim 8.5$).
7. $\Delta\chi^2 > 4$, calculated as the difference between the best-fitting χ^2 value for the low-redshift restricted model to the fiducial model. By requiring a value greater than 4, we require the low-redshift model to be ruled out at $\geqslant 2\sigma$ significance. We note that Harikane et al. (2023) required a more conservative $\Delta\chi^2 > 9$, based on results from the CEERS simulated imaging. We explore the impacts of such a cut in Section 7.2.

4.3. Sample Vetting

This automated selection resulted in a sample of 64 candidate galaxies with $z \gtrsim 8.5$. To explore the impact of potential contamination by spurious (e.g., nonastrophysical) sources, we visually inspected each object in all images to search for obvious diffraction spikes, sources on image edges (as the dithering is different in different filters, the edges do not necessarily line up between different images), and unflagged cosmic rays.

To perform this inspection, three authors (S.F., J.K., and P.A.H.) viewed "bio" plots that showed small 1.″5 cutouts of the candidate in all images with two different stretches, a 5″ cutout in the F200W and detection images, the fiducial and $z < 7$ restricted $\mathcal{P}(z)$, and the spectral energy distribution (SED) plotted against both fiducial and low-z EAZY models. From this process, we identified 36 spurious sources: six diffraction spikes, 11 unflagged cosmic rays, and 19 sources on image edges. We remove these 36 sources from our galaxy sample, show images of all of them in Figure 20 in Appendix C, and list their positions in Table 6. The cosmic rays were predominantly an issue in the LW image in pointings 3 and 6; these have longer exposure times than pointings 1 and 2, amplifying the impact of cosmic rays. We note that future versions of our images and catalogs should be able to minimize these types of spurious sources by improving the cosmic-ray removal and using the JWST data model "Context map" to identify pixels close to the image edge.

Upon visual inspection, we noticed that a small fraction of the sample (four objects) had Kron apertures that were drawn too large due to the presence of bright, nearby galaxies. To explore the impact of potential light from neighboring galaxies affecting the colors, we performed an additional run of EAZY using the colors measured in 0.″3 diameter apertures to more accurately measure the colors of the candidate galaxies. For this run, we used the PSF-matched photometry but did not apply any aperture correction, as the photometric redshift depends only on colors and not luminosity. We found that two sources, CEERS3_1537 (initial $z_{\rm phot} = 10.5$) and CEERS6_7478 (initial $z_{\rm phot} = 8.6$), had their photometric redshifts shift to lower redshifts with these smaller apertures; these were the same two sources where the Kron aperture was most significantly stretched. We thus remove these two objects from our sample, though we show their SEDs and image stamps in Figure 19 in the Appendix.

The remaining two sources (CEERS1_3908 and CEERS1_3910) had no significant change in their photometric redshift when measured with these small apertures. For these two objects, we correct their fluxes by a single factor to account for potential excess brightness from the neighbor. We do this by deriving a flux correction as the ratio between the median total flux for objects with similar 0.″3 diameter aperture fluxes as a given source and the actual Kron-based (aperture-corrected) total flux for those comparison sources and dividing this quantity by the same quantity for the sources in question. We apply this correction to all filters such that the colors do not change. For CEERS1_3908, this correction factor was 0.6; for CEERS1_3910, it was close to unity, so we applied no correction.

5. An Early Redshift >9 JWST Galaxy Sample

Our final sample consists of 26 galaxies. We divide them into redshift bins for analysis. The high redshifts now accessible with JWST coupled with the broad filter transmission functions make it harder to place galaxies in often-used $\Delta z = 1$ redshift bins, and such bins span progressively smaller epochs of cosmic time (for example, a $\Delta z = 1$ bin centered at $z = 12$ would cover the Lyα break over only ∼40 Myr of cosmic time, compared to 200 Myr for one centered at $z = 6$). We thus split our sample into three redshift bins: a $z \sim 9$ sample from $z = 8.5$ to 10, a $z \sim 11$ sample from $z = 10$ to 12, and a $z > 12$ sample from $z = 12$ to 17. We place candidate galaxies in the sample bin based on where their integrated $\mathcal{P}(z)$ across the bin is the largest. These three bins cover roughly similar ranges of cosmic time (116, 105, and 122 Myr, respectively). We show the distribution of the photometric redshifts and apparent magnitudes of our sample in Figure 3, and we tabulate the sample in Table 2. We show cutout images of all candidates in Figures 5, 6, 16, and 17 and SEDs in Figures 5, 7, and 18. We also present a rest-UV color montage of our sample in descending redshift order in Figure 4.

As a check on our fiducial EAZY photometric redshifts, we also measured photometric redshifts with Cigale (Burgarella et al. 2005). We find excellent agreement, with a median difference of 0.2 (with EAZY preferring the slightly higher redshifts). Only the source CEERS1_4143 has a difference in best-fitting redshift of >2, as Cigale prefers $z = 5.4$ for this source, whereas EAZY finds $z = 9.0$, though this source lies right at the boundary of our $\Delta\chi^2$ selection criterion. Finally, we inspect the spatial distribution of our full sample of galaxies across the full CEERS field and do not visually see any evidence of strong clustering (which would not be expected with such a small sample covering a broad redshift range).

5.1. The z > 12 Sample

This highest-redshift sample contains two galaxies, CEERS2_2159 and CEERS1_1730, with 68% confidence ranges on their photometric redshifts of 16.0–16.6 and 12.3–14.2 and F277W magnitudes of 26.5 and 27.7, respectively. We will compare our sample in detail to those previously presented in the literature in Section 6 below, but we note here that CEERS2_2159 was first presented as a robust ultrahigh-redshift candidate in D23, with some caveats discussed in Zavala et al. (2023; a marginal 2.6σ dust emission detection) and Naidu et al. (2022a; environmental evidence). CEERS1_1730 is presented here for the first time.

The cutout images for these two objects are shown in the top panel of Figure 5. It is apparent that there is no significant flux at the positions of both candidates in the stacked ACS F606W +F814W images or the F115W image. For CEERS2_2159, no flux is also seen in F150W, and the F200W flux is noticeably fainter than in F277W, consistent with its photometric redshift of $z \sim 16$–16.6 (where the Lyα break would be in the red half of the F200W bandpass). For CEERS1_1730, faint flux is seen in F150W. However, this is measured at just 1σ significance in the Kron aperture. This leads to a fairly broad $\mathcal{P}(z)$, spanning $z \sim 12$–14. If this flux is real, then $z \sim 12$ is more likely. If it is not significant, then $z \sim 14$ is possible. The SEDs, shown in the bottom panel of Figure 5, show the strong observed Lyα break in both objects, with fairly blue UV spectral slopes redward of the break. For both objects, the high-redshift model is a

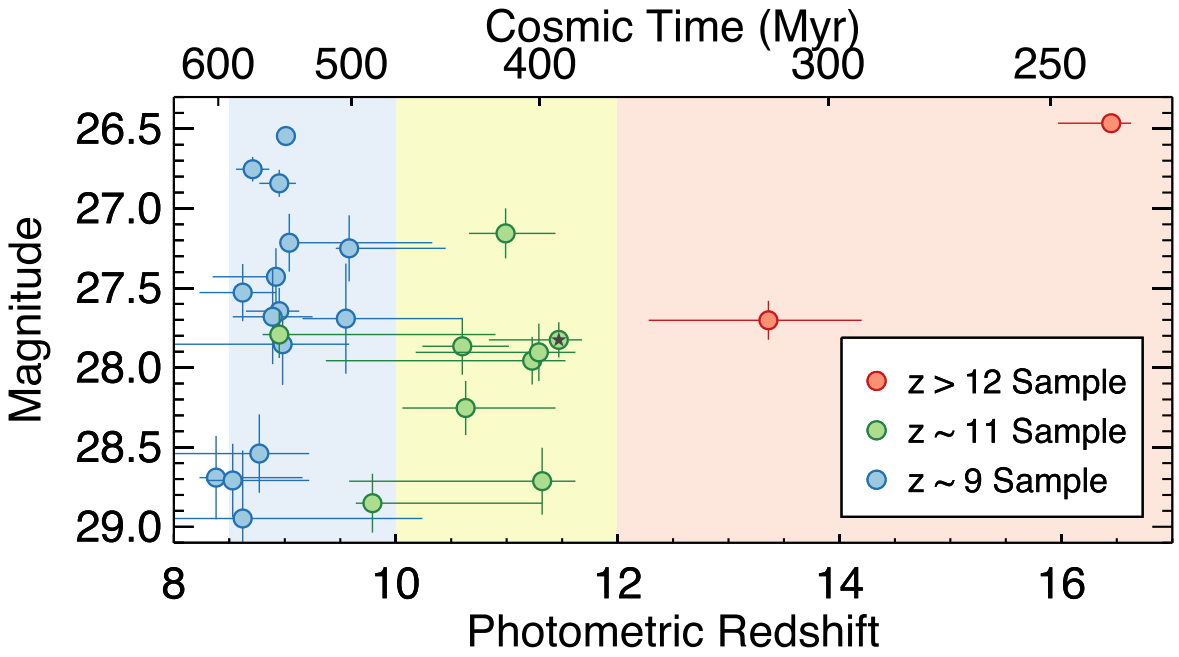

Figure 3. Distribution of the best-fitting (minimized χ^2) photometric redshift vs. apparent magnitude. Each object is color-coded by its redshift sample (with the background shading denoting the redshift range where the $\mathcal{P}(z)$ was integrated to determine the sample placement; the green circle far into the blue region [CEERS6_7641] has a double-peaked $\mathcal{P}(z)$). The magnitude plotted is F150W for $z\sim9$, F200W for $z\sim11$, and F277W for $z>12$ galaxies. The small black star denotes Maisie's Galaxy (Finkelstein et al. 2022b), known here as CEERS2_5429. The top axis shows the cosmic time (the time since the Big Bang) for our adopted cosmology.

significantly better fit to the photometry than the best-fitting low-redshift model, though for CEERS1_1730, $\Delta\chi^2$ is only 4.4; this low-redshift solution is visible as a small peak in the fiducial $\mathcal{P}(z)$. CEERS2_2159 shows no significant lower-redshift peak in its $\mathcal{P}(z)$ due to its brighter magnitude, leading to more robust constraints on the full shape of the SED.

5.2. The z ~ 11 Sample

The $z\sim11$ sample contains nine galaxies. As shown by their cutout images in Figure 6, all show no significant flux in the ACS stack and F115W images, consistent with $z>9.5$. Many exhibit a red F150W−F200W color, suggesting that the Lyα break is in F150W. The SEDs and $\mathcal{P}(z)$s for all nine are shown in Figure 7. First looking at the $\mathcal{P}(z)$s in the upper left of each panel, the amplitude of the detected Lyα break is strong enough to either eliminate or leave a very small low-redshift solution. As shown in Table 2, the integrated $\mathcal{P}(z>7)$ is $\geqslant$0.98 for eight out of nine of these sources, with the remaining source (CEERS1_7227) having a value of 0.85, significantly greater than our sample limit of 0.7 (this object also just barely satisfies our $\Delta\chi^2$ criterion, as only the F200W flux is discrepant with the best-fitting lower-redshift model, and it has a 2.2σ significance detection in F115W, which could indicate a redshift closer to $z\sim9$).

CEERS6_7641 shows a double-peaked high-redshift $\mathcal{P}(z)$ with peaks at $z\sim9$ and 11. As its integrated $\mathcal{P}(z)$ at $z=10$–12 is larger than at $z=8.5$–10, this object was placed in the $z\sim11$ sample (this is the green $z\sim11$ symbol in Figure 3 that is in the $z\sim9$ sample region), though clearly, it has a near-equal probability of being at a slightly lower redshift. Finally, we note that, similar to the $z>12$ galaxies, the UV spectral slopes for these nine sources all appear fairly blue. Though a detailed analysis of this quantity is beyond the scope of this paper, it is clear that these objects all appear fairly low in dust attenuation. We acknowledge that Lyα break selection can be biased against red sources (e.g., Dunlop et al. 2012); thus, a full quantitative analysis of the colors of these galaxies is reserved for future work.

5.3. The z ~ 9 Sample

The remaining 15 candidate high-redshift galaxies fall into our $z\sim9$ sample. We show the cutout images in Figures 16 and 17 in the Appendix. At $z<9.5$, the Lyα break is in the F115W band; thus, we expect to see signal in that image for all but the highest-redshift sources in this subsample, though the F115W flux should be fainter than the F150W flux, which is apparent in the images. Likewise, we see no significant flux in the ACS filters, as expected. Examining their SEDs and $\mathcal{P}(z)$s in Figure 18 in the Appendix, we see that nearly all show a tight high-redshift peak with very little probability of being at $z<7$. Similar to the $z\sim11$ sample, the lowest integrated $\mathcal{P}(z>7)$ is 0.84, with 13 out of 15 galaxies having integrated $\mathcal{P}(z>7)>.9$ and eight out of 15 having integrated $\mathcal{P}(z>7)\geqslant0.99$.

The object with the most complex $\mathcal{P}(z)$ constraints is CEERS2_2274, which shows small peaks at $z\sim2$ and 6, a dominant peak at $z\sim9$, and a modest peak at $z\sim10.5$, with this uncertainty due to the lower S/N on this object's faint fluxes. While the majority of these $z\sim9$ galaxies show blue rest-UV spectral slopes, the lower redshift means that the reddest filters probe the 4000 Å breaks, and it is clear that some objects have redder colors in these reddest bands. This may be due to significant stellar mass in (somewhat) older stellar populations (e.g., Labbe et al. 2022), though it could also be due to the presence of very strong rest-frame optical nebular emission lines, which early JWST results indicate are ubiquitous in high-redshift star-forming galaxies (Endsley et al. 2022; C. Papovich et al. 2022, in preparation).

Table 2
Summary of $z > 9$ Candidate Galaxies

ID	R.A. (J2000)	Decl. (J2000)	m_{F277W} (mag)	$\int_7^{20} \mathcal{P}(z)$	Sample $\mathcal{S}_z$	Photometric Redshift	M_{1500} (mag)	R_h (pixels)	$\Delta\chi^2$
CEERS2_2159	214.914548	52.943023	26.5	1.00	>12	$16.45^{+0.18}_{-0.48}$	$-21.7^{+0.1}_{-0.1}$	4.4 ± 6.0	4.4
CEERS1_1730	215.010022	53.013641	27.7	0.97	>12	$13.36^{+0.84}_{-1.08}$	$-20.4^{+0.2}_{-0.2}$	5.3 ± 1.0	15.0
CEERS2_588	214.906633	52.945504	26.9	0.98	11	$10.99^{+0.45}_{-0.33}$	$-20.6^{+0.2}_{-0.2}$	4.0[a]	7.4
CEERS1_8817	215.043999	52.994302	28.1	1.00	11	$10.60^{+0.42}_{-0.36}$	$-19.7^{+0.2}_{-0.2}$	2.4 ± 0.7	12.1
CEERS2_7929	214.922783	52.911528	28.1	0.99	11	$11.29^{+0.33}_{-1.11}$	$-19.8^{+0.4}_{-0.1}$	4.1 ± 1.0	9.1
CEERS6_7641	214.897232	52.843854	28.1	0.99	11	$8.95^{+1.95}_{-0.15}$	$-19.3^{+0.1}_{-0.5}$	2.0 ± 0.6	8.8
CEERS2_5429[b]	214.943148	52.942442	28.3	1.00	11	$11.47^{+0.21}_{-0.63}$	$-19.9^{+0.2}_{-0.0}$	4.6 ± 1.2	25.8
CEERS1_7227	215.037504	52.999394	28.3	0.85	11	$11.23^{+0.30}_{-1.86}$	$-19.7^{+0.3}_{-0.1}$	6.4 ± 2.6	4.0
CEERS6_7603	214.901252	52.846997	28.9	0.98	11	$11.32^{+0.30}_{-1.74}$	$-18.8^{+0.4}_{-0.2}$	2.1 ± 0.8	7.1
CEERS6_4407	214.869661	52.843646	29.0	0.99	11	$10.63^{+0.81}_{-0.57}$	$-19.1^{+0.2}_{-0.3}$	4.2 ± 1.2	9.3
CEERS6_8056	214.850129	52.808052	29.1	0.99	11	$9.79^{+1.53}_{-0.15}$	$-18.4^{+0.1}_{-0.4}$	2.0 ± 0.7	7.2
CEERS2_2402	214.844766	52.892104	26.7	1.00	9	$8.71^{+0.15}_{-0.15}$	$-20.5^{+0.1}_{-0.0}$	4.9 ± 0.7	63.9
CEERS1_6059	215.011706	52.988303	27.0	1.00	9	$9.01^{+0.06}_{-0.06}$	$-20.6^{+0.0}_{-0.1}$	2.6 ± 0.4	48.0
CEERS1_1875	214.951936	52.971742	27.1	0.92	9	$8.92^{+0.06}_{-0.57}$	$-19.9^{+0.3}_{-0.1}$	8.4 ± 1.3	5.4
CEERS1_3858	214.994402	52.989379	27.2	0.99	9	$8.95^{+0.15}_{-0.18}$	$-20.4^{+0.1}_{-0.1}$	2.1 ± 0.5	7.6
CEERS2_7534	214.876144	52.880826	27.4	1.00	9	$8.95^{+0.18}_{-0.30}$	$-19.5^{+0.2}_{-0.1}$	1.4 ± 0.3	39.6
CEERS1_3908	215.005189	52.996580	27.3	0.96	9	$9.04^{+1.29}_{-0.06}$	$-20.1^{+0.1}_{-0.3}$	10.0*	5.3
CEERS6_4012	214.888127	52.858987	27.6	0.92	9	$8.89^{+0.36}_{-0.36}$	$-20.0^{+0.2}_{-0.2}$	5.5 ± 2.0	4.8
CEERS2_2324	214.861602	52.904603	27.6	1.00	9	$9.58^{+0.87}_{-0.12}$	$-20.1^{+0.2}_{-0.2}$	2.0 ± 0.8	10.7
CEERS1_3910	215.005365	52.996697	28.0	0.96	9	$9.55^{+1.05}_{-0.39}$	$-19.5^{+0.5}_{-0.2}$	8.6 ± 1.4	5.7
CEERS1_5534	214.950078	52.949267	27.9	1.00	9	$8.62^{+0.30}_{-0.39}$	$-19.5^{+0.2}_{-0.2}$	3.6 ± 0.7	12.6
CEERS1_4143	214.966717	52.968286	28.1	0.84	9	$8.98^{+0.60}_{-1.23}$	$-19.5^{+0.3}_{-0.2}$	4.8 ± 1.5	4.0
CEERS3_1748	214.830685	52.887771	28.5	0.93	9	$8.77^{+0.45}_{-1.08}$	$-18.3^{+1.0}_{-0.2}$	1.8 ± 0.7	4.2
CEERS2_1298	214.902236	52.939370	28.6	1.00	9	$8.53^{+0.69}_{-0.24}$	$-18.4^{+0.2}_{-0.4}$	2.2 ± 0.8	12.7
CEERS2_2274	214.846173	52.894001	29.1	0.85	9	$8.62^{+1.62}_{-1.41}$	$-18.6^{+0.3}_{-0.4}$	1.8 ± 0.8	4.2
CEERS2_1075	214.907627	52.944611	29.1	1.00	9	$8.38^{+0.78}_{-0.15}$	$-18.4^{+0.2}_{-0.3}$	1.6 ± 0.7	12.7

Notes. The horizontal lines divide our three redshift samples (given by the sixth column). The photometric redshift is "za" from EAZY, which is the redshift where the χ^2 is minimized. The $\Delta\chi^2$ in the final column compares the best-fitting low-redshift ($0.5 < z < 7$) model to the best-fitting high-redshift model; a value of $\geqslant 4$ was required for selection.

[a] GALFIT did not converge for these two sources, so we list their SE half-light radii.

[b] Previously published as Maisie's Galaxy by Finkelstein et al. (2022b). The half-light radii are listed in pixels; our pixel scale is 30 mas.

Finally, we discuss the interesting pair of objects CEERS1_3908 and CEERS1_3910. As discussed in Section 4.3, the presence of a bright neighbor skewed their Kron apertures, but even after correcting for this, they appear as robust $z \sim 9$ candidates. In inspecting their cutout images, it is apparent that these two candidate galaxies are very close together, with each consisting of a small knot of emission with centroids $0\farcs6$ ($\sim$2.5 kpc) apart. It is possible that these are two star-forming regions of the same host galaxy. This hypothesis could be supported by the apparent very faint emission between the two clumps. Deeper imaging could resolve this, though as they are resolved in our catalog, we do not merge them here.

As one additional check, we explored the impact of our photometric correction factors (Section 3.6), which we had applied to our catalog. We reran EAZY on our final sample of 26 high-redshift galaxy candidates removing this factor. We find that this makes effectively no change in the high-redshift solution, with the best-fitting redshifts unchanged to more than the 1%–2% level and all candidates continuing to satisfy our $\mathcal{P}(z)$ selection criteria. We did notice a slight change in the best-fitting low-redshift model, leading to the median value of $\Delta\chi^2$ being reduced by 5%. This impacts sources that were close to our $\Delta\chi^2$ threshold, with CEERS1_1730 ($\Delta\chi^2 = 4.4 \rightarrow 3.5$), CEERS1_4143 ($\Delta\chi^2 = 4.0 \rightarrow 3.7$), and CEERS1_7227 ($\Delta\chi^2 = 4.0 \rightarrow 3.7$) falling below our threshold of $\Delta\chi^2 > 4$. While these sources remain in the sample following our fiducial selection, their presence near this threshold leaves their inclusion more sensitive to these small correction factors.

5.4. Galaxy Sizes

We derive the sizes of all 26 galaxies in the sample by performing parametric fits on the F200W NIRCam images using GALFIT[61] (Peng et al. 2002, 2010). GALFIT finds the optimum Sérsic fit to a galaxy's light profile using a least-squares fitting algorithm. As input, we use a 100 × 100 pixel cutout of the F200W image for each source, the corresponding error array (the ERR extension) as the input sigma image, and the empirically derived PSFs. We use the source location, magnitude, size, position angle, and axis ratios from the SE catalog as initial guesses. We allow the Sérsic index to vary between 0.01 and 8, the magnitude of the galaxy between zero and 45, the half-light radius (r_h) between 0.3 and 100 pixels, and the axis ratio between 0.0001 and 1. We also allow GALFIT to oversample the PSF by a factor of 9. We then visually

[61] https://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html

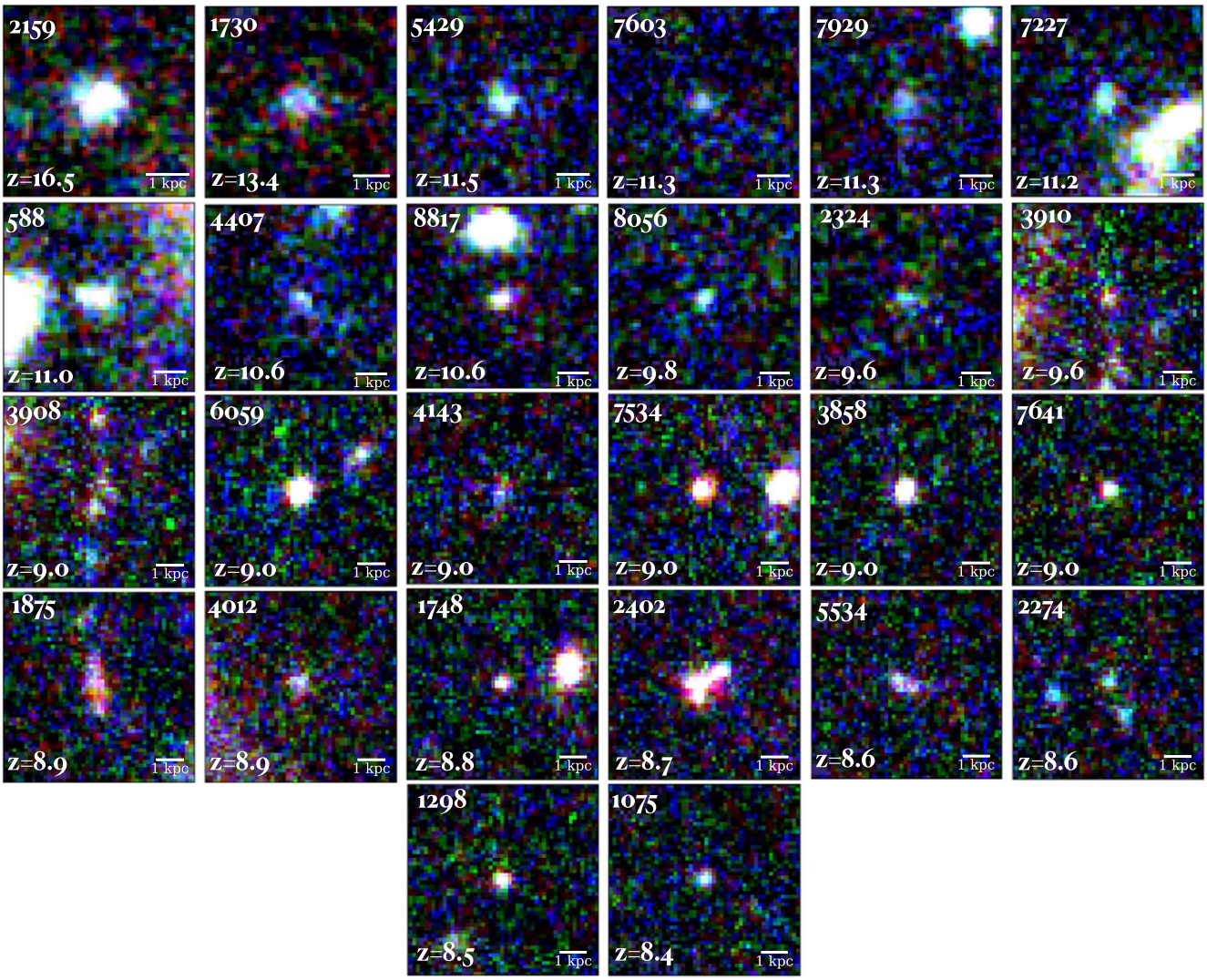

Figure 4. Montage of images for the galaxy candidates identified here. For each galaxy, the image shows as blue–green–red the three NIRCam bandpasses closest to the Lyα break, which includes F150W, F200W, and F277W for galaxies at $z = 9$; F200W, F277W, and F356W for galaxies in the $z = 10$–12 samples; and F277W, F356W, and F444W for galaxies in the $z \geqslant 12$ samples. In each case, the images are $1\farcs5 \times 1\farcs5$. The inset scale bar shows 1 (physical) kpc. We list the numerical IDs and best-fitting photometric redshift values (the redshift uncertainties are listed in Table 2).

inspected the best-fit model and image residual for each source to ensure that the fits were reasonable and that minimal flux remained in the residual. We also performed a fit on the PSF image itself in order to determine the smallest resolvable size. This value is 1.18 ± 0.01 pixels. Two sources (CEERS1_3908, a member of the pair discussed above, and CEERS2_588) failed to converge on a fit.

In Figure 8, we show these measured sizes, highlighting that the majority of the sample is significantly spatially resolved. The measured r_h values range from 0.41 to 8.59 pixels, with a median value of 3.6 pixels ($0\farcs11$ or 0.46 kpc). These sizes are consistent with the rest-UV sizes found in the GLASS survey by Yang et al. (2022), who found a median half-light radius of 0.45 ± 0.13 kpc for galaxies at $7 < z < 15$.

5.5. Stellar Contamination Screening

It is important to analyze whether any objects in our sample could potentially be low-mass stars or brown dwarfs, as they can have similar colors as high-redshift galaxies when observed in broadband filters (e.g., Yan et al. 2003; Ryan et al. 2005; Caballero et al. 2008; Wilkins et al. 2014). We first analyze the GALFIT-produced half-light radii of our candidates, as any resolved objects cannot be stellar in origin. Figure 8 shows the GALFIT-measured radii for the 24 out of 26 of our candidate high-redshift galaxies where GALFIT converged (showing the SE values for the remaining two). We find that the majority of the sample is obviously resolved, with only five of the faintest sources having sizes within the 1σ error bars that allow for a point source.

For these five unresolved sources, we then follow the methodology of Finkelstein et al. (2022a, 2022b) to explore whether the colors of any of the few unresolved galaxy candidates could be consistent with stars. We derive a grid of models for the colors of low-mass stars and brown dwarfs (spectral types of M4–T8) in the NIRCam filters by integrating the IRTF SpEX brown dwarf templates (Burgasser 2014). As these spectra end at 2.5 μm, we use the tabulated 2MASS photometry to link each SpeX model with Spitzer/IRAC photometry from Patten et al. (2006). Following Finkelstein et al. (2022b), we assume we can map IRAC 3.6 μm onto F356W and 4.5 μm onto F444W, though future spectroscopic

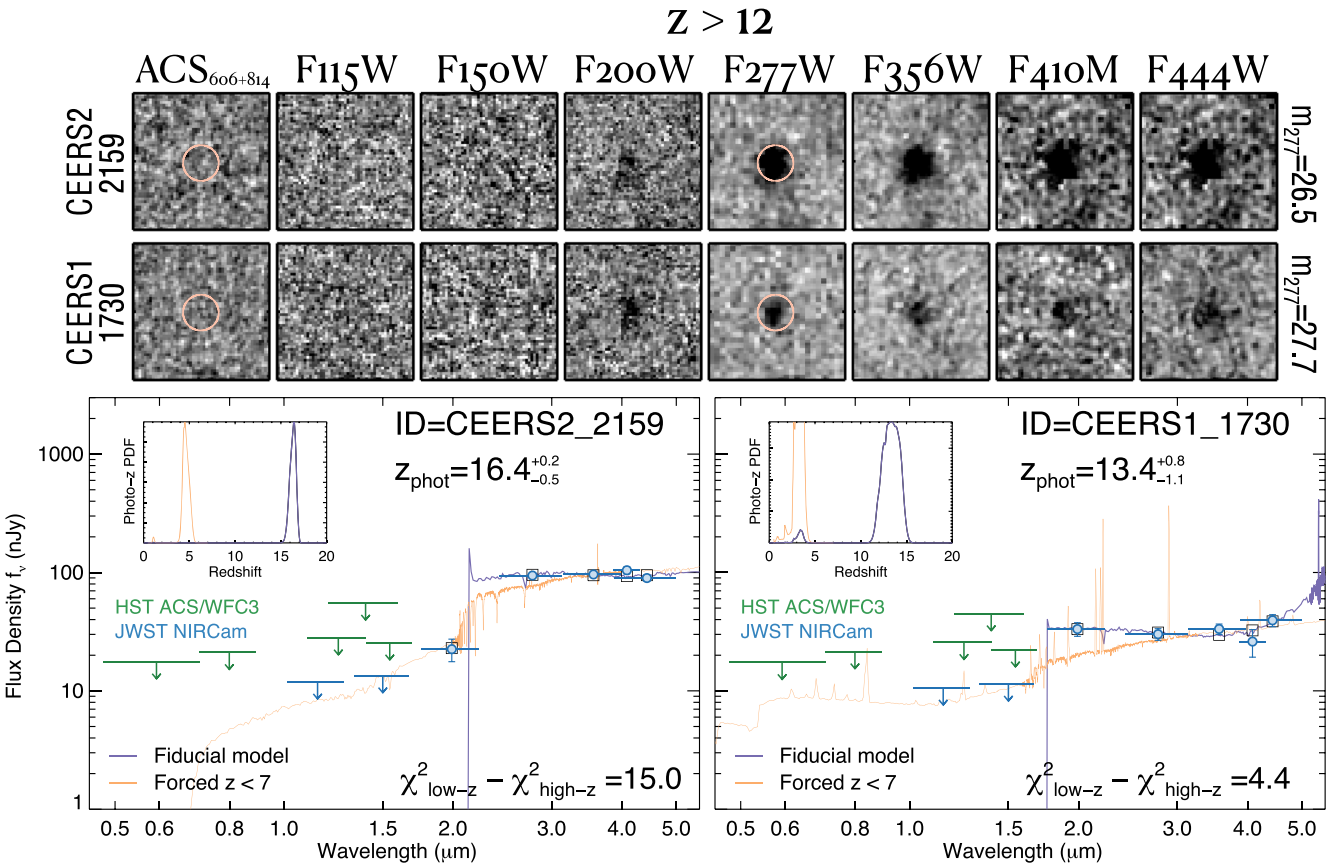

Figure 5. Top: stamp images, 1.″5 on a side, of the two galaxies that are best fit with $z > 12$ (shown from the non-PSF-matched images). The red circle denotes a 0.″4 diameter region around the source (which we show in only two bands for clarity). CEERS1_1730 has a best-fit photo-z of 13.4, albeit with a wide 68% confidence range of 12.3–14.2. CEERS2_2159 is best fit by $z = 16.5$ (16.0–16.6) and was first identified by D23. The bottom panels show the best-fitting EAZY models (both overall and constrained to $z < 7$), the model bandpass fluxes (open squares), and the observed photometry (circles; upper limits are 2σ). The inset panels show the $\mathcal{P}(z)$ distributions. Both sources exhibit well-constrained Lyα breaks, implying that the redshifts are $z > 11$. CEERS1_1730 does show a small low-redshift solution, and its primary $\mathcal{P}(z)$ peak extends to $z \sim 10.5$.

observations of brown dwarfs with JWST at $\lambda \gtrsim 2.5\,\mu$m will improve this methodology. To explore which model is preferred, we use the Bayesian information criterion (BIC), which includes the goodness of fit (χ^2), the number of photometric constraints (five for stars and 12 for galaxies due to the use of HST for the latter), and the number of free parameters (one for stars and 19 for galaxies, with 18 templates fit simultaneously, plus the redshift). We find that two of these five sources have BIC values that indicate a preference for the best-fitting stellar model.

Although these two objects (CEERS2_2274 and CEERS2_1075) are formally unresolved and have optical/near-IR colors consistent with brown dwarfs (Patten et al. 2006), we nevertheless doubt this interpretation, as they are very faint ($m_{\rm F277W} = 29.1$ mag). The colors imply a very late spectral type of $\gtrsim$T5, which corresponds to an absolute magnitude of $K = 16.1$ AB mag (Patten et al. 2006). Such a brown dwarf would therefore be at a heliocentric distance of 3.6 kpc, or 3.1 kpc above the Galactic plane, which would place this object approximately nine scale heights out of the thin disk (e.g., Ryan et al. 2011; Holwerda et al. 2014). These objects are therefore highly unlikely to be part of the Population I thin disk stars but could be members of the thick disk or Galactic halo (Ryan & Reid 2016). At present, there are very few constraints on the density of brown dwarfs in these more distant Galactic components at these extremely low effective temperatures, but based on more massive main-sequence stars, these are expected to have many orders of magnitude fewer stars than the thin disk (e.g., Jurić et al. 2008). A rigorous Bayesian model of the halo brown dwarfs would provide the strongest statistical evidence and prediction for or against this source, but this is beyond the scope of the present work. We also note that CEERS2_2274 appears to have a very nearby neighbor with a similar SED, which could indicate that it is a merging system at high redshift. We note that both sources are below the brightness limit ($m < 28.5$) we apply when analyzing our sample in Section 7, so their inclusion (or exclusion) does not affect any of our conclusions.

5.6. Examining Ancillary Multiwavelength Observations

We have examined the positions of our 26 candidate $z \gtrsim 9$ galaxies at both shorter and longer wavelengths and find no significant detections, increasing confidence in the very high redshift interpretation. While the depth of these data is not necessarily sufficient to completely rule out low-redshift solutions, these nondetections do increase our confidence in the fidelity of our sample. First, we searched for X-ray emission coincident with our candidate positions using Chandra imaging from the AEGIS-XD survey (Nandra et al. 2015), which has a flux limit of 1.5×10^{-16} erg cm^{-2} s^{-1} in the 0.5–10 keV band. There was no emission detected, with a Poisson false

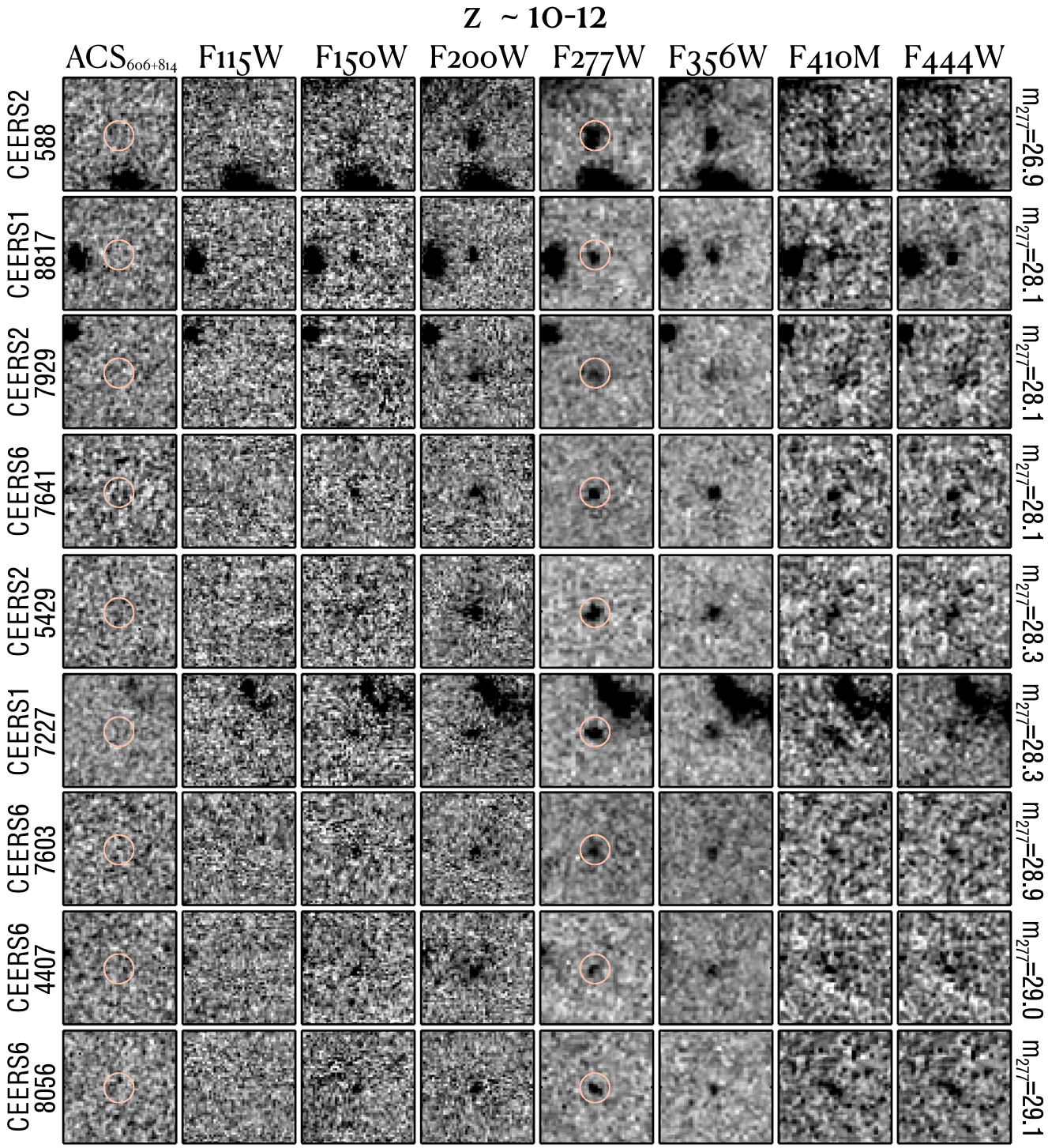

Figure 6. Similar to Figure 5 for the nine galaxies in the $z \sim 11$ sample.

probability of less than 4×10^{-6} in the soft (0.5–2 keV), hard (2–7 keV), full (0.5–7 keV), and ultrahard (4–7 keV) energy bands.

We then investigated possible far-infrared emission at the position of our high-redshift galaxy candidates using the superdeblending catalog technique from Jin et al. (2018) and Liu et al. (2018), which was adapted for the EGS field. To summarize, this multiwavelength fitting technique is meant to optimize the number of priors fitted at each band to extract the deepest possible information. For this multiwavelength fitting, we use data from Spitzer (24 μm from FIDEL; Dickinson & FIDEL Team 2007), Herschel 100 μm and 160 μm from PEP (Lutz et al. 2011) and 250 μm, 350 μm, 500 μm from HerMES (Oliver et al. 2012), JCMT/SCUBA-2, including 850 μm from S2CLS (Geach et al. 2017) and 450 μm and 850 μm from Zavala et al. (2017) and AzTEC 1.1 mm from Aretxaga (2015). The key is to obtain an adaptive balance as a function of wavelength/data set between the number of priors fitted, the quality of the fit, and the achievable deblending given the PSF sizes. We start

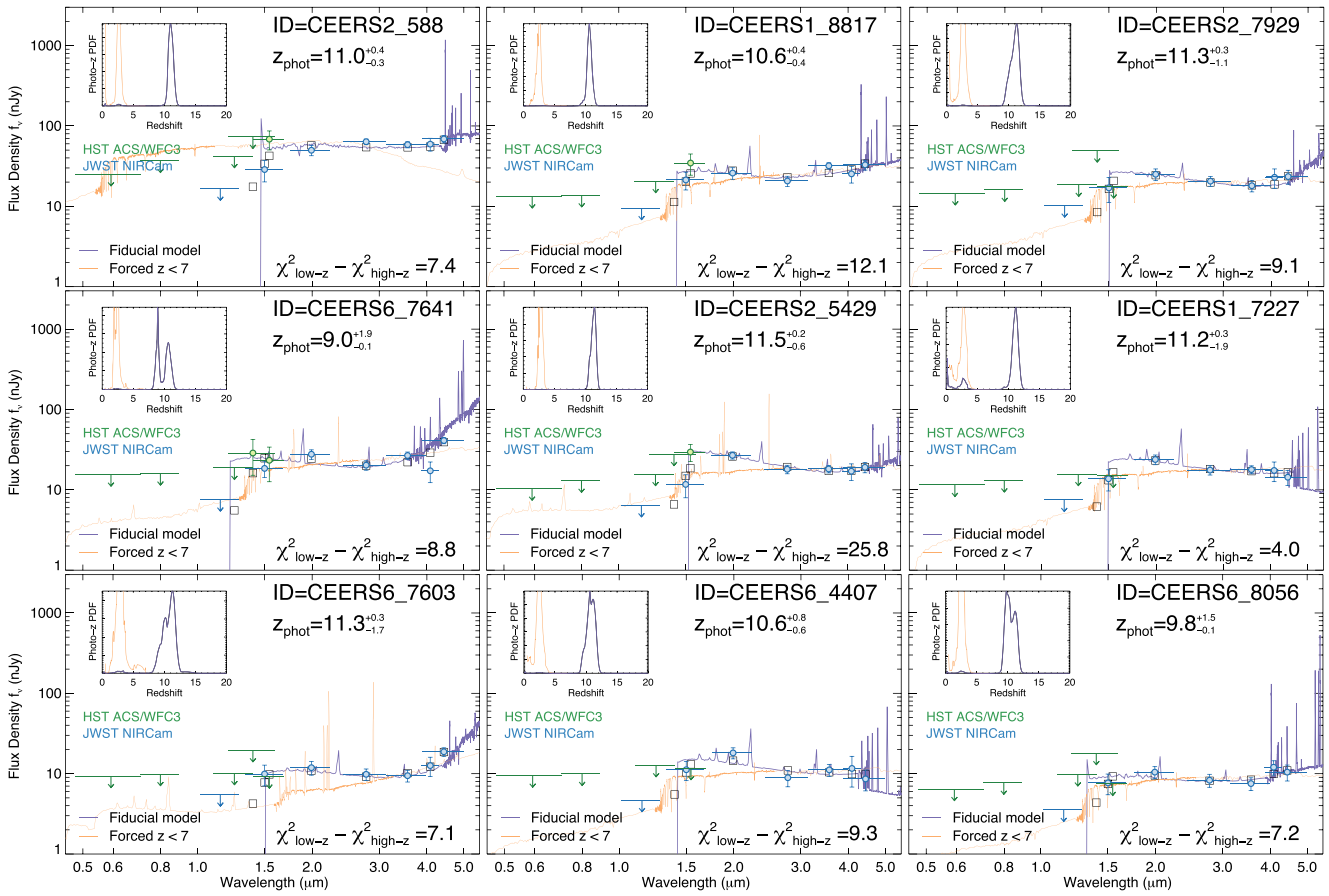

Figure 7. The SEDs and $\mathcal{P}(z)$s for the $z \sim 10$–12 sample, with lines and symbols the same as in Figure 5.

with the deepest images and fit each band from deeper to shallower images.

In general, the high-redshift galaxy candidates are not already contained in our prior lists, with the exception of two sources, for which all measurements are already contained in the forthcoming catalog (A. Le Bail et al. 2022, in preparation). For the other candidates, we consider them one at a time and add a specific prior at its position that we fit together with the rest of the priors that are relevant for each band. Extensive Monte Carlo simulations ensure that the uncertainties associated with the flux measurements are "quasi-Gaussians" (see Jin et al. 2018; Liu et al. 2018; A. Le Bail et al. 2022, in preparation). None of the candidates are significantly detected at any band between 24 μm and 1.1 mm. We note that for one galaxy, CEERS1_1730, this method implies an $\sim 3\sigma$ detection at 500 μm. However, inspecting the images, it is clear at 160, 250, and 350 μm, there is significant emission from a bright neighbor to the south, which overlaps our source's position at 500 μm due to the worsening PSF. We conclude that it is highly unlikely that this emission is associated with our candidate.

We also note that our object CEERS2_2159, first published as ID = 93316 by D23, has a formal 2.6σ detection in the SCUBA-2 850 μm data, as noted by Zavala et al. (2023). However, that marginal 850 μm detection could plausibly be associated with other nearby $z \sim 5$ galaxies within the beam, as discussed in Zavala et al. (2023); higher-resolution millimeter interferometry is being carried out on this system to gain further insight (S. Fujimoto et al. 2022, in preparation).

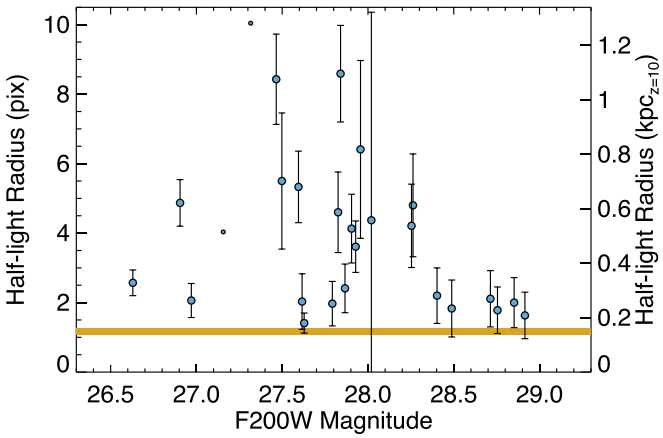

Figure 8. The F200W half-light radii of our high-redshift galaxy candidates, measured with GALFIT. The shaded region denotes the half-light radius for the F200W PSF of 1.18 ± 0.01 pixels (our pixel scale is 30 mas). While the galaxies are compact, all but the faintest appear to be resolved. Our candidate galaxies have a median half-light radius of 0.46 kpc. The small circles denote the two objects where GALFIT did not converge; thus, we show the SE half-light radius.

6. Previously Identified $z \gtrsim 9$ Candidates in CEERS

6.1. Candidates Identified with HST

The area presently covered by CEERS covers two $z \sim 9$–10 galaxies previously identified by Finkelstein et al. (2022a), EGS_z910_40898 and EGS_z910_65860 (none of the candidate $z \gtrsim 9$ galaxies from Bouwens et al. 2019 fall in this first

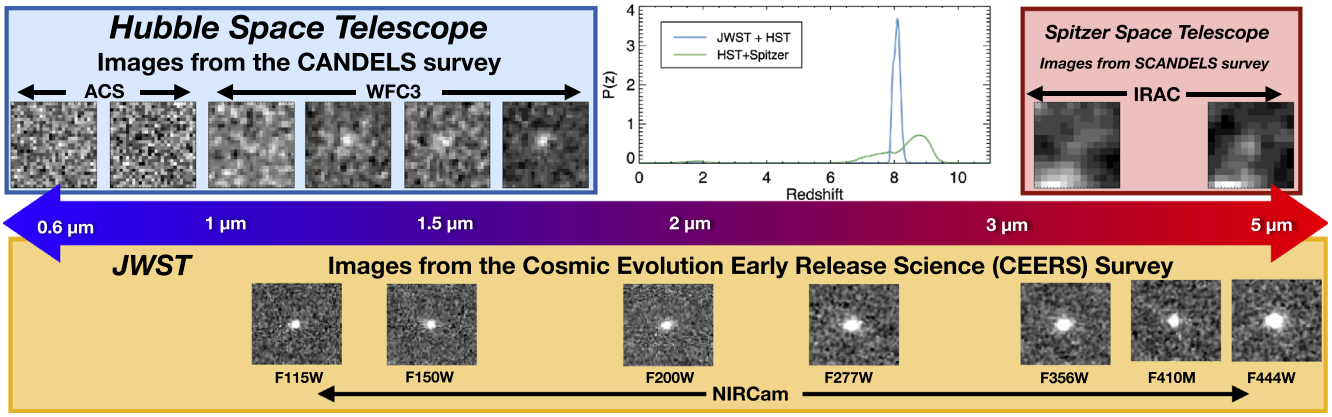

Figure 9. Candidate galaxy EGS_z910_40898, first published by Finkelstein et al. (2022a) as a $z \sim 9$ galaxy candidate. This source was detectable by HST, albeit in only three filters and at only modest significance. Even in 50 hr of Spitzer/IRAC imaging, it was marginally detected. In just <3000 s with JWST/CEERS, this source is extremely well detected in all seven NIRCam filters, highlighting the power of JWST to probe the very early universe. In the top middle panel, we compare the previous $\mathcal{P}(z)$ to that now possible with JWST, finding a much sharper peak centered at $z \sim 8.1$ (the F115W−F150W color is only marginally red, implying a Lyα break at the blue edge of F115W).

epoch of CEERS NIRCam imaging). In Figure 9, we show a comparison of the first object in NIRCam versus HST and Spitzer imaging. While the object was only detected at modest significance in the ∼few-orbit HST imaging and barely at all in the 50 hr depth Spitzer/IRAC imaging, it is extremely well detected (S/N > 30) in <3000 s in all seven NIRCam filters.

In this figure, we also compare constraints on the photometric redshift from the previous HST + Spitzer photometry of EGS_z910_40898 to the present JWST + HST photometry. While this object was previously selected as a $z \sim 9$ candidate, as the majority of the $\mathcal{P}(z)$ was at $z \gtrsim 9$, the primary peak extended down to $z < 7$, and there was a nonnegligible secondary peak at $z \sim 2$. With the significant improvement in photometric constraints, there is now a single narrow peak at $z \sim 8.1$. The best-fitting redshift is modestly lower than the previous value due to the only marginally red F115W−F150W color, implying that the Lyα break is toward the blue end of the F115W filter. Observations of this object highlight the capabilities of JWST to significantly improve constraints on photometric redshifts compared to the pre-JWST era.

The second object, EGS_z910_65860, however, shows no significant flux at all at the expected position (Figure 10). This implies that the source identified in the HST imaging was either spurious or a transient phenomenon. Finkelstein et al. (2022a) performed a detailed vetting process to remove all forms of spurious sources, including persistence; thus, this seems like an unlikely solution. We therefore consider whether the observations are consistent with a transient source. This object showed significant detections in HST F125W and F160W imaging, with no detections in F606W, F814W, F105W, or F140W. While 3D-HST F140W preimaging was shallower than the CANDELS imaging, it was curious that this object showed no detection, as its SED (anchored by F125W and F160W) should have been detectable.

We thus investigate the date of each of these observations. The F125W and F160W images were taken in 2013, with two images taken ∼50 days apart (April 2 and May 24). We made updated HST mosaics (following the procedure in Koekemoer et al. 2011) around the position of this source in each epoch separately, and we see a clear detection in both bands in both epochs. This further refutes the hypothesis that this disappearing source was spurious. Adding to the likelihood of a transient explanation is that the source is fading between the two epochs, with the ratio of fluxes from the first to second epoch being 2.5 and 2.0 for F125W and F160W, respectively (fluxes were measured with SE on each epoch separately using MAG_AUTO to approximate total fluxes). The nondetection of this source in F140W is easily explained, in the context of a transient interpretation, by the acquisition date of those images, which was 1.5 yr earlier than F125W and F160W (2011 December 2).

While the apparent Lyα break between F105W and F125W could imply that this object is a $z \sim 9$ transient, the F105W imaging was obtained much later, on 2015 April 1. Likewise, the F606W images were taken in 2004 and 2011, while the F814W images were taken in 2004, 2011, and 2013. The SCANDELS Spitzer/IRAC imaging was obtained over many years from 2003 to 2012, but the bulk was obtained between 2010 and 2012 (Ashby et al. 2015). We therefore do not have the contemporaneous photometry needed for a reliable redshift of this source. What we do know is that the host galaxy is fainter than the limit of our NIRCam imaging. Taking the limit of our deepest image in Table 2, we find that the observed flux in the first F160W epoch (23.6 nJy) is ∼150 times brighter than the 1σ upper limit of our NIRCam imaging. The compact morphology of this source, along with the lack of apparent proper motion between the two CANDELS epochs, leaves a supernova as the most likely explanation (a proper motion of less than the F160W PSF in 50 days indicates a distance of more than 1 pc, ruling out a solar system object). We show the observed fluxes/limits versus time in the left panel of Figure 10, while in the right panel, we show constraints on host galaxy absolute magnitude.

While nearly any redshift is plausible from a luminosity standpoint (from a low-redshift dwarf galaxy to a very high redshift $\sim L^*$ galaxy), the significant fading over a 50 day time period argues against a very high redshift due to time dilation of the supernova decay curve. From the observed (sparse) light curve, we can only place loose constraints on the nature of the transient, following the methods of Rodney et al. (2014). The data favor classification as a core-collapse supernova with a redshift ranging from $z \simeq 0.2$ to 1.2, with maximum likelihood at $z \simeq 1.1$ and a slight preference for Type Ib/c supernovae

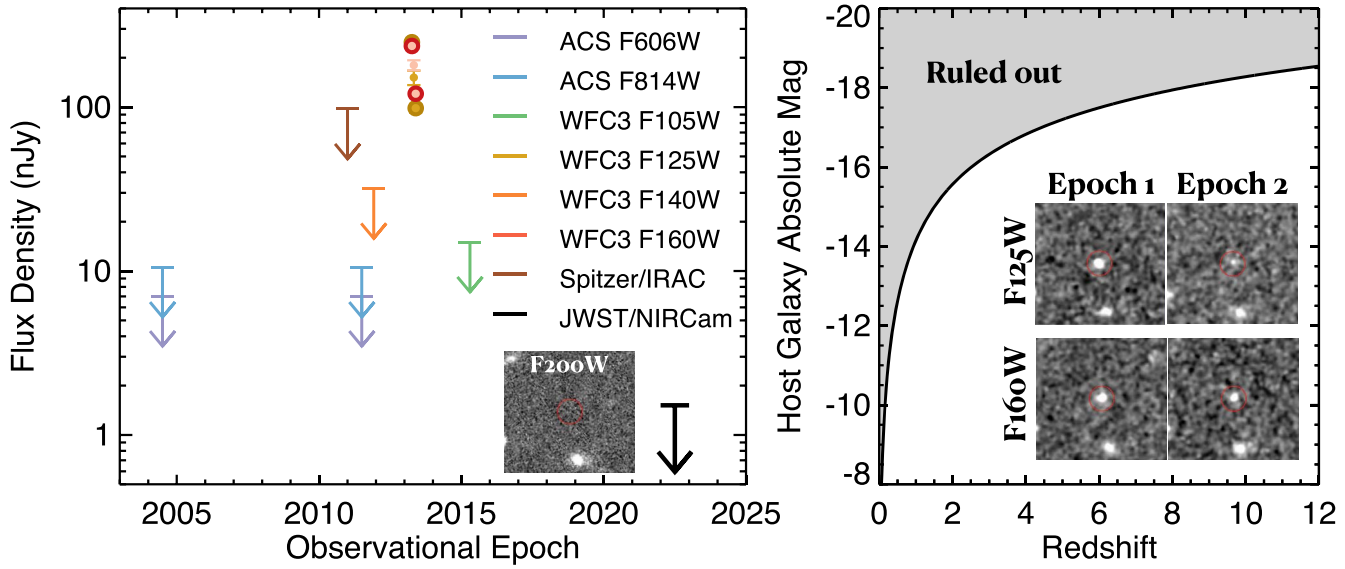

Figure 10. (Left) Flux vs. observational epoch for the object EGS_z910_68560, first published by Finkelstein et al. (2022a) as a $z \sim 9$ galaxy candidate. The colors denote the filter. The small circles denote the integrated CANDELS F125W and F160W fluxes, while the larger circles denote the fluxes measured in each epoch, 50 days apart. The object is detected in both epochs in both bands, fading by a factor of ∼2.5 and 2.0 in F125W and F160W, respectively, across this time interval. The object is not detected in any of the CEERS imaging; the inset image shows this position in the CEERS F200W image. This implies that the object is >150× fainter now than it was when it was detected. (Right) Limits on the host galaxy absolute magnitude, taken by applying the cosmological distance modulus to the CEERS upper limit. With the exception of the extremely low-redshift universe, a wide range of host galaxy absolute magnitudes are plausible, making it difficult to constrain the redshift of this supernova, though analysis of the light curve implies a likely redshift of $z \sim 0.2$–1.2. The inset panels show the images of this transient source in both CANDELS epochs in F125W and F160W.

over Type II. Nevertheless, the data are also compatible with Type Ia at $z \simeq 1.45$.

6.2. Candidates Identified with CEERS

At the time of this writing, there have been five previously submitted papers that have identified $z \gtrsim 9$ candidates from these CEERS imaging data (D23; Finkelstein et al. 2022b; Harikane et al. 2023; Labbe et al. 2022; Whitler et al. 2023). In this section, we compare our current analysis to these previous results, noting that differences in selection techniques may bias the redshift ranges that a particular study is sensitive to. We also emphasize that while the selection of a candidate by more than one study does increase the confidence in the object's fidelity, the presence or lack of a galaxy in a given sample is often easily attributable to differences in data reduction, photometric methodology, and selection criteria.

First, Finkelstein et al. (2022b) published a paper on a single galaxy, Maisie's Galaxy, which they identified as a $z \sim 12$ galaxy candidate. While they used the same full CEERS imaging data set as we use here (albeit with an earlier version of the data reduction), they employed highly conservative selection criteria to identify a single extremely robust galaxy candidate due to the nascent state of the data reduction pipeline and photometric calibration c. 2022 July. This object is in our galaxy sample, here known as CEERS2_5429. Our photometric redshift constraints for this source are $z_{\rm phot} = 11.5^{+0.2}_{-0.6}$. This is consistent at the $\sim 1\sigma$ level with the value of $z_{\rm phot} = 11.8^{+0.3}_{-0.2}$ from Finkelstein et al. (2022b), with the small differences in the redshift attributable to small changes in photometry due to the updated photometric calibration now available.

The largest previous sample of high-redshift galaxies in the CEERS field comes from D23, who selected 19 high-redshift galaxies over the CEERS field. They required dropouts in the F115W (or redder) bands to be selected, biasing the sample to $z > 9.5$ (where Lyα redshifts out of F115W). Here we compare to their revised sample (c. 2022 October). We find nine galaxies in common between our two samples. Of the 17 candidates in our sample but not in D23, 14 are at $z \lesssim 9.5$; thus, their absence from the D23 sample is expected via their requirement of no significant flux in F115W. The remaining three candidates in our sample and not in D23, CEERS1_1730, CEERS1_7227, and CEER6_4407, have $z_{\rm best} > 10$; thus, in principle, they could have been selected by D23. Of the nine sources in common, the median difference in the photometric redshifts is 0.3 ± 0.7, with all redshifts agreeing within $\Delta z < 1$ (Figure 11), with the exception of our source CEERS6_7641 (D23 ID 30585), which we find has $z_{\rm best} = 9.0$ compared to their value of $z_{\rm best} = 10.6$. However, as shown in Figure 7, the $\mathcal{P}(z)$ for this source is quite broad and contains a second peak at $z \sim 10.4$; thus, our results are fully consistent with those of D23.

Of note is that our sample contains D23 object 93316 as object CEERS2_2159, and our photometric redshift is nearly precisely equal to the D23 value of $z = 16.4$. This object is remarkably bright ($m_{\rm F277W} = 26.5$), and some evidence is present from both submillimeter imaging and environmental studies that indicate $z \sim 5$ (Naidu et al. 2022a; Zavala et al. 2023). The JWST photometry alone very strongly prefers this ultrahigh redshift. Follow-up spectroscopy and millimeter interferometry will soon reveal the true nature of this potentially record-breaking system.

We next explore the properties of the 10 galaxies selected by D23 that are not in our final high-redshift galaxy sample. Of these 10, we find that five are faint enough that they did not meet our source detection significance criteria. Some of these are faint enough in our catalog that their photometric redshifts are not well constrained, while others show peaks at $z > 10$. Four more sources show primary peaks at $z > 10$ but have

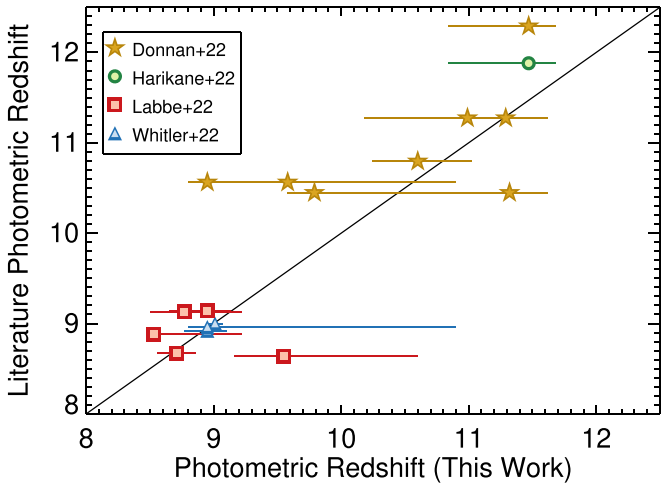

Figure 11. Comparison of our photometric redshifts to those for objects in our sample that were previously published as $z > 8.5$ galaxy candidates from CEERS data in the literature, showing good agreement (not shown is the $z \sim 16$ candidate, where the agreement is good between the three studies who have published it). As discussed in Section 6.2, for literature sources not present in our sample, we find in our catalog that they typically miss our detection significance and/or $\Delta\chi^2$ criteria, though often the $\mathcal{P}(z > 10)$ is nonzero.

secondary peaks at $z < 4$ that are large enough for our selection criteria to remove these sources (these D23 IDs 1434_2, 26409_4, 5628_2, and 6647 all have $\Delta\chi^2 < 4$ in our catalog). For only one source, D23 ID 61486, do we find a strong low-redshift solution. This object exhibits a fairly flat SED in our catalog, though it exhibits a red F115W−F150W color, which could be indicative of a true $z > 9$ Lyα break. Overall, we find strong consistency between objects in common in both our sample and that of D23, and for the D23 sources not in our sample, a high-redshift nature is possible given our photometry.

We next compare to Harikane et al. (2023). At the time of this writing, only the original sample, pre–photometric calibration update, was available for comparison. In the CEERS field, they selected galaxies as F150W or F200W dropouts, restricting their sample to $z \gtrsim 12$ and selecting six high-redshift candidates. Of these six, two are in common with our sample: CEERS2_5429 (Maisie's Galaxy; Harikane ID CR2-z12-1) and CEERS2_2159 (D23 93316; Harikane ID CR2-z17-1), with Harikane et al. having $z_{\rm best} = 11.88$ and 16.45 for these two objects, respectively. Of the four sources in their sample we do not recover, for CR2-z12-3, CR3-z12-1, and CR6-z12-1, our $\mathcal{P}(z)$ shows a primary peak at $z > 10$, but the low-redshift peak is significant, with $\Delta\chi^2 < 4$ for all three. All three sources are also faint and do not meet our detection significance criteria. For their source CR2-z12-1, we find $z < 6$, driven by an $\sim 6\sigma$ detection in F115W, and a blue F115W −F150W color.

Labbe et al. (2022) selected candidate massive galaxies at $7 < z < 11$ as those with detectable Lyα and Balmer breaks in their photometry. For this comparison, we use an updated sample made available following improvements in photometric calibration (I. Labbe 2022, private communication). Of their 13 sources, they reported photometric redshifts of >8.5 for five of them. Promisingly, all five of these sources are in our sample, with our Labbe et al. IDs of CEERS1_3910 (39575), CEERS2_1298 (21834), CEERS2_2402 (16624), CEERS2_7534 (14924), and CEERS3_1748 (35300), with the best-fitting photometric redshifts agreeing to $\Delta z < 1$ (and $\lesssim 0.3$ for 4/5).

Finally, Whitler et al. (2023) studied the stellar populations of bright galaxies at high redshift in CEERS. We compare to an updated sample, again following improved photometric calibration (L. Whitler 2022, private communication). In this sample, they have six galaxies with $z_{\rm best} > 8.5$. Three of these sources are in our sample (CEERS1_3858, CEERS1_6059, and CEERS6_7641, with Whilter et al. IDs of 37135, 37400, and 9711, respectively). The photometric redshifts for these three agree extremely well ($\Delta z < 0.05$) with our estimates. Of the three sources we do not recover, we find a strong $z \sim 10$ peak for their ID 7860; this source narrowly misses our sample with $\Delta\chi^2 = 3.7$. The ID 14506 does show a red F115W−F150W color, but our analysis prefers a 4000 Å break rather than a Lyα break, though a very small peak is present at $z \sim 11$. For ID 34362, a $z \sim 9$ peak is dominant in our measurements, but ∼50% of the integrated $\mathcal{P}(z)$ is contained in a $z \sim 2$ peak.

In Figure 11, we compare the photometric redshifts from our study to these previous works for sources in common. As discussed above, especially considering the mix of photometric procedures and photometric redshift techniques, the agreement is generally good.

As one final comparison, we compare our estimated total magnitudes with the published magnitudes for those candidates in common (where, using the information available, we compare F200W for D23 and Whitler et al. 2023, F356W for Harikane et al. 2023, and F444W for Labbe et al. 2022). We find the largest discrepancy with D23, where for the nine sources in common, the median magnitude differences between our cataloged total magnitudes and those of D23 is −0.3 mag (meaning that our fluxes are brighter); however, there is significant scatter, with one source being as much as 1 mag brighter in our catalog and another being 0.4 mag fainter in our catalog. Further exploring this discrepancy requires a deeper comparison between the procedures adopted to estimate total fluxes. Comparing to the other studies, we find results more in agreement. For the four sources in common with Whitler et al. (2023), the median magnitude offset is −0.1, with individual objects ranging from −0.2 to zero. For the two objects in common with Harikane et al. (2023), our magnitudes are fainter by 0.1 and 0.3 mag. Finally, for the five sources in common with Labbe et al. (2022), our magnitudes are all fainter by 0.05–0.2 mag, with a median of 0.15 mag. We conclude that while there is significant scatter in the photometry due to the various processes used to reduce and analyze the data, we find no evidence that our photometry is significantly systematically shifted compared to other studies, especially by the large amount needed to explain the observed luminosity function excess (Section 7.1). Clearly, improving photometric agreement between different photometric catalogs is a key goal for the near future.

7. The Evolution of Galaxies in the First 500 Myr

Here we investigate what constraints we can place on galaxy evolution with our sample of 26 candidate $z \sim 8.5$–16.5 galaxies. While it is early in the JWST mission and the sample size is modest, the fact that our sample contains a large number of galaxies at $z > 9$ allows us to investigate what constraints are possible. We first measure the rest-frame UV luminosity function at $z \sim 11$ in comparison to both previous results and expectations from empirical extrapolations. We then compare

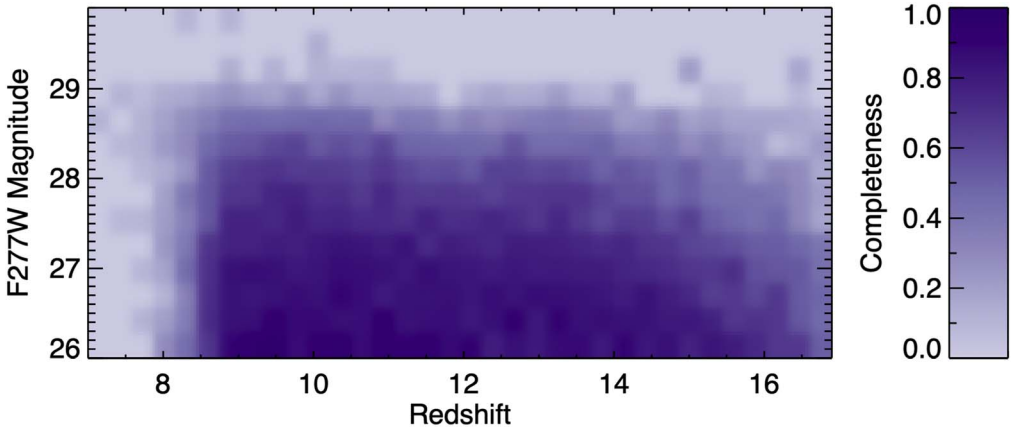

Figure 12. Total completeness as a function of input redshift and F277W apparent magnitude. These values were derived by simulations where we inject compact sources with realistic (modestly blue) SEDs into our imaging and attempt to recover them with our analysis pipeline. The shading denotes the fraction of sources in each bin of input redshift and magnitude, which were both detected by SE and satisfied our sample selection. The incompleteness at $z < 8.5$ is expected via our sample selection. We see that the CEERS imaging allows recovery of galaxies across the entire redshift range of interest to very high completeness at $m_{\rm F277W} < 28.0$, with the completeness dropping steadily from $m = 28$ to 29.

the cumulative number of galaxies to predictions from a suite of theoretical models to explore the accuracy with which these models predicted galaxy abundances in this early epoch.

7.1. The z ∼ 11 Rest-UV Luminosity Function

The UV luminosity distribution function is one of the key observational diagnostics of galaxy evolution in the early universe. With each technological leap leading to a new redshift era being observable, this quantity is always of immense interest (e.g., Bouwens et al. 2004; McLure et al. 2013; Finkelstein et al. 2015), as it is directly comparable to simulation predictions, helping to constrain the physical mechanisms regulating galaxy evolution. While the extensive amount of deep-field imaging data available from the full Cycle 1 data set will lead to excellent constraints on this quantity, here we gain a first look by exploring constraints placed by the CEERS data.

For this first-look luminosity function, we focus on the specific redshift range of $z \sim 9.5$–12. We exclude $z = 8.5$–9.5 for two reasons. First, there is a likely overdensity at $z = 8.7$ (e.g., Finkelstein et al. 2022a; Larson et al. 2022a) that would bias this quantity high. Second, at $z > 9.5$, the Lyα break is fully redward of F115W, providing a true JWST dropout sample that is not dependent on the shallower HST imaging. We choose the upper bound of $z \sim 12$, as only two galaxies in our sample have higher photometric redshifts.

To measure the UV luminosity function, we require an estimate of the effective volume over which we are sensitive to galaxies. This is a function of both redshift and source brightness and accounts for incompleteness due to both photometric measurements and sample selection. Following Finkelstein et al. (2022a), we estimate the completeness by injecting mock sources of known brightness into our images. We do this separately for each of the four fields, injecting 10^3 sources per iteration, with ∼30 iterations per field, to avoid crowding.

We generate source morphologies with GALFIT (Peng et al. 2002). While the completeness can depend sensitively on source size due to surface brightness dimming, our sources are quite compact (Figure 8); thus, we choose a lognormal half-light radius distribution such that the size distribution of the recovered sources matches well that of our galaxy sample (noting that if there were highly extended sources not present in our sample, we would be underestimating the incompleteness). While we choose Sérsic profiles (with a lognormal distribution peaked at $n = 1$), our size assumption results in a peak of unresolved sources, with a tail toward modestly resolved sources. We generate galaxy SEDs assuming a lognormal distribution in magnitude and a flat distribution in redshift using Bruzual & Charlot (2003) templates with stellar population properties tuned to generate fairly blue galaxies (this results in a median UV spectral slope (β) of −2.3, with a tail to −3 and −1). We note that while we use a specific SPS model here, the completeness simulations are broadly dependent on the UV colors produced; thus, the completeness results would be broadly similar using a different SPS model provided the UV slope range was similar.

After sources are added to the images, the images are processed through our entire analysis pipeline in the same way as our real data, measuring photometry with SE, applying all aperture corrections (though we did not apply the small zero-point offsets, as those may be due to real instrumental offsets not captured in these simulations), measuring photometric redshifts with EAZY, and applying our sample selection. In Figure 12, we show our completeness as a function of input redshift and F277W magnitude. This highlights that our sample selection accurately begins to select galaxies at $z \gtrsim 8.5$. Our completeness remains high to $m \sim 28.0$, beginning to fall off at $m > 28.5$, consistent with prelaunch expectations.

To calculate rest-UV absolute magnitudes for both our real and simulated sources, we follow Finkelstein et al. (2015), performing a basic round of SED fitting and measuring M_{1500} from the bandpass-averaged flux over a top-hat bandpass spanning 1450–1550 Å in the rest frame. We use 10^3 Monte Carlo simulations to obtain uncertainties on these magnitudes. For the completeness for our luminosity function estimate, we require the simulated sources to have best-fit photometric redshifts spanning 9.5–12, to match those of our galaxy subsample. We then calculate our completeness as a function of absolute magnitude and measure effective volumes in bins of absolute magnitude by integrating over the comoving volume element multiplied by the completeness for a given magnitude

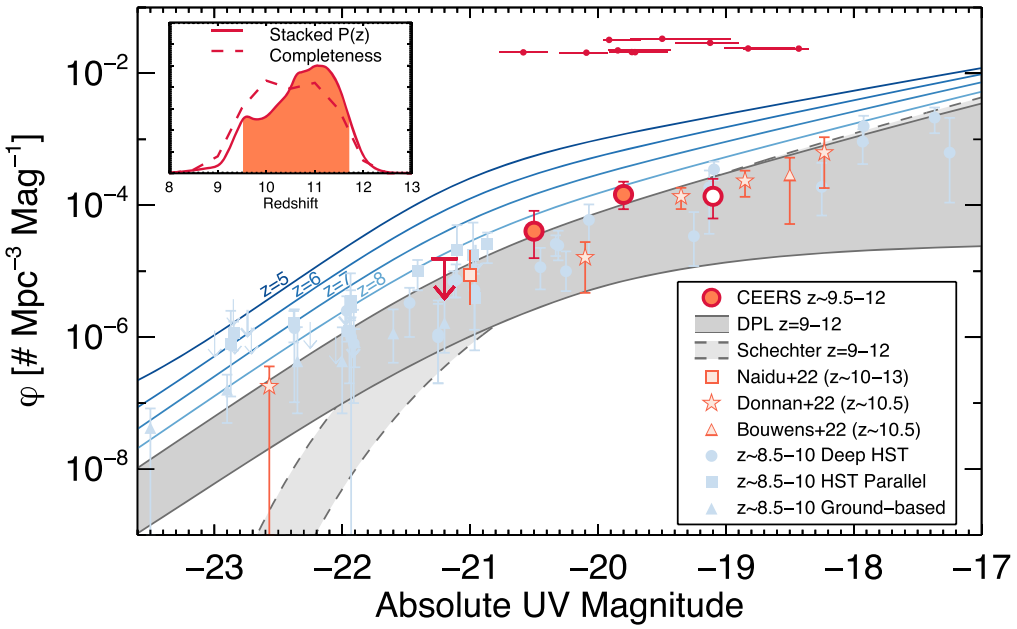

Figure 13. Rest-frame UV luminosity function at $z \sim 11$, shown as the red circles (the open circle denotes our faintest bin, where we are <30% complete). Each galaxy's magnitude and magnitude uncertainty are denoted by a small circle and line at the top of the figure. The light red symbols show literature constraints from JWST data from GLASS (Naidu et al. 2022b), CEERS+GLASS (D23, who also used UltraVISTA), and the HUDF (Bouwens et al. 2022b). The light blue points show a compilation of data from the literature at $z \sim 9$–10. Circles denote results from studies that used (modestly) deep imaging from surveys such as CANDELS and the HFF, including McLeod et al. (2015), Oesch et al. (2018), and Bouwens et al. (2019, 2021). The squares denote studies making use of HST pure parallel surveys, including Bernard et al. (2016), Morishita et al. (2018), and Rojas-Ruiz et al. (2020). The triangles denote results from the wide-area ground-based studies of Stefanon et al. (2019) and Bowler et al. (2020). The blue lines show the evolving DPL luminosity functions from Finkelstein & Bagley (2022) at $z = 4$–8 (this model was fit to data at $z = 3$–9). The dark gray shaded region shows the predictions from these DPL fits extrapolated to $z = 9$ (upper bound)—12 (lower bound); the light gray shaded region (outlined with dashed lines) shows a similar extrapolation from the evolving Schechter function fits from Finkelstein (2016; this model was fit to data at $z = 4$–8). The inset shows the stacked $\mathcal{P}(z)$ of the galaxies used in this luminosity function, as well as the redshift distribution estimated from the completeness simulations at $M_{\rm UV} = -20$. The observed $z \sim 11$ luminosity function is consistent with the top end of both smooth extrapolations, implying that the observed smoothly UV luminosity function evolution from $z = 4$ to 9 may be slowing at $z \sim 11$.

Table 3
$z \sim 11$ UV Luminosity Function

M_{1500} (mag)	Number of Galaxies	Effective Number	$\phi \times 10^{-5}$ (Mpc^{-3} mag^{-1})	$V_{\rm eff}$ (Mpc^3)
−21.2	0	0.0 ± 0.2	<2.8	92,200
−20.5	1	1.4 ± 0.6	$4.0\ ^{+4.2}_{-2.5}$	74,300
−19.8	6	4.8 ± 1.0	$14.6\ ^{+8.2}_{-5.9}$	51,900
−19.1[a]	2	2.4 ± 1.0	$13.7\ ^{+11.5}_{-7.4}$	30,600

Notes. The number densities are derived via an MCMC method that includes photometric uncertainties; thus, galaxies can contribute to the number density in more than one bin. The effective number column lists the mean and standard deviation of the number of galaxies per bin from these MCMC simulations, while the number column gives the actual value based on the measured magnitude.

[a] Our data are <30% complete in this faintest bin, so we do not consider this value reliable; it is indicated as an open symbol in Figure 13.

and redshift over our current survey area of 35.5 arcmin^2. We list our effective volumes in Table 3.

We measure our luminosity function following the methodology of Finkelstein et al. (2015), adopting a bin size of 0.7 mag. The number density in each bin is estimated via Markov Chain Monte Carlo (MCMC; see details in Finkelstein et al. 2015), sampling galaxies' absolute magnitude posterior distributions, such that galaxies can fractionally span multiple magnitude bins. We note that at $M_{\rm UV} > -19.5$, our completeness falls below 30% of its maximum value; thus, our faintest bin is shaded in white to indicate that the value is dominated by the completeness correction.

In Figure 13, we show our luminosity function results. In the top left panel, we show the stacked $\mathcal{P}(z)$ of the 10 sources in our sample at $9.5 < z_{\rm best} < 12.0$. The FWHM of this normalized distribution spans $z = 9.5$–11.7, with a peak at $z \sim 11$. The dashed line shows the completeness from our simulations as a function of redshift at the median absolute magnitude of our sample, which probes a broadly similar redshift range, albeit with a flatter distribution. In this figure, we also plot recent $z \sim 11$ results from D23, Naidu et al. (2022b), and Bouwens et al. (2022a). While the numbers in all studies are fairly small, the agreement between all studies is encouraging. We also compare to a wealth of studies in the literature from HST at $z = 8.5$–11.

Broadly speaking, our results at $z \sim 11$ do not show significant evolution from these modestly lower redshifts. This is consistent with the conclusion from Bowler et al. (2020), who noted that the bright end of the UV luminosity function ($M_{\rm UV} = -22$ to -23) shows little evolution from $z = 8$–10. However, here we are beginning to probe fainter, yet we still see little evolution.

Finally, we comment on both the shape of the UV luminosity function and the overall evolution. In Figure 13, we plot two empirical extrapolations. The first is a Schechter function from Finkelstein (2016), who measured the evolution of the Schechter function parameters as a linear function of $(1 + z)$ from $z = 4$–8; we plot this function evolved to $z = 9$ and 12, shown by the light gray shaded region. The dark gray shaded region is a similar empirical evolution, this time using data from $z = 3$–9 and assuming a double power-law (DPL) form from Finkelstein & Bagley (2022). Over the magnitude range

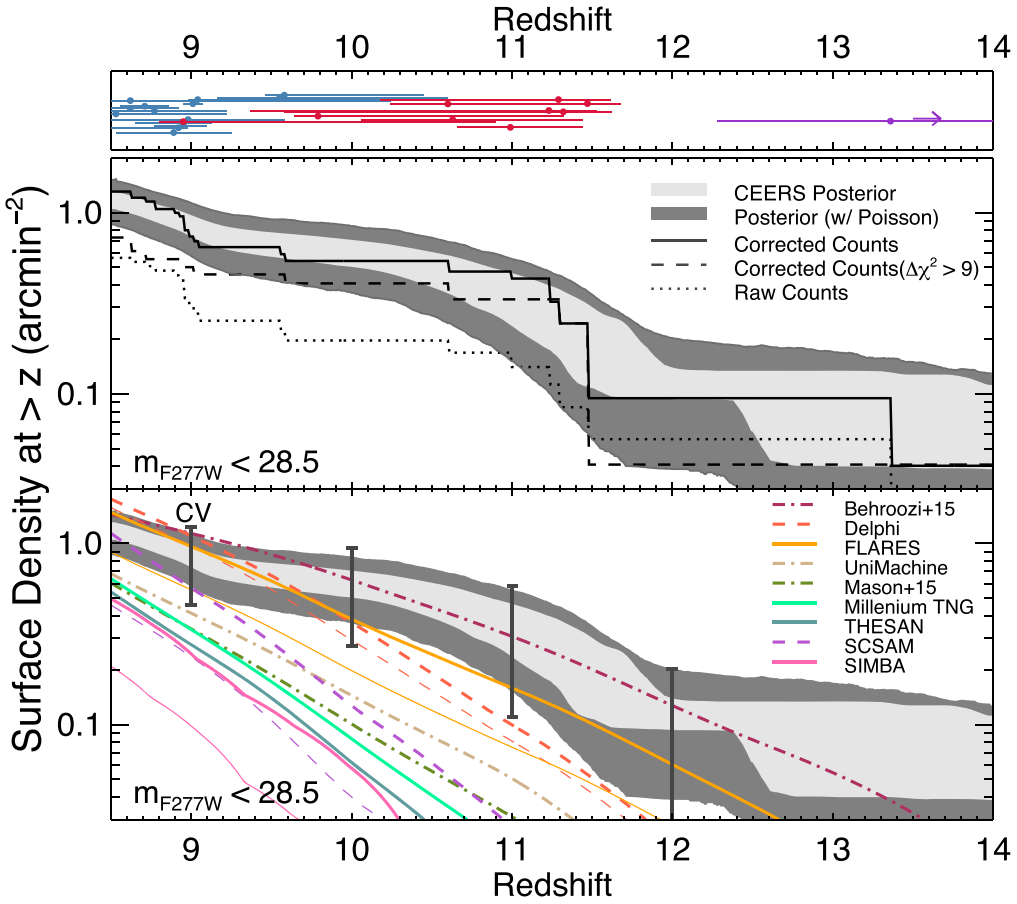

Figure 14. Cumulative surface density of sources with $m_{F277W} < 28.5$ at redshifts greater than a given x-axis value, starting at $z \geqslant 8.5$. The top panel shows the redshifts of individual objects (with blue, red, and purple denoting the $z \sim 9$, 11, and >12 samples). In the middle panel, the solid line shows the observed surface density after applying a correction for incompleteness; the dotted line shows the uncorrected (incomplete) values. The light shaded region shows the posterior on the distribution of the completeness-corrected surface density derived from Monte Carlo simulations marginalizing over the uncertainties in magnitude and photometric redshift; the dark shaded region includes Poisson uncertainty in this marginalization. The dashed line shows the completeness-corrected surface density if we had applied a more conservative sample selection criterion of $\Delta\chi^2_{\text{low-}z-\text{high-}z} > 9$. In the bottom panel, we show the same shaded regions, now comparing several recent model predictions, shown by the various colored lines (with solid, dotted–dashed, and dashed denoting predictions from hydrodynamical, semiempirical, and semianalytic models, respectively). For four models, we plot thicker lines for predictions with no dust attenuation and thinner lines for attenuated predictions. In this panel, we also show an estimate of the cosmic variance uncertainty using the method from Bhowmick et al. (2020), though this is likely an upper limit on this quantity. Even including all sources of uncertainty, our observed surface densities are higher than nearly all predictions, with the exception of the Behroozi & Silk (2015) semiempirical model. This apparent excess of high-redshift galaxies holds true at all $z = 8.5$–14, regardless of the $\Delta\chi^2$ cut, and is even true of the raw, noncorrected counts at $z > 10.5$. These results strongly imply that these predictions lack the full complement of physics describing star formation in the early universe, which we discuss in Section 8.2.

of our observed sources, these two functions agree, and our observations are consistent with, albeit at the high end of, these empirical extrapolations, implying that pure observed $z \sim 11$ UV luminosity function is similar to the $z = 9$ DPL luminosity functional fit from Finkelstein & Bagley (2022).

These results suggest that the evolution of the UV luminosity function, which had been smoothly declining from $z \sim 4$ to 8, begins to slow by $z \sim 11$. The luminosity function decline has been debated in the literature prior to the JWST era, notably by Oesch et al. (2018) and Bouwens et al. (2019), who found evidence for an accelerated decline in the UV luminosity function at $z > 8$. While our small sample cannot conclusively distinguish between these two scenarios, should future luminosity function efforts validate our observed number densities, it provides further evidence that there is significant star formation in our universe at $z > 10$. Previous efforts to recover the star formation histories of galaxies detected at $z \sim 9$–11 with HST and Spitzer/IRAC (Tacchella et al. 2022) and analysis of early JWST luminosity functions from D23 also arrived at a similar conclusion.

7.2. The Cumulative Surface Density of Galaxies at z > 8.5

As another view into the evolution of galaxies in the first 500 Myr, in Figure 14, we plot the cumulative surface density of the sources in our sample. This plot shows the integrated surface density for sources at redshift greater than z, limiting to $m_{F277W} < 28.5$ to avoid fainter luminosities where we are highly incomplete. We correct for incompleteness by counting each galaxy as one divided by the estimated completeness at the redshift and magnitude of a given galaxy. The solid black line shows the completeness-corrected value from our sample, while the dotted line shows the results with no correction. We note that across the redshift range considered here, the completeness correction is typically a factor of ∼2, which is typical of most analyses of modestly faint galaxies. This, of course, means that the accuracy of the completeness corrections remains an important systematic, one we will explore in more detail in future works. However, we note that with the exception of the lowest redshifts studied here, even the uncorrected counts exceed most of the model predictions;

therefore, this is a reasonable lower limit on the true surface density.

To encompass the uncertainty in both flux and photometric redshift, we run Monte Carlo simulations sampling both the F277W flux and photometric redshift posterior distributions and plotting the 68% confidence range on the surface density as the 68% spread in values from these 10^3 simulations as the light gray shaded region. We also run a set of Monte Carlo simulations additionally sampling the Poisson uncertainty, shown by the wider dark gray region. We find that at $z > 8.5$ and $m_{\rm F277W} < 28.5$, our results suggest >one galaxy per square arcminute. This trends downward when integrating from higher redshifts, albeit slowly, with this quantity not dropping to 0.1 arcmin^{-2} until $z > 11.5$–12. We note that these surface densities are only slightly reduced when removing sources with $4 < \Delta\chi^2 < 9$ (dashed line).

We also show the cosmic variance calculated based on the BLUETIDES simulation (Bhowmick et al. 2020). While these uncertainties appear comparable to the combined redshift, photometric, and Poisson uncertainties, these cosmic variance uncertainties are likely an upper limit. As they are simulation-based, they rely on the predicted abundance of galaxies, and as we show here, most simulations underpredict the abundance of $z > 9$ galaxies. As the relation between UV luminosity and halo mass in this epoch is very uncertain (Section 8.2), this leads to uncertainty in these calculations. We show this upper limit on the true cosmic variance uncertainty to highlight that it implies that our results are still significantly above the predictions.

In the bottom panel of Figure 14, we also compare to nine recent model-based predictions, including from the Santa Cruz (Yung et al. 2019a, 2020) and DELPHI (Dayal et al. 2017) SAMs, empirical models by Behroozi & Silk (2015), Mason et al. (2015), and the UNIVERSEMACHINE (Behroozi et al. 2019, 2020), and the cosmological hydrodynamic simulations FLARES (Wilkins et al. 2023, 2022), THESAN (Kannan et al. 2022), MillenniumTNG (Kannan et al. 2022), and SIMBA (Davé et al. 2019). Interestingly, our results are noticeably higher than most predictions from physics-based models. At $z < 9.5$, our observations are consistent with both the Behroozi & Silk (2015) and DELPHI models (and the FLARES predictions with no dust attenuation), while at $z > 10$, our results lie significantly above all predictions, with the exception of Behroozi & Silk (2015). We include results with and without dust attenuation for the Santa Cruz SAM, DELPHI, FLARES, and SIMBA. We note that while there is likely an overdensity at $z \sim 8.7$ in the observed field, this has no impact on the surface density measured at $z > 9$, and the excess of our observed galaxy numbers over nearly all theoretical predictions persists at all redshifts. We discuss possible explanations for this discrepancy in Section 8.2.

It is important to note that the compilation of theory results included in this comparison is made utilizing several different modeling approaches and with some different modeling assumptions (see Somerville & Davé 2015 for a thorough review). For instance, cosmological hydrodynamic simulations (e.g., MillenniumTNG, THESAN, and SIMBA) are carried out by solving the equations of hydrodynamics, thermodynamics, and gravity for large numbers (typically billions) of dark matter particles, gas particles, and star particles. These simulations are capable of self-consistently tracking various properties (e.g., stellar and gas mass) and morphology, as well as their correlation with the large-scale environment. Cosmological zoom-in simulations (e.g., FLARES, based on the EAGLE model) track galaxy evolution at higher resolution by zooming into regions of interest in a cosmological simulation and resimulating at higher mass resolution and, in some cases, with more fundamental treatments of baryonic processes.

All numerical cosmological hydrodynamic simulations require the use of "subgrid" prescriptions to represent physical processes that operate at physical scales below the resolution limit of these simulations (e.g., star formation, stellar feedback, black hole growth and feedback, etc.). Hydrodynamic simulations are subject to tension between the simulated volume and resolution. Large volumes are required to capture the rare overdense regions that host massive galaxies (the ones of interest in this comparison), and high mass resolution is required to properly resolve the evolution and assembly history of these galaxies. Thus, it is extremely challenging to capture the rare peaks where high-redshift galaxies form while simultaneously resolving these low-mass halos with sufficient numbers of particles.

On the other hand, the semianalytic modeling technique uses phenomenological recipes to track the formation and evolution of galaxies in dark matter halo merger trees and is able to predict a wide variety of physical and observable properties of galaxies with a relatively low computational cost. These models are typically capable of simulating galaxies over a wider mass range and are less susceptible to the volume-resolution tension faced by hydrodynamic simulations (though even obtaining an adequate dynamic range for dark matter–only simulations and halo merger trees is challenging; see Section 8.2 for a more detailed discussion). For both the SAMs and the hydrodynamic simulations described above, the predictive power of these models relies on the assumption that the physical recipes within, often derived from or calibrated to nearby observations, accurately represent the processes that drive galaxy formation in the high-redshift universe.

A different class of models are the (semi)empirical methods, also known as subhalo abundance matching or halo occupation distribution, which efficiently map the properties of a large ensemble of galaxies onto the properties of dark matter halos guided by a set of observed quantities and scaling relations (see Wechsler & Tinker 2018 for a detailed review). This approach is independent of specific galaxy formation models and guaranteed to match the observational constraints to which these models are calibrated. However, extrapolating these models to redshifts different from those where they were calibrated has less physical grounding.

This is only a broad overview of the simulation methods that produced the results presented in Figure 14, with the aim of emphasizing that these results are produced with different methods, each having their own advantages and disadvantages. We refer the reader to these works for a full description of the design of these simulations and their performance against observational constraints.

8. Discussion

8.1. Observational Effects

In the previous sections, we showed that the evolution of the UV luminosity function appears to be "slowing" at $z > 10$ and that the abundance of $z > 9$ galaxies significantly exceeds predictions by most physically based theoretical models. Here we explore several possible reasons for these unexpected

results. At this early time with JWST, we must first acknowledge that the purity of the sample is uncertain. Should the majority of our galaxy sample turn out to be low-redshift interlopers, it would put our results more in line with the theoretical predictions. While this is unlikely given our detailed photometry and selection process, these strong claims require strong evidence; thus, spectroscopic confirmation is a must. Without this, it remains possible that heretofore unexpected contaminants could be dominating our sample. The CEERS spectroscopic program will observe a number of these sources, and thus we may soon be able to gain further confidence in our galaxy sample. We note that nondetections in ALMA dust-continuum follow-up of three $z \gtrsim 11$ sources in other early JWST fields strengthens the high-redshift solutions in all four cases (e.g., Bakx et al. 2023; Fujimoto et al. 2022; Yoon et al. 2022).

One other observational systematic effect that could affect these results is that of aperture corrections. In the above sections, we described our multistep approach to deriving total fluxes, accounting for what can be detected in the images and then using simulations to correct for any missing flux from the wings of the PSF. As one additional check, in Section 6.2, we explored whether our fluxes are significantly systematically brighter (or fainter) than other objects published in the literature, finding that while there was significant scatter, there was no evidence that our fluxes were systematically brighter (especially by the large amount needed to explain the luminosity function excess). Nonetheless, future improvements in photometry can increase confidence in these results.

8.2. Theoretical Implications

Here we speculate on potential physical causes for the observed abundances should future observations validate our results. One potential explanation would be a complete absence of dust attenuation (e.g., Ferrara et al. 2022; Mason et al. 2023). As shown in Figure 14, for those models where we plot the (un) attenuated predictions as thicker (thinner) lines (when available), higher number densities are predicted. While a full analysis of the colors of these galaxies is beyond the scope of this paper, their SEDs do appear quite blue; thus, it is possible that they have little dust attenuation. Of note is that bright $z \sim 9$–10 galaxies exhibit only marginally redder spectral slopes. For example, Tacchella et al. (2022) found a median UV spectral slope of ~ -2 for $M_{\rm UV} = -21$, implying that a small amount of dust attenuation was present. Simulations (FLARES, SIMBA) predict a large, $\sim 3\times$ reduction in surface densities due to dust, while the DELPHI SAM predicts much less. Should a lack of dust attenuation at $z \sim 10$–12 be responsible for our observed evolution, one may expect to see the luminosity function at higher redshifts continue to decline. However, we note that even those models plotted without dust attenuation mostly fail to match our observed galaxy numbers at $z \gtrsim 10$, though the FLARES model is only below at the 1σ–2σ level.

Most models we have compared to employ some kind of Kennicutt–Schmidt (KS) star formation law, relating the star formation rate surface density to the dense gas surface density, assuming a constant efficiency per freefall time. While models such as the Santa Cruz SAM (Yung et al. 2019a, 2019b) adopt a KS law that becomes steeper at higher gas surface densities, leading to higher star formation efficiencies at early times, when the typical gas densities are higher, these models still imply a fairly long gas depletion time compared to the age of the universe at these extremely high redshifts. For example, in the local universe, the depletion time for dense (molecular) gas is $\sim$1.5 Gyr. This timescale appears to be shorter at higher redshifts, as short as $\sim$0.7 Gyr at $z = 2$ and perhaps even shorter at higher redshifts (e.g., Sommovigo et al. 2022). The Behroozi & Silk (2015) model would require stars to form at the same rate that gas is funneled into halos, effectively assuming a negligibly short gas depletion time. It is also possible that the star formation efficiency could be higher in low-metallicity gas, though the opposite trend has been proposed (e.g., Krumholz & Dekel 2012), or that stellar-driven winds are weaker. Additionally, given the extremely high gas densities at these high redshifts, the cold neutral medium itself may also participate in star formation in addition to the molecular phase typically tracked by some simulations (Bialy & Sternberg 2019).

Another explanation for the poor match between the predictions and observations in SAMs is the mass and temporal resolution of halo merger trees. The halo merger and assembly histories, as an important component of SAMs and some empirical models, can either be constructed with the extended Press–Schechter (EPS) formalism (e.g., Somerville & Kolatt 1999) or extracted from *N*-body cosmological simulations (e.g., Behroozi et al. 2013). However, even the current-generation state-of-the-art cosmological simulations only saved a few dozen snapshots at $z > 6$ and among them only a handful at $z \gtrsim 13$, which is insufficient to properly capture the merger histories of these early-forming halos. Furthermore, the bulk of these high-redshift halos have masses near the resolution limit of these cosmological simulations, which further limits their ability to resolve halo merger trees at these extreme redshifts. On the other hand, while EPS-based merger trees have been compared to and shown to be in good statistical agreement with *N*-body merger trees, they are untested in the mass and redshift ranges that are explored here (see Yung et al. 2020 for a detailed discussion).

Conversely, the lower resolution of the gas content in cosmological hydrodynamic simulations that does not resolve the multiphase nature of the interstellar medium (ISM) could contribute to the discrepancy between predictions and observations. It has been shown (e.g., Sparre & Springel 2016) that increasing the resolution typically leads to more dense gas and thus burstier star formation (without adjusting any free parameters, although this depends on the nature of the subgrid recipes used for star formation and stellar feedback). Thus, solely increasing the resolution in cosmological hydro simulations, such as in the Lyra cosmological zoom simulation set (Gutcke et al. 2022), and/or improving the subgrid recipes for the ISM, star formation, and stellar feedback may partly alleviate the poor match between the simulations and these observations. Should future zoom-in simulations (such as those in Ma et al. 2018) make predictions to these high observable epochs, they may be better able to match observations.

Finally, and perhaps most interestingly, would be an evolution in the IMF. It has long been predicted that the first stars would have a top-heavy IMF (e.g., Bromm et al. 2001; Clarke & Bromm 2003) due not only to extremely low metallicities (e.g., Chon et al. 2021; Sharda & Krumholz 2022) but also to the rising cosmic microwave background temperature floor (e.g., Larson 1998). While the relatively massive systems we see here are unlikely to be dominated by metal-free

stars, it is possible that their metallicities are low enough to affect the characteristic mass of their IMF. In fact, these observations now probe so close to the Big Bang that it would be surprising if the IMF did not begin to evolve at some point.

We therefore explore what excess UV luminosity is needed to match our observation. We first do a simple test by exploring how much we would have to shift our observed luminosity function data in Figure 13 in UV magnitude such that they would match the DPL luminosity function fits from lower redshifts extrapolated to $z \sim 11$. We find that our data would need to shift $\sim$1 mag fainter to match these empirical predictions. Therefore, if an evolving IMF was the sole explanation for the higher-than-expected luminosities, the UV luminosity would need to be boosted by about a factor of $\sim$2.5.

As a more complex version of this analysis, we estimate the total masses for the host halos of galaxies across the luminosity function via abundance matching. We follow the procedures of Reddick et al. (2013), assuming a 0.2 dex scatter in UV luminosity at fixed halo mass, showing the observed relations between halo mass and UV absolute magnitude in Figure 15; qualitatively similar findings applied for scatter ranging from zero to 0.4 dex. For $z \leqslant 9$, we use the DPL UV luminosity functions from Finkelstein & Bagley (2022). For the expected evolution to $z \sim 11$, we extrapolate their fits to $z \sim 12$ and show the volume-averaged extrapolated luminosity function for $z = 9.5$–12 with the bright blue line labelled as $z \sim 11$ Extrapolated. For our actual observed values, as we have not fit a functional form here due to our limited dynamic range in luminosity, we assume the $z = 9$ DPL fit from Finkelstein & Bagley (2022), as this is most consistent with our observed number densities in Figure 13. We show in red the $M_{\rm halo}$–$M_{\rm UV}$ relation when abundance matching the average of the $z \sim 9.5$–12 halo mass functions to this assumed observed luminosity function. In this figure, we highlight $M_{\rm UV} = -20$, which is where our $z \sim 11$ observations are most constraining. Our abundance-matching analysis predicts that based on our observations, these galaxies are hosted in halos with $\log(M_h/M_\odot) = 10.55$ (upper light red bar). If we had instead used the extrapolated $z \sim 11$ results, these halos should host galaxies with $M_{\rm UV} = -19.3$ to -19.4 (lower light red bar). This implies that our observed galaxies are ~ -0.6 to -0.7 mag brighter (1.8$\times$ in UV luminosity) than expected for a galaxy in their host halo.

A UV luminosity boost of $\sim$1.8–2.5 is not unexpected in the context of a top-heavy IMF. Raiter et al. (2010) explored the impact of differing IMFs on the UV luminosity (specifically investigating the conversion from UV luminosity to star formation rate). They found that the top-heavy IMFs proposed by Tumlinson (2006), which peak at $\sim$10–40 $M_\odot$, yield a UV luminosity $\sim$0.4 dex higher than Salpeter at a stellar metallicity of $\log(Z/Z_\odot) = -2$. Such a brightness increase is exactly what we find would be needed to bring our observed luminosity function in line with expectations. This interpretation was also considered by Harikane et al. (2023), who found that some top-heavy IMF models can boost the UV luminosity by $\sim$3–4$\times$ (e.g., Zackrisson et al. 2011). The top end of the expected UV luminosity boost was discussed by Pawlik et al. (2011), who predicted up to a factor of 10 increase in UV luminosity for a top-heavy IMF in a zero-metallicity system.

If the interpretation of invoking a top-heavy IMF were correct, there would be a number of additional effects and consequences that may be detectable, thus providing an independent cross-check. One such diagnostic is the predicted boost to nebular line emission, including the occurrence of strong He II 1640 Å features (e.g., Bromm et al. 2001; Raiter et al. 2010), accessible via our upcoming CEERS spectroscopic follow-up. A qualitatively different signature of a top-heavy IMF would be a supernova rate that may be enhanced by up to an order of magnitude per unit stellar mass, possibly resulting in stronger supernova-driven galactic winds (e.g., Dayal & Ferrara 2018; Jaacks et al. 2019), with alternative explanations for any such signature possible as well (e.g., Naab & Ostriker 2017).

We note that the high observed galaxy number densities do not represent a fundamental problem for hierarchical structure formation models. For instance, using a SIMBA simulation with all explicit feedback processes turned off and no extinction results in $\sim$2$\times$ higher number densities than observed (even with a standard IMF), and the empirical Behroozi & Silk (2015) model can match the data. Hence, the results do not necessarily challenge the ΛCDM paradigm but rather our understanding of the physics of early galaxy evolution within that paradigm.

9. Summary and Conclusions

We have presented the results from a study of galaxies at $z > 9$ in the first epoch of NIRCam imaging from the CEERS program. This imaging covers $\sim$35 arcmin^2 with seven photometric filters (six broadband and one medium-band) covering $\sim$1–5 μm, reaching $m \gtrsim 29$. In addition to being some of the widest moderate-depth imaging available early in Cycle 1, these data probe a parameter space in wavelength and depth optimal for studying these early redshifts.

Following a detailed reduction of the data (with full details available in Bagley et al. 2022b), we measure photometry from all sources in the field. Using a combination of the F277W and F356W images as the detection images, we create a 13-band photometric catalog, inclusive of imaging in six HST filters in this field. We emphasize the calculation of robust colors, total fluxes, and flux uncertainties using simulations to validate our choices in cataloging. We explore any systematic offsets in our photometry by comparing to the best-fitting model templates for $\sim$800 spectroscopically confirmed sources in our field, finding that the large zero-point offsets that affected early JWST/NIRCam studies have been resolved. We do measure (and apply) offsets from $\sim$1% to 5%, which are reflective of potential residual zero-point corrections in the NIRCam imaging, adjustments to our estimations of the total fluxes, and potential mismatches between the used template set and the SEDs of the real galaxies.

We estimate photometric redshifts for all sources in our catalog using EAZY. In addition to the standard set of 12 FSPS templates, we add six new templates designed to span the blue colors expected in very early galaxies, as well as very strong emission lines expected at low metallicities. We designed a set of selection criteria to robustly select galaxies at $z > 9$, balancing our desire to maximize completeness with our need to minimize contamination. These selection criteria demand both robust detection significance (S/N > 5.5 in at least two filters, measured in 0.″2 diameter apertures) and photometric redshift PDFs that are strongly constrained to $z > 8.5$.

Following visual inspection to remove any remaining spurious sources (shown in the Appendix), our selection resulted in a sample of 26 robust candidate galaxies at $z > 8.5$.

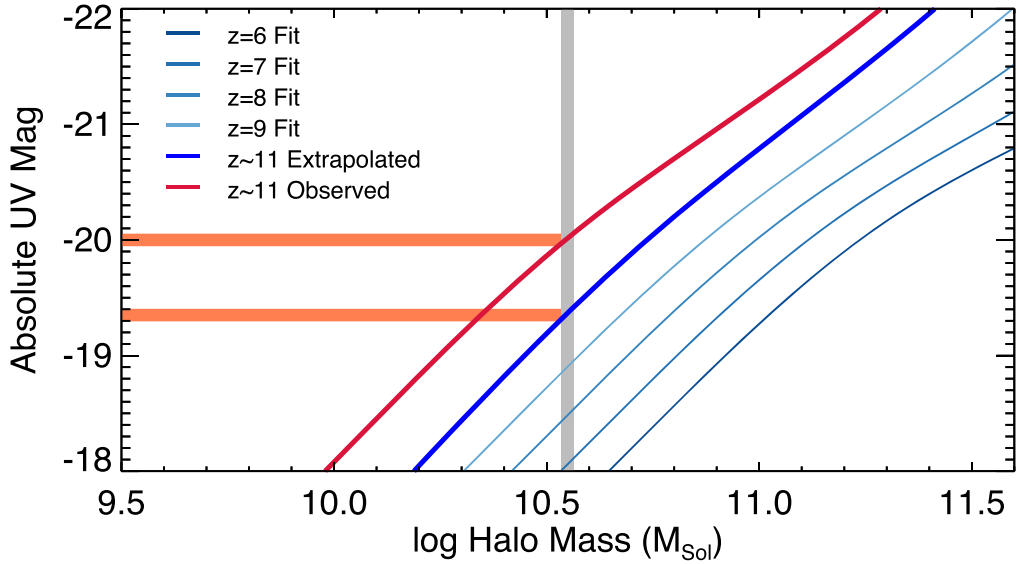

Figure 15. Relation between rest-UV absolute magnitude and halo mass, obtained via abundance matching the observed UV luminosity functions assuming the DPL fits from Finkelstein & Bagley (2022) for $z = 4$–8 and extrapolated to $z \sim 11$ (the thick blue curve is equivalent to the average of the dark gray shaded region in Figure 13). For our observed $z \sim 11$ UV luminosity function, we use the $z = 9$ DPL fit from Finkelstein & Bagley (2022), as it is very consistent with our observations (Figure 13). This relation implies that our observed galaxies at $M_{\rm UV} = -20$ (where our observations place the tightest constraints) have a host halo mass of $\log(M_h/M_\odot) = 10.55$. At this halo mass, the expected UV luminosity based on the expected UV luminosity function is 0.6–0.7 mag fainter. This implies that based on the observed abundances, our observed galaxies are 1.8× more UV luminous than expected from extrapolation of HST results. While larger sample sizes and spectroscopic confirmations are needed to have greater confidence in the $z \sim 11$ luminosity function, should these results be confirmed, it implies that $z > 10$ galaxies are more luminous in the rest-UV than expected. One possible explanation could be an increasing prevalence at higher redshifts of a top-heavy IMF, which is predicted to dominate at very low metallicities.

The majority of the sample (15 galaxies) lies at $8.5 < z < 10$, while nine galaxies lie at $10 < z < 12$. Two candidates lie at higher redshifts: the previously published $z \sim 16.5$ source -D23 and a new candidate at $z \sim 13$. While a full analysis of the properties of these galaxies is reserved for future work, their rest-frame UV SEDs are fairly uniformly blue, and the galaxies are compact, with a median half-light radius of 3.6 pixels (0.″11, or 0.46 kpc). We explore information from multi-wavelength constraints on these galaxies and find no significant detections in either X-ray or submillimeter/millimeter wavelengths, strengthening the conclusion that these sources reside at $z > 9$. We compare to the few previous searches for $z > 9$ sources in this field and find that for sources in common between studies, the photometric redshifts agree quite well. Of the two known HST sources previously published in this field, we very robustly constrain the SED of one, placing it at the slightly lower redshift of $z \sim 8.1$. The other is completely absent in our NIRCam data, and we conclude that it was a likely supernova serendipitously captured by the HST data taken in 2013.

We estimate our sample completeness as a function of redshift and magnitude using source-injection simulations and present an early look at the $z \sim 11$ rest-UV luminosity function. We find that the abundance of modestly bright ($M_{\rm UV} \sim -20$) galaxies at $z \sim 11$ does not appear to be evolving from $z \sim 9$–11, which is unexpected given the steady decline in the abundance of such galaxies from lower redshifts to $z = 8$, though such nonevolution at the very bright end from $z = 8$ to 10 had previously been discussed by Bowler et al. (2020). We then compare the surface density of our sources to a variety of model predictions, finding that even after accounting for several sources of systematic and random uncertainty, the observed abundance of galaxies is in significant excess of these predictions.

We explore several potential explanations for these unexpected results. While not the most exciting, significant sample contamination cannot be conclusively ruled out. These data represent our first foray into a new cosmic epoch, and spectroscopic confirmation of the redshifts to at least a subset of these ultrahigh-redshift sources is necessary to gain confidence in our sample selection processes. However, such data will begin to flow soon, with CEERS scheduled to spectroscopically observe ∼10 of these sources in late 2022 (though these high redshifts may necessitate longer exposure times for future cycle programs).

Should these high abundances of $z = 9$–13 galaxies be confirmed, we explore what possible changes in the models could bring their predictions into agreement with observations. One very exciting possibility is that we are beginning to probe an era where star formation in galaxies is dominated by a top-heavy IMF due to the presence of very low metallicities, which could increase the ratio of UV luminosity per unit halo mass. We explore the "excess" UV luminosity from our observations, both by comparing to the expected UV luminosity function based on extrapolations from lower redshift and through an abundance-matching exercise, and we find that our UV luminosities may be enhanced by 1.8–2.5×. This is very similar to the predicted excess UV emission from a low-metallicity stellar population where the IMF peaks at 10–40 $M_\odot$ (Tumlinson 2006) of a factor of ∼2.5×, implying that a top-heavy IMF is a plausible physical explanation. We also discuss how potential changes to the dust attenuation, star formation law, galactic feedback, and resolution of numerical simulations could collectively contribute to reconciling our observations with model predictions. As all of these potential explanations could bring the model predictions into agreement with the observations, our results are not inconsistent with the ΛCDM paradigm.

These possibilities are exciting, and while one might expect that at $z > 10$, we should expect to see changes in star formation physics, such as a top-heavy IMF, our results are just a first glimpse, and the data available in the near future will provide much stronger constraints. Specifically, the remainder of Cycle 1 will see the creation of high-redshift galaxy samples orders of magnitude larger than we present here from the combination of the full CEERS survey, COSMOS-Web (PIs: Casey & Kartaltepe), PRIMER (PI: Dunlop), and NGDEEP (PIs: Finkelstein, Papovich & Pirzkal), along with JADES (PI: Rieke). Additionally, Cycle 1 will also see spectroscopic follow-up with NIRSpec of NIRCam-identified sources from CEERS and JADES, and several Cycle 2 programs will certainly target these sources with deep observations. While it is early days with JWST, our first-look CEERS results provide an enthralling glimpse of the potential secrets the early universe has that our observations can unlock.

We acknowledge that the location where this work took place, the University of Texas at Austin, sits on indigenous land. The Tonkawa lived in central Texas, and the Comanche and Apache moved through this area. We pay our respects to all of the American Indian and indigenous peoples and communities who have been or have become a part of these lands and territories in Texas on this piece of Turtle Island. We acknowledge support from NASA through STScI ERS award JWST-ERS-1345. We thank Marcia Rieke, Daniel Schaerer, Volker Bromm, and Mike Boylan-Kolchin for helpful conversations. Some/all of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific observations analyzed can be accessed via 10.17909/z7p0-8481.

Facilities: HST (ACS, WFC3), JWST (NIRCam).

Appendix A
Full Photometry

In Tables 4 and 5, we list the measured photometry for our final sample.

Table 4
NIRCam Photometry for the $z > 8.5$ Galaxy Sample

ID	F115W	F150W	F200W	F277W	F356W	F410M	F444W
CEERS2_2159	-4.1 ± 5.9	7.0 ± 6.7	22.5 ± 4.9	94.2 ± 4.6	96.3 ± 3.7	104.4 ± 7.3	89.7 ± 5.4
CEERS1_1730	7.7 ± 5.3	6.0 ± 5.7	33.3 ± 4.4	30.1 ± 3.4	33.3 ± 3.3	26.1 ± 6.8	39.4 ± 4.6
CEERS2_588	-5.3 ± 8.3	28.7 ± 8.5	49.8 ± 7.2	63.6 ± 5.1	58.5 ± 5.1	59.4 ± 10.8	68.8 ± 7.3
CEERS1_8817	-0.4 ± 4.7	21.3 ± 5.3	25.9 ± 4.3	20.7 ± 3.0	31.9 ± 2.9	25.3 ± 5.9	32.8 ± 4.6
CEERS2_7929	-6.8 ± 5.1	17.0 ± 5.8	25.0 ± 4.2	20.5 ± 2.9	18.0 ± 2.9	22.9 ± 6.2	23.4 ± 4.6
CEERS6_7641	4.2 ± 3.8	18.5 ± 4.2	27.7 ± 3.8	20.3 ± 2.7	26.8 ± 2.8	17.3 ± 5.0	41.0 ± 3.6
CEERS2_5429	-7.0 ± 3.2	11.7 ± 3.7	26.9 ± 2.7	18.0 ± 2.2	18.2 ± 1.9	17.1 ± 4.0	19.1 ± 2.8
CEERS1_7227	5.5 ± 3.7	13.7 ± 4.1	23.8 ± 3.3	17.6 ± 2.3	17.9 ± 2.0	17.4 ± 4.8	14.4 ± 3.0
CEERS6_7603	1.9 ± 2.7	9.9 ± 2.9	11.9 ± 2.3	9.8 ± 1.6	9.4 ± 1.6	12.6 ± 3.4	18.7 ± 2.3
CEERS6_4407	-0.5 ± 2.3	11.2 ± 3.0	18.1 ± 2.8	8.9 ± 2.0	11.1 ± 1.9	11.6 ± 4.8	8.6 ± 2.5
CEERS6_8056	-1.5 ± 1.8	7.7 ± 2.2	10.5 ± 1.8	8.3 ± 1.5	7.5 ± 1.4	12.0 ± 2.5	10.5 ± 2.3
CEERS2_2402	25.4 ± 4.3	72.2 ± 5.1	62.8 ± 3.9	76.3 ± 4.0	89.1 ± 3.6	121.7 ± 6.5	205.0 ± 5.8
CEERS1_6059	21.5 ± 3.8	87.5 ± 4.5	80.9 ± 3.7	58.7 ± 2.6	58.1 ± 2.4	49.6 ± 4.7	61.6 ± 3.4
CEERS1_1875	10.3 ± 5.5	38.7 ± 6.4	37.5 ± 4.9	54.8 ± 4.6	70.1 ± 4.1	64.9 ± 7.1	109.9 ± 6.2
CEERS1_3858	21.3 ± 4.3	66.5 ± 5.2	59.1 ± 4.1	46.7 ± 3.5	54.7 ± 2.9	52.1 ± 5.7	50.2 ± 4.2
CEERS2_7534	12.4 ± 3.3	31.8 ± 4.3	32.3 ± 3.0	40.3 ± 2.3	57.8 ± 2.1	107.5 ± 3.9	163.0 ± 3.1
CEERS1_3908	-1.8 ± 6.9	47.2 ± 7.9	43.0 ± 6.9	44.2 ± 4.7	54.0 ± 4.9	41.1 ± 8.9	73.2 ± 7.8
CEERS6_4012	9.6 ± 7.3	30.7 ± 8.4	36.4 ± 6.7	32.3 ± 5.9	31.5 ± 4.8	27.7 ± 8.9	29.8 ± 7.6
CEERS2_2324	-9.4 ± 8.0	45.7 ± 8.7	32.6 ± 7.9	32.2 ± 7.0	28.1 ± 5.1	24.1 ± 10.8	18.3 ± 8.6
CEERS1_3910	-8.8 ± 7.0	30.4 ± 9.7	26.5 ± 6.3	23.1 ± 4.7	49.6 ± 4.9	71.6 ± 11.3	76.9 ± 6.9
CEERS1_5534	14.8 ± 5.2	35.4 ± 5.9	24.5 ± 4.6	25.7 ± 3.8	37.5 ± 3.3	25.4 ± 6.7	66.0 ± 5.3
CEERS1_4143	5.0 ± 5.3	26.2 ± 6.2	18.0 ± 4.9	20.6 ± 4.0	16.3 ± 3.7	7.1 ± 6.4	16.7 ± 5.4
CEERS3_1748	3.2 ± 2.8	13.9 ± 3.2	11.4 ± 3.0	14.9 ± 2.1	34.6 ± 2.1	62.7 ± 4.3	78.0 ± 3.3
CEERS2_1298	6.5 ± 2.2	11.9 ± 2.5	15.8 ± 2.1	13.6 ± 1.6	30.6 ± 1.5	41.9 ± 3.1	68.5 ± 2.2
CEERS2_2274	5.5 ± 3.3	9.6 ± 3.8	14.6 ± 2.9	8.4 ± 2.8	17.1 ± 2.7	11.5 ± 5.3	22.9 ± 5.1
CEERS2_1075	5.4 ± 2.5	12.1 ± 2.9	9.9 ± 2.1	8.2 ± 1.8	9.5 ± 1.6	8.0 ± 3.6	21.2 ± 2.2

Note. All fluxes are in nJy. The horizontal lines distinguish our $z > 12$, 10–12, and 8.5–10 samples.

Table 5
HST/ACS and WFC3 Photometry for the $z > 8.5$ Galaxy Sample

ID	F606W	F814W	F125W	F140W	F160W
CEERS2_2159	2.7 ± 8.8	3.9 ± 10.5	2.3 ± 14.0	-23.6 ± 27.6	6.4 ± 12.7
CEERS1_1730	15.0 ± 8.7	-8.0 ± 10.7	-13.6 ± 12.8	33.1 ± 22.4	-13.5 ± 11.1
CEERS2_588	-5.2 ± 12.5	-20.2 ± 18.3	5.8 ± 21.0	-13.0 ± 36.6	68.5 ± 18.2
CEERS1_8817	-2.0 ± 6.6	-7.4 ± 6.8	-24.9 ± 10.3	⋯	34.6 ± 10.2
CEERS2_7929	-6.6 ± 7.2	0.1 ± 8.0	2.6 ± 9.2	-7.9 ± 24.7	15.9 ± 8.8
CEERS6_7641	2.8 ± 7.7	1.6 ± 8.0	10.8 ± 9.4	28.7 ± 13.6	23.3 ± 10.7
CEERS2_5429	8.1 ± 5.2	1.4 ± 6.5	-8.1 ± 7.8	-31.8 ± 14.0	29.6 ± 7.1
CEERS1_7227	1.5 ± 5.8	1.6 ± 6.5	-6.5 ± 7.9	⋯	13.1 ± 7.4
CEERS6_7603	-2.2 ± 4.7	-0.6 ± 4.9	-0.1 ± 5.1	5.9 ± 9.8	1.1 ± 4.6
CEERS6_4407	1.0 ± 4.7	-1.5 ± 5.0	0.8 ± 6.4	⋯	6.8 ± 5.9
CEERS6_8056	-0.1 ± 3.2	-0.5 ± 3.8	3.7 ± 4.9	2.3 ± 9.0	6.4 ± 3.7
CEERS2_2402	6.2 ± 7.1	-6.0 ± 8.9	50.5 ± 11.3	70.3 ± 13.3	65.0 ± 9.5
CEERS1_6059	-2.5 ± 5.5	1.6 ± 5.5	38.3 ± 9.0	90.5 ± 16.7	44.2 ± 7.9
CEERS1_1875	6.5 ± 9.2	16.9 ± 13.1	17.2 ± 14.5	65.0 ± 24.9	39.0 ± 13.1
CEERS1_3858	2.5 ± 7.0	4.6 ± 8.6	35.8 ± 10.2	46.9 ± 18.9	50.7 ± 8.7
CEERS2_7534	-0.3 ± 4.8	0.8 ± 5.1	21.7 ± 7.0	20.3 ± 9.1	24.0 ± 6.1
CEERS1_3908	14.7 ± 11.5	19.2 ± 13.9	46.7 ± 18.7	14.1 ± 33.7	53.8 ± 16.8
CEERS6_4012	4.9 ± 12.6	-4.0 ± 17.1	29.9 ± 19.4	144.9 ± 26.2	71.0 ± 18.6
CEERS2_2324	5.7 ± 16.9	-0.2 ± 16.9	29.8 ± 20.6	-23.1 ± 32.6	39.3 ± 17.2
CEERS1_3910	8.6 ± 10.9	10.5 ± 12.5	-2.8 ± 17.8	10.1 ± 31.3	24.2 ± 16.0
CEERS1_5534	10.9 ± 7.0	-2.7 ± 8.3	0.2 ± 11.6	26.7 ± 21.3	37.2 ± 10.4
CEERS1_4143	9.1 ± 9.4	1.4 ± 12.2	24.6 ± 13.7	44.4 ± 23.6	30.5 ± 11.7
CEERS3_1748	-5.6 ± 4.4	5.7 ± 5.7	18.5 ± 7.9	-1.6 ± 13.9	1.4 ± 6.9
CEERS2_1298	0.3 ± 4.2	-0.3 ± 4.6	6.4 ± 5.6	7.0 ± 9.7	17.5 ± 4.9
CEERS2_2274	-10.3 ± 6.6	0.9 ± 8.0	1.1 ± 8.5	22.4 ± 10.6	18.6 ± 7.8
CEERS2_1075	-4.8 ± 4.0	-6.2 ± 5.7	3.0 ± 5.8	18.3 ± 10.4	8.6 ± 5.2

Note. All fluxes are in nJy. The horizontal lines distinguish our $z > 12$, 10–12, and 8.5–10 samples. We do not include a column for F105W, as none of these candidate galaxies were covered by the sparse amount of F105W imaging in this field.

Appendix B
$z \sim 9$ Sample Plots

Here we show the cutout images (Figures 16 and 17) and SED plots (Figure 18) for the $z \sim 9$ sample as described in Section 5.3, presented here for clarity in the main text.

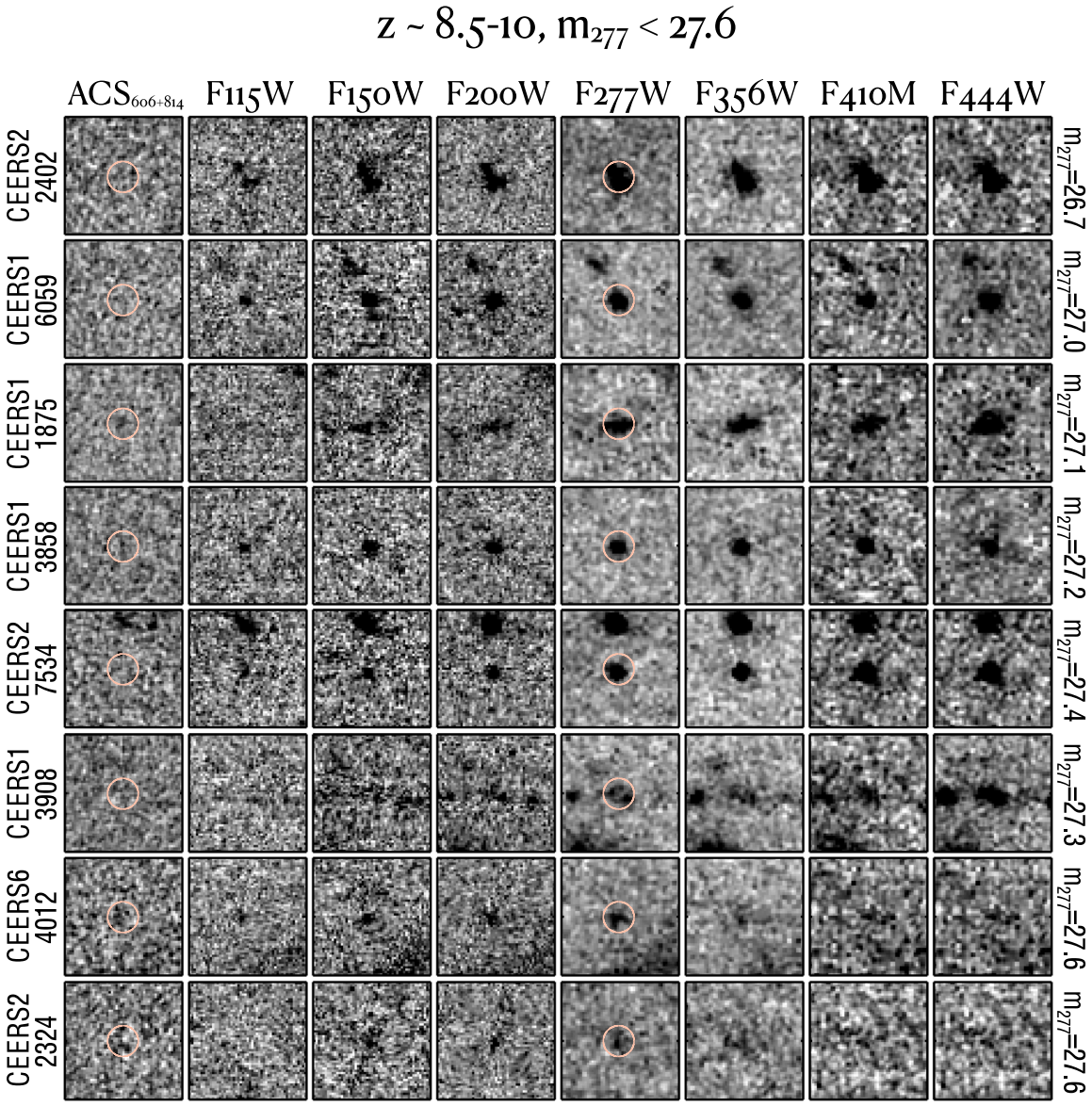

Figure 16. Similar to Figure 5, showing the eight candidates with best-fit photometric redshifts of $z \sim 9$ and an F277W magnitude brighter than 27.6.

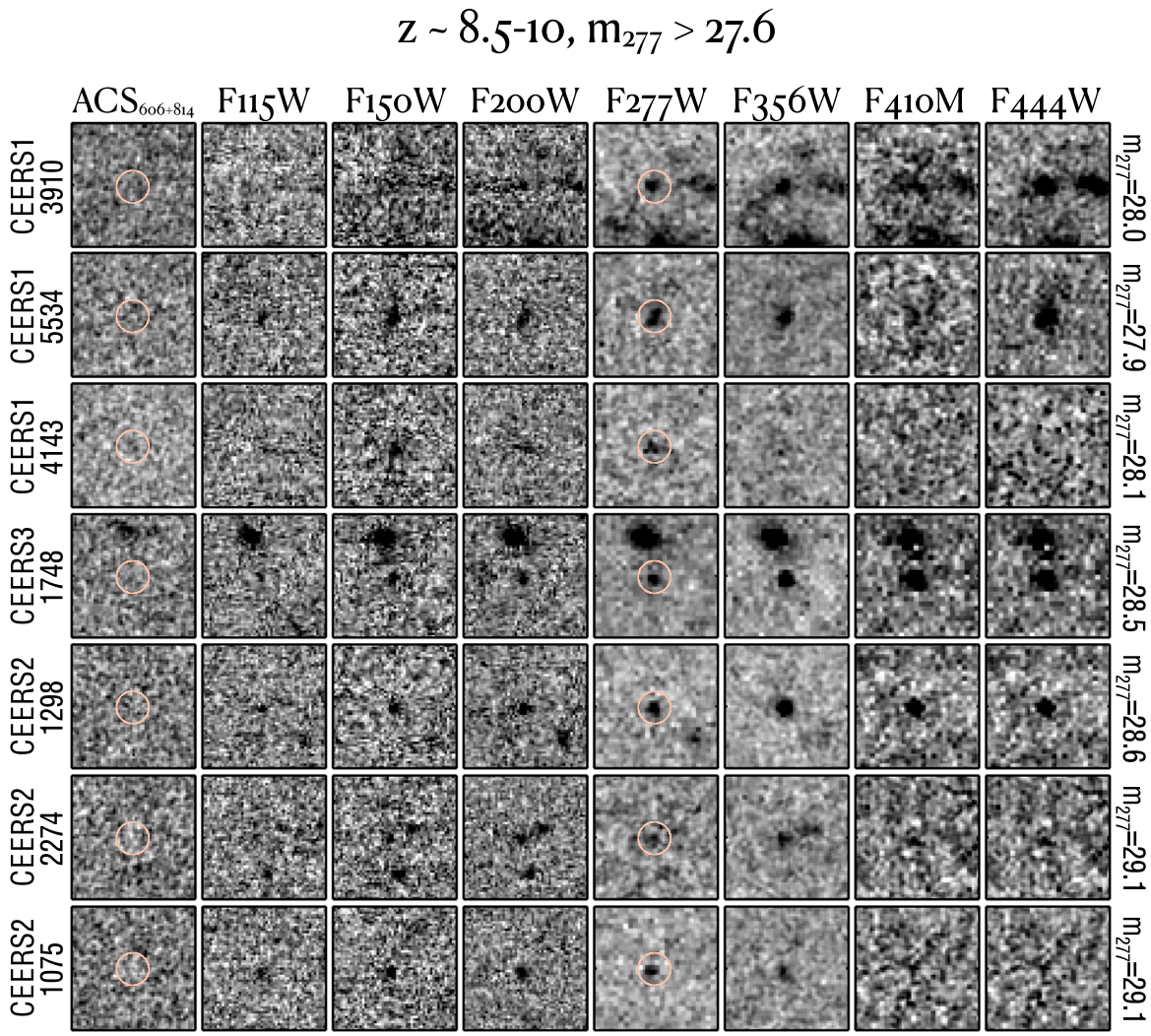

Figure 17. Similar to Figure 5, showing the nine candidates with best-fit photometric redshifts of $z \sim 9$ and an F277W magnitude fainter than 27.6.

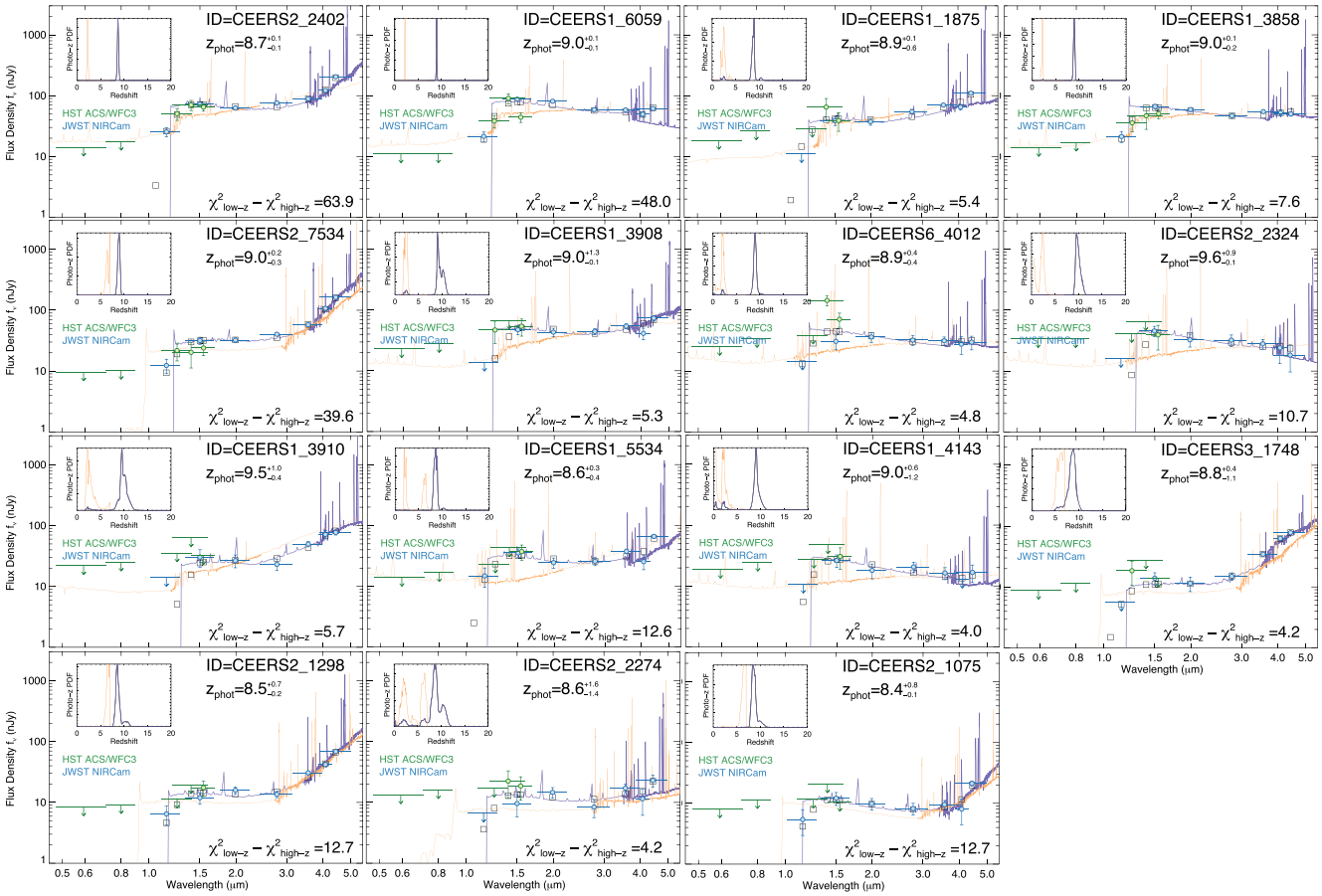

Figure 18. The SEDs and $\mathcal{P}(z)$s of the $z \sim 9$ sample.

Appendix C
Sources Removed from the Sample

Figure 19 shows two objects removed from our sample after remeasuring the colors in smaller apertures due to the stretching of their Kron apertures by nearby bright sources. Figure 20 shows cutout images for the 36 sources removed as spurious sources following the visual inspection described in Section 4.3. We list the coordinates for all removed sources in Table 6.

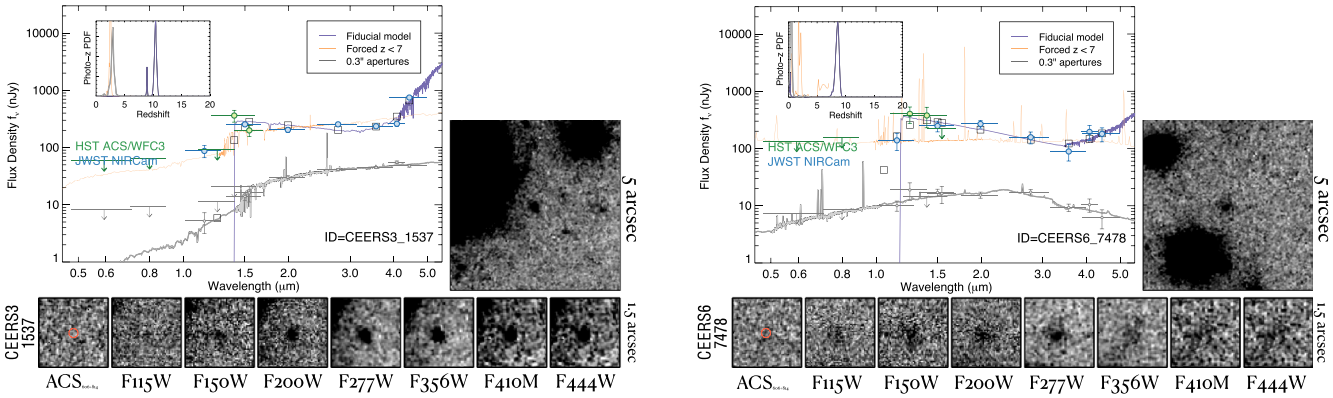

Figure 19. Two sources that had incorrectly drawn Kron apertures due to the presence of nearby bright sources. The large plot shows the SED, with the Kron NIRCam (HST) photometry in blue (green). The best-fit EAZY model to these data is shown in purple (with squares denoting the model bandpass-averaged fluxes), with the best-fit $z < 7$ solution shown in orange. The small gray circles show the fluxes measured in $0\farcs3$ diameter apertures, with the gray line showing the EAZY fit to these small-aperture fluxes. The $P(z)$s for all three EAZY runs are shown in the top left. The large image shows a $5'' \times 5''$ cutout around each source, highlighting the nearby neighbor responsible for stretching the Kron aperture, leading to much brighter Kron fluxes. The small images show a $1\farcs5$ region around each source in the seven NIRCam bands (and the stacked ACS dropout bands), with the red circle denoting a $0\farcs3$ diameter region around the source. The $P(z)$ from the colors measured in this small circular aperture prefers lower redshifts in both cases; thus, these sources were removed from the sample.

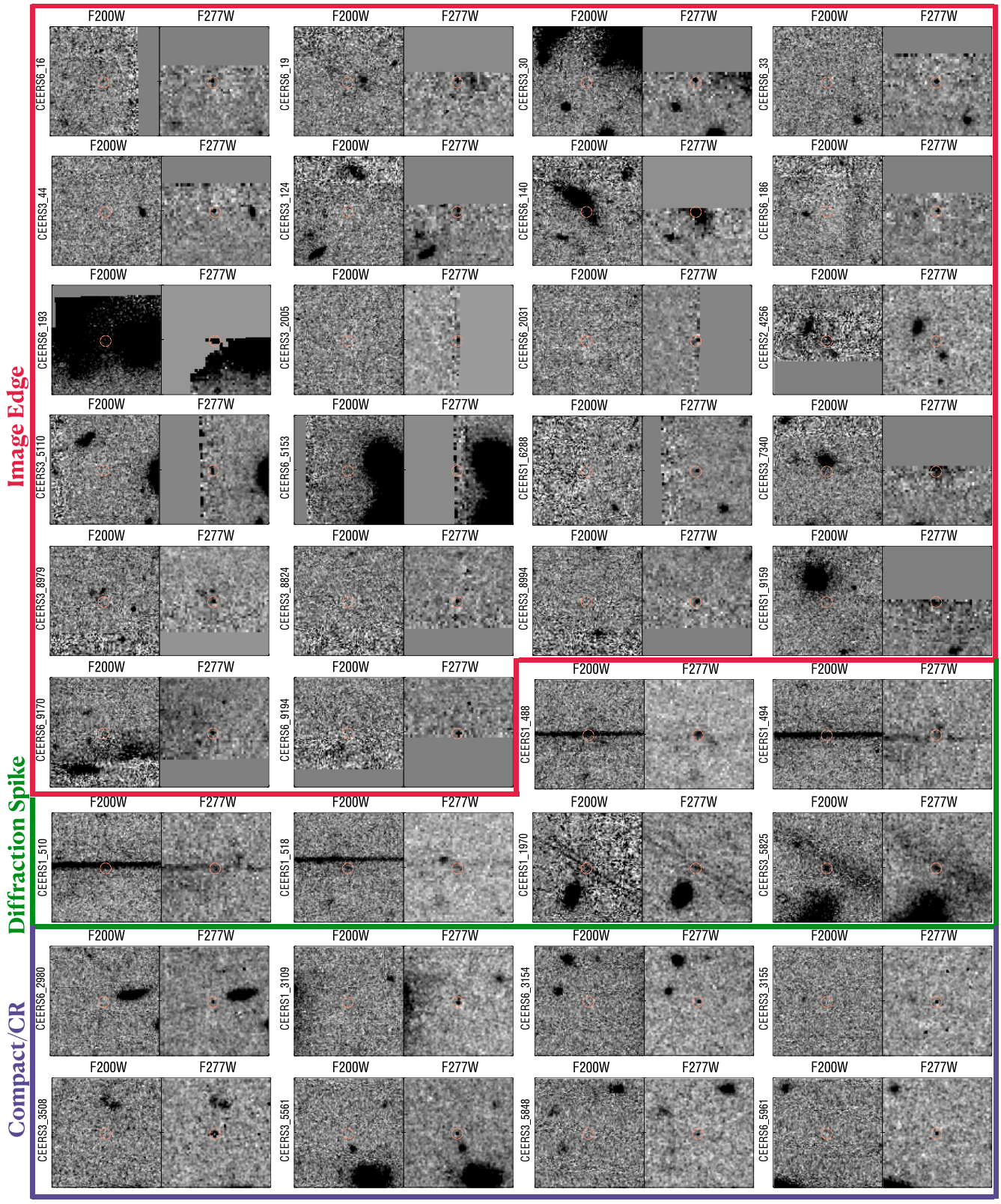

Figure 20. This compilation shows 3″ × 3″ cutout images of the 36 sources identified as spurious from our visual inspection of the initial list of candidates. The majority of these spurious detections come from very near the image edges, which are easily identifiable (and can be automated in the future). Six objects are obvious diffraction spikes. The remaining eight sources are very compact and boxy and visible only in the LW channels. We conclude that these are highly likely to be unflagged cosmic rays, which will be better flagged in future reductions (M. Bagley et al. 2022, in preparation). We note that seven out of eight are in the CEERS3 or 6 pointings, which had fewer images with longer exposure times than CEERS1 and 2, leading to more numerous cosmic-ray hits.

Table 6
List of Removed Sources

ID	R.A.	Decl.	Field	Reason
488	215.005247	53.017772	1	Diffraction spike
494	215.006368	53.018553	1	Diffraction spike
510	215.006574	53.018658	1	Diffraction spike
518	215.005202	53.017651	1	Diffraction spike
1970	214.945032	52.966214	1	Diffraction spike
3109	214.937432	52.953783	1	CR residual
6288	214.985433	52.968347	1	Image edge
9159	214.947769	52.980439	1	Image edge
4256	214.878918	52.904385	2	Image edge
30	214.735947	52.832000	3	Image edge
44	214.768557	52.855112	3	Image edge
124	214.798821	52.875759	3	Image edge
1537	214.750580	52.829452	3	Bad Kron
2005	214.797283	52.862516	3	Image edge
3155	214.758502	52.829340	3	CR residual
3508	214.775758	52.838390	3	CR residual
5110	214.798858	52.844780	3	Image edge
5561	214.779455	52.827964	3	CR residual
5825	214.775453	52.823445	3	Diffraction spike
5848	214.780253	52.826894	3	CR residual
7340	214.749582	52.841873	3	Image edge
8824	214.811070	52.829804	3	CR residual
8979	214.834667	52.845385	3	CR residual
8994	214.861852	52.864635	3	CR residual
16	214.806895	52.826457	6	Image edge
19	214.788359	52.813215	6	Image edge
33	214.820973	52.836319	6	Image edge
140	214.864192	52.866051	6	Image edge
186	214.871337	52.870820	6	Image edge
193	214.878573	52.875993	6	Image edge
2031	214.849667	52.843711	6	Image edge
2980	214.811286	52.810462	6	CR residual
3154	214.811199	52.809295	6	CR residual
5153	214.851208	52.826106	6	Image edge
5961	214.832642	52.808095	6	CR residual
7478	214.851426	52.812204	6	Bad Kron
9170	214.914199	52.845844	6	Image edge
9194	214.897088	52.833510	6	Image edge

Note. ID and coordinates of sources shown in Figures 19 and 20.

ORCID iDs

Steven L. Finkelstein https://orcid.org/0000-0001-8519-1130
Micaela B. Bagley https://orcid.org/0000-0002-9921-9218
Henry C. Ferguson https://orcid.org/0000-0001-7113-2738
Stephen M. Wilkins https://orcid.org/0000-0003-3903-6935
Jeyhan S. Kartaltepe https://orcid.org/0000-0001-9187-3605
Casey Papovich https://orcid.org/0000-0001-7503-8482
L. Y. Aaron Yung https://orcid.org/0000-0003-3466-035X
Pablo Arrabal Haro https://orcid.org/0000-0002-7959-8783
Peter Behroozi https://orcid.org/0000-0002-2517-6446
Mark Dickinson https://orcid.org/0000-0001-5414-5131
Dale D. Kocevski https://orcid.org/0000-0002-8360-3880
Anton M. Koekemoer https://orcid.org/0000-0002-6610-2048
Rebecca L. Larson https://orcid.org/0000-0003-2366-8858
Aurélien Le Bail https://orcid.org/0000-0002-9466-2763
Alexa M. Morales https://orcid.org/0000-0003-4965-0402
Pablo G. Pérez-González https://orcid.org/0000-0003-4528-5639
Denis Burgarella https://orcid.org/0000-0002-4193-2539
Romeel Davé https://orcid.org/0000-0003-2842-9434
Michaela Hirschmann https://orcid.org/0000-0002-3301-3321
Rachel S. Somerville https://orcid.org/0000-0002-6748-6821
Stijn Wuyts https://orcid.org/0000-0003-3735-1931
Volker Bromm https://orcid.org/0000-0003-0212-2979
Caitlin M. Casey https://orcid.org/0000-0002-0930-6466
Adriano Fontana https://orcid.org/0000-0003-3820-2823
Seiji Fujimoto https://orcid.org/0000-0001-7201-5066
Jonathan P. Gardner https://orcid.org/0000-0003-2098-9568
Mauro Giavalisco https://orcid.org/0000-0002-7831-8751
Andrea Grazian https://orcid.org/0000-0002-5688-0663
Norman A. Grogin https://orcid.org/0000-0001-9440-8872
Nimish P. Hathi https://orcid.org/0000-0001-6145-5090
Taylor A. Hutchison https://orcid.org/0000-0001-6251-4988
Saurabh W. Jha https://orcid.org/0000-0001-8738-6011
Shardha Jogee https://orcid.org/0000-0002-1590-0568
Lisa J. Kewley https://orcid.org/0000-0001-8152-3943
Allison Kirkpatrick https://orcid.org/0000-0002-1306-1545
Arianna S. Long https://orcid.org/0000-0002-7530-8857
Jennifer M. Lotz https://orcid.org/0000-0003-3130-5643
Laura Pentericci https://orcid.org/0000-0001-8940-6768
Justin D. R. Pierel https://orcid.org/0000-0002-2361-7201
Nor Pirzkal https://orcid.org/0000-0003-3382-5941
Swara Ravindranath https://orcid.org/0000-0002-5269-6527
Russell E. Ryan, Jr. https://orcid.org/0000-0003-0894-1588
Jonathan R. Trump https://orcid.org/0000-0002-1410-0470
Guang Yang https://orcid.org/0000-0001-8835-7722
Rachana Bhatawdekar https://orcid.org/0000-0003-0883-2226
Laura Bisigello https://orcid.org/0000-0003-0492-4924
Véronique Buat https://orcid.org/0000-0003-3441-903X
Antonello Calabrò https://orcid.org/0000-0003-2536-1614
Marco Castellano https://orcid.org/0000-0001-9875-8263
Nikko J. Cleri https://orcid.org/0000-0001-7151-009X
M. C. Cooper https://orcid.org/0000-0003-1371-6019
Darren Croton https://orcid.org/0000-0002-5009-512X
Emanuele Daddi https://orcid.org/0000-0002-3331-9590
Avishai Dekel https://orcid.org/0000-0003-4174-0374
David Elbaz https://orcid.org/0000-0002-7631-647X
Maximilien Franco https://orcid.org/0000-0002-3560-8599
Eric Gawiser https://orcid.org/0000-0003-1530-8713
Benne W. Holwerda https://orcid.org/0000-0002-4884-6756
Marc Huertas-Company https://orcid.org/0000-0002-1416-8483
Anne E. Jaskot https://orcid.org/0000-0002-6790-5125
Gene C. K. Leung https://orcid.org/0000-0002-9393-6507
Ray A. Lucas https://orcid.org/0000-0003-1581-7825
Bahram Mobasher https://orcid.org/0000-0001-5846-4404
Viraj Pandya https://orcid.org/0000-0002-2499-9205
Sandro Tacchella https://orcid.org/0000-0002-8224-4505
Benjamin J. Weiner https://orcid.org/0000-0001-6065-7483
Jorge A. Zavala https://orcid.org/0000-0002-7051-1100

References

Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755
Aretxaga, I. 2015, IAUGA, 29, 2258051
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, ApJS, 218, 33
Bagley, M., Finkelstein, S. L., Finkelstein, S. L., et al. 2022a, ApJ, submitted, arXiv:2205.12980
Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2022b, arXiv:2211.02495
Bakx, T. J. L. C., Zavala, J. A., Mitsuhashi, I., et al. 2023, MNRAS, 519, 5076
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Behroozi, P., Conroy, C., Wechsler, R. H., et al. 2020, MNRAS, 499, 5702
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., & Silk, J. 2015, ApJ, 799, 32
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013, ApJ, 763, 18
Bernard, S. R., Carrasco, D., Trenti, M., et al. 2016, ApJ, 827, 76
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bhowmick, A. K., Somerville, R. S., Di Matteo, T., et al. 2020, MNRAS, 496, 754
Bialy, S., & Sternberg, A. 2019, ApJ, 881, 160
Boucaud, A., Bocchio, M., Abergel, A., et al. 2016, PyPHER: Python-based PSF Homogenization kERnels, Astrophysics Source Code Library, record ascl:1609.022
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, ApJL, 709, L133
Bouwens, R. J., Illingworth, G. D., & Oesch, P. A. 2015, ApJ, 803, 34
Bouwens, R. J., Illingworth, G. D., van Dokkum, P. G., et al. 2022a, ApJ, 927, 81
Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, AJ, 162, 47
Bouwens, R. J., Stefanon, M., Brammer, G., et al. 2022b, arXiv:2211.02607
Bouwens, R. J., Stefanon, M., Oesch, P. A., et al. 2019, ApJ, 880, 25
Bouwens, R. J., Thompson, R. I., Illingworth, G. D., et al. 2004, ApJL, 616, L79
Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., et al. 2020, MNRAS, 493, 2059
Boyer, M. L., Anderson, J., Gennaro, M., et al. 2022, RNAAS, 6, 191
Boylan-Kolchin, M. 2022, arXiv:2208.01611
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bridge, J. S., Holwerda, B. W., Stefanon, M., et al. 2019, ApJ, 882, 42
Bromm, V., Kudritzki, R. P., & Loeb, A. 2001, ApJ, 552, 464
Bromm, V., & Larson, R. B. 2004, ARA&A, 42, 79
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413
Burgasser, A. J. 2014, in Astronomical Society of India Conference Series, The SpeX Prism Library: 1000+ low-resolution, near-infrared spectra of ultracool M, L, T and Y dwarfs, 11, ed. H. P. Singh, P/ Prugniel, & I. Vauglin, 7
Caballero, J. A., Burgasser, A. J., & Klement, R. 2008, A&A, 488, 181
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Casertano, S., de Mello, D., Dickinson, M., et al. 2000, AJ, 120, 2747
Castellano, M., Fontana, A., Treu, T., et al. 2022, ApJL, 938, L15
Chon, S., Omukai, K., & Schneider, R. 2021, MNRAS, 508, 4175
Clarke, C. J., & Bromm, V. 2003, MNRAS, 343, 1224
Conroy, C., & Gunn, J. E. 2010, FSPS: Flexible Stellar Population Synthesis, Astrophysics Source Code Library, record ascl:1010.043
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1
Dayal, P., Choudhury, T. R., Bromm, V., & Pacucci, F. 2017, ApJ, 836, 16
Dayal, P., & Ferrara, A. 2018, PhR, 780, 1
Dickinson, M. & FIDEL Team 2007, AAS Meeting Abstracts, 211, 52.16
Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2023, MNRAS, 518, 6011
Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, MNRAS, 420, 901
Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019
Endsley, R., Stark, D. P., Whitler, L., et al. 2022, arXiv:2208.14999
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Ferrara, A., Pallottini, A., & Dayal, P. 2022, arXiv:2208.00720
Finkelstein, S. L. 2016, PASA, 33, e037
Finkelstein, S. L., Bagley, M., Song, M., et al. 2022a, ApJ, 928, 52
Finkelstein, S. L., & Bagley, M. B. 2022, ApJ, 938, 25
Finkelstein, S. L., Bagley, M. B., Arrabal Haro, P., et al. 2022b, ApJL, 940, L44
Finkelstein, S. L., D'Aloisio, A., Paardekooper, J. P., et al. 2019, ApJ, 879, 36
Finkelstein, S. L., Papovich, C., Giavalisco, M., et al. 2010, ApJ, 719, 1250
Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015, ApJ, 810, 71
Fruchter, A. S., & Hook, R. N. 2002, PASP, 114, 144
Fujimoto, S., Finkelstein, S. L., Burgarella, D., et al. 2022, arXiv:2211.03896
Gawiser, E., van Dokkum, P. G., Herrera, D., et al. 2006, ApJS, 162, 1
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gutcke, T. A., Pakmor, R., Naab, T., & Springel, V. 2022, MNRAS, 513, 1372
Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, ApJS, 265, 5
Holwerda, B. W., Trenti, M., Clarkson, W., et al. 2014, ApJ, 788, 77
Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, ApJ, 799, 12
Jaacks, J., Finkelstein, S. L., & Bromm, V. 2019, MNRAS, 488, 2202
Jiang, L., Kashikawa, N., Wang, S., et al. 2021, NatAs, 5, 256
Jin, S., Daddi, E., Liu, D., et al. 2018, ApJ, 864, 56
Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864
Kannan, R., Garaldi, E., Smith, A., et al. 2022, MNRAS, 511, 4005
Kannan, R., Springel, V., Hernquist, L., et al. 2022, arXiv:2210.10066
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Krumholz, M. R., & Dekel, A. 2012, ApJ, 753, 16
Labbe, I., van Dokkum, P., Nelson, E., et al. 2022, arXiv:2207.12446
Larson, R. B. 1998, MNRAS, 301, 569
Larson, R. L., Finkelstein, S. L., Hutchison, T. A., et al. 2022a, ApJ, 930, 104
Larson, R. L., Hutchison, T. A., Bagley, M., et al. 2022b, arXiv:2211.10035
Liu, D., Daddi, E., Dickinson, M., et al. 2018, ApJ, 853, 172
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Ma, X., Hopkins, P. F., Garrison-Kimmel, S., et al. 2018, MNRAS, 478, 1694
Mason, C. A., Trenti, M., & Treu, T. 2023, MNRAS, Advance Access
Mason, C. A., Trenti, M., Treu, T., et al. 2015, ApJ, 813, 21
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, MNRAS, 450, 3032
McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, MNRAS, 459, 3812
McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Morishita, T., Trenti, M., Stiavelli, M., et al. 2018, ApJ, 867, 150
Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59
Naidu, R. P., Oesch, P. A., Setton, D. J., et al. 2022a, arXiv:2208.02794
Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022b, ApJL, 940, L14
Nandra, K., Laird, E. S., Aird, J. A., et al. 2015, ApJS, 220, 10
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2010, ApJL, 709, L16
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, ApJ, 855, 105
Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614
Papovich, C., Shipley, H. V., Mehrtens, N., et al. 2016, ApJS, 224, 28
Patten, B. M., Stauffer, J. R., Burrows, A., et al. 2006, ApJ, 651, 502
Pawlik, A. H., Milosavljević, M., & Bromm, V. 2011, ApJ, 731, 54
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H. W. 2002, AJ, 124, 266
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H. W. 2010, AJ, 139, 2097
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Raiter, A., Schaerer, D., & Fosbury, R. A. E. 2010, A&A, 523, A64
Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Rieke, M. J., Kelly, D., & Horner, S. 2005, Proc. SPIE, 5904, 1
Robertson, B. E. 2022, ARA&A, 60, 121
Rodney, S. A., Riess, A. G., Strolger, L. G., et al. 2014, AJ, 148, 13
Rojas-Ruiz, S., Finkelstein, S. L., Bagley, M. B., et al. 2020, ApJ, 891, 146
Ryan, R. E., Jr., & Reid, I. N. 2016, AJ, 151, 92
Ryan, R. E., Thorman, P. A., Yan, H., et al. 2011, ApJ, 739, 83
Ryan, R. E. J., Hathi, N. P., Cohen, S. H., & Windhorst, R. 2005, ApJL, 631, L159
Sharda, P., & Krumholz, M. R. 2022, MNRAS, 509, 1959
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 31
Somerville, R. S., & Kolatt, T. S. 1999, MNRAS, 305, 1
Sommovigo, L., Ferrara, A., Pallottini, A., et al. 2022, MNRAS, 513, 3122
Sparre, M., & Springel, V. 2016, MNRAS, 462, 2418
Stark, D. P. 2016, ARA&A, 54, 761
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2019, ApJ, 883, 99
Tacchella, S., Finkelstein, S. L., Bagley, M., et al. 2022, ApJ, 927, 170
Trenti, M., Bradley, L. D., Stiavelli, M., et al. 2011, ApJL, 727, L39
Tumlinson, J. 2006, ApJ, 641, 1
Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
Whitler, L., Endsley, R., Stark, D. P., et al. 2023, MNRAS, 519, 157
Wilkins, S. M., Stanway, E. R., & Bremer, M. N. 2014, MNRAS, 439, 1038
Wilkins, S. M., Vijayan, A. P., Lovell, C. C., et al. 2023, MNRAS, 519, 3118

Wilkins, S. M., Vijayan, A. P., & Lovell, C. C. 2022, MNRAS, 517, 3227
Yan, H., Windhorst, R. A., & Cohen, S. H. 2003, ApJL, 585, L93
Yang, L., Morishita, T., Leethochawalit, N., et al. 2022, ApJL, 938, L17
Yoon, I., Carilli, C. L., Fujimoto, S., et al. 2022, arXiv:2210.08413
Yung, L. Y. A., Somerville, R. S., Ferguson, H. C., et al. 2022, MNRAS, 515, 5416
Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., et al. 2020, MNRAS, 496, 4574
Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Popping, G., & Davé, R. 2019a, MNRAS, 483, 2983
Yung, L. Y. A., Somerville, R. S., Popping, G., et al. 2019b, MNRAS, 490, 2855
Zackrisson, E., Rydberg, C. E., Schaerer, D., Östlin, G., & Tuli, M. 2011, ApJ, 740, 13
Zavala, J. A., Aretxaga, I., Geach, J. E., et al. 2017, MNRAS, 464, 3369
Zavala, J. A., Buat, V., Casey, C. M., et al. 2023, ApJL, 943, L9