Publication Year	2023
Acceptance in OA	2025-03-03T13:09:57Z
Title	CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST
Authors	Finkelstein, Steven L., Bagley, Micaela B., Ferguson, Henry C., Wilkins, Stephen M., Kartaltepe, Jeyhan S., Papovich, Casey, Yung, L. Y. Aaron, Haro, Pablo Arrabal, Behroozi, Peter, DICKINSON, MARK, Kocevski, Dale D., Koekemoer, Anton M., Larson, Rebecca L., Le Bail, Aurélien, Morales, Alexa M., Pérez-González, Pablo G., Burgarella, Denis, Davé, Romeel, HIRSCHMANN, Michaela Monika, Somerville, Rachel S., Wuyts, Stijn, Bromm, Volker, Casey, Caitlin M., FONTANA, Adriano, Fujimoto, Seiji, Gardner, Jonathan P., Giavalisco, Mauro, GRAZIAN, Andrea, Grogin, Norman A., Hathi, Nimish P., Hutchison, Taylor A., Jha, Saurabh W., Jogee, Shardha, Kewley, Lisa J., Kirkpatrick, Allison, Long, Arianna S., Lotz, Jennifer M., PENTERICCI, Laura, Pierel, Justin D.R., Pirzkal, Nor, Ravindranath, Swara, Ryan, Russell E., Trump, Jonathan R., Yang, Guang, Bhatawdekar, Rachana, BISIGELLO, Laura, Buat, Véronique, CALABRO', Antonello, CASTELLANO, Marco, Cleri, Nikko J., Cooper, M. C., Croton, Darren, Daddi, Emanuele, Dekel, Avishai, Elbaz, David, Franco, Maximilien, Gawiser, Eric, Holwerda, Benne W., Huertas-Company, Marc, Jaskot, Anne E., Leung, Gene C.K., Lucas, Ray A., Mobasher, Bahram, Pandya, Viraj, Tacchella, Sandro, Weiner, Benjamin J., Zavala, Jorge A.
Publisher's version (DOI)	10.3847/2041-8213/acade4
Handle	http://hdl.handle.net/20.500.12386/36377
Journal	THE ASTROPHYSICAL JOURNAL LETTERS
Volume	946

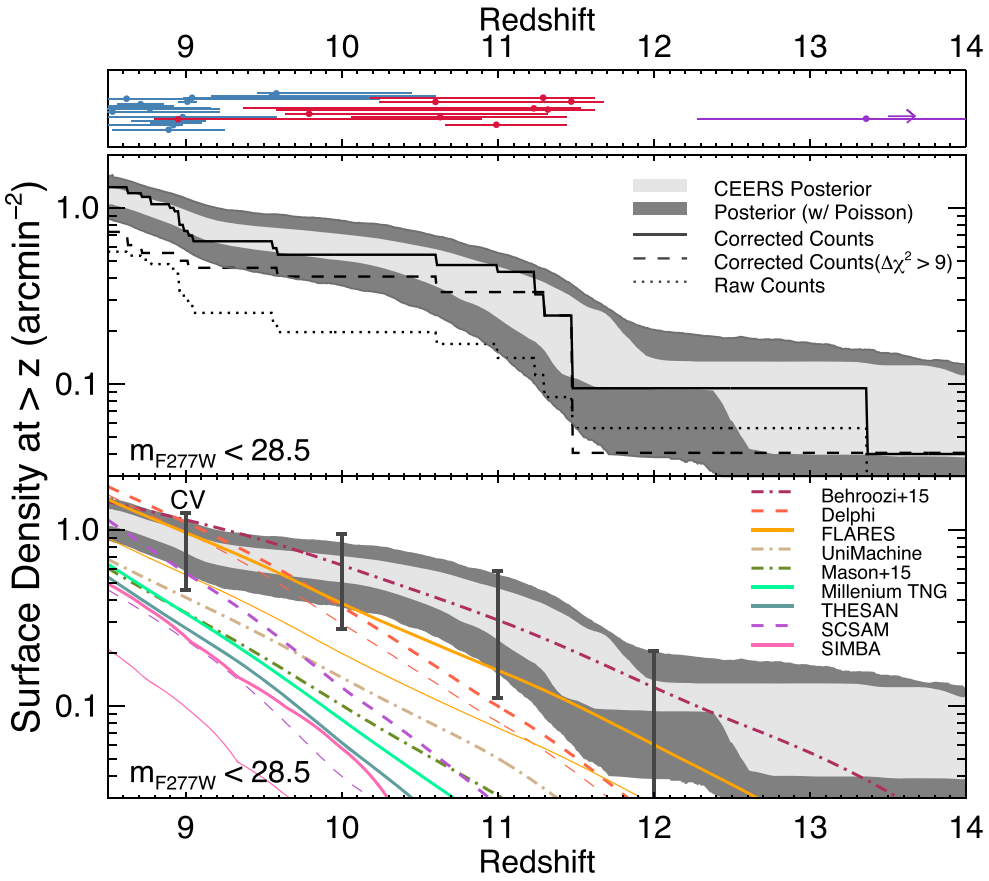

Figure 14. Cumulative surface density of sources with $m_{F277W} < 28.5$ at redshifts greater than a given x-axis value, starting at $z \geqslant 8.5$. The top panel shows the redshifts of individual objects (with blue, red, and purple denoting the $z \sim 9$, 11, and >12 samples). In the middle panel, 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 ~ 1.5 Gyr. This timescale appears to be shorter at higher redshifts, as short as ~ 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 $z > 8.5$.

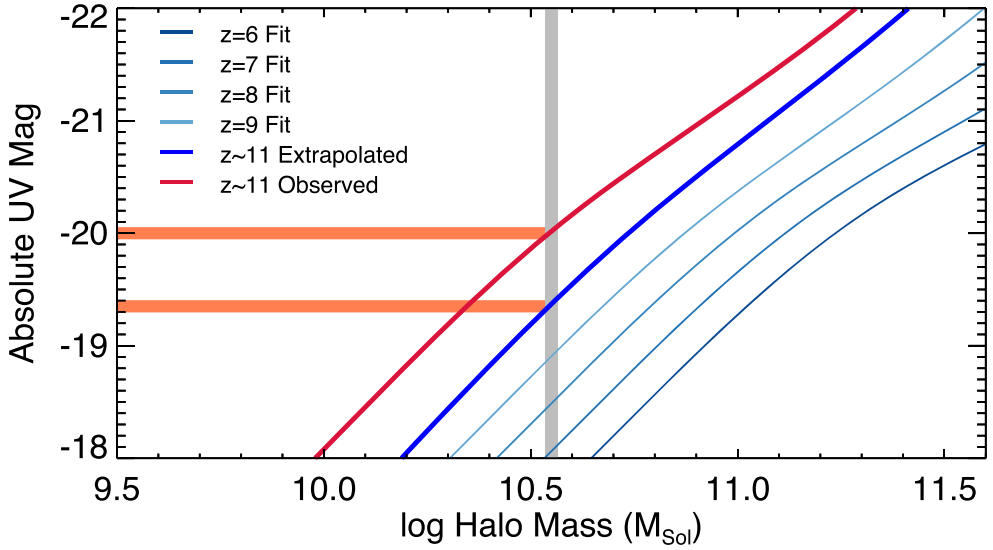

Figure 15. Relation between rest-UV absolute magnitude and halo mass, obtained via abundance matching the observed UV luminosity functions assuming the DPL fits from Finkelstein & Bagley (2022) for $z = 4$–8 and extrapolated to $z \sim 11$ (the thick blue curve is equivalent to the average of the dark 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 ($0\farcs11$, 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.

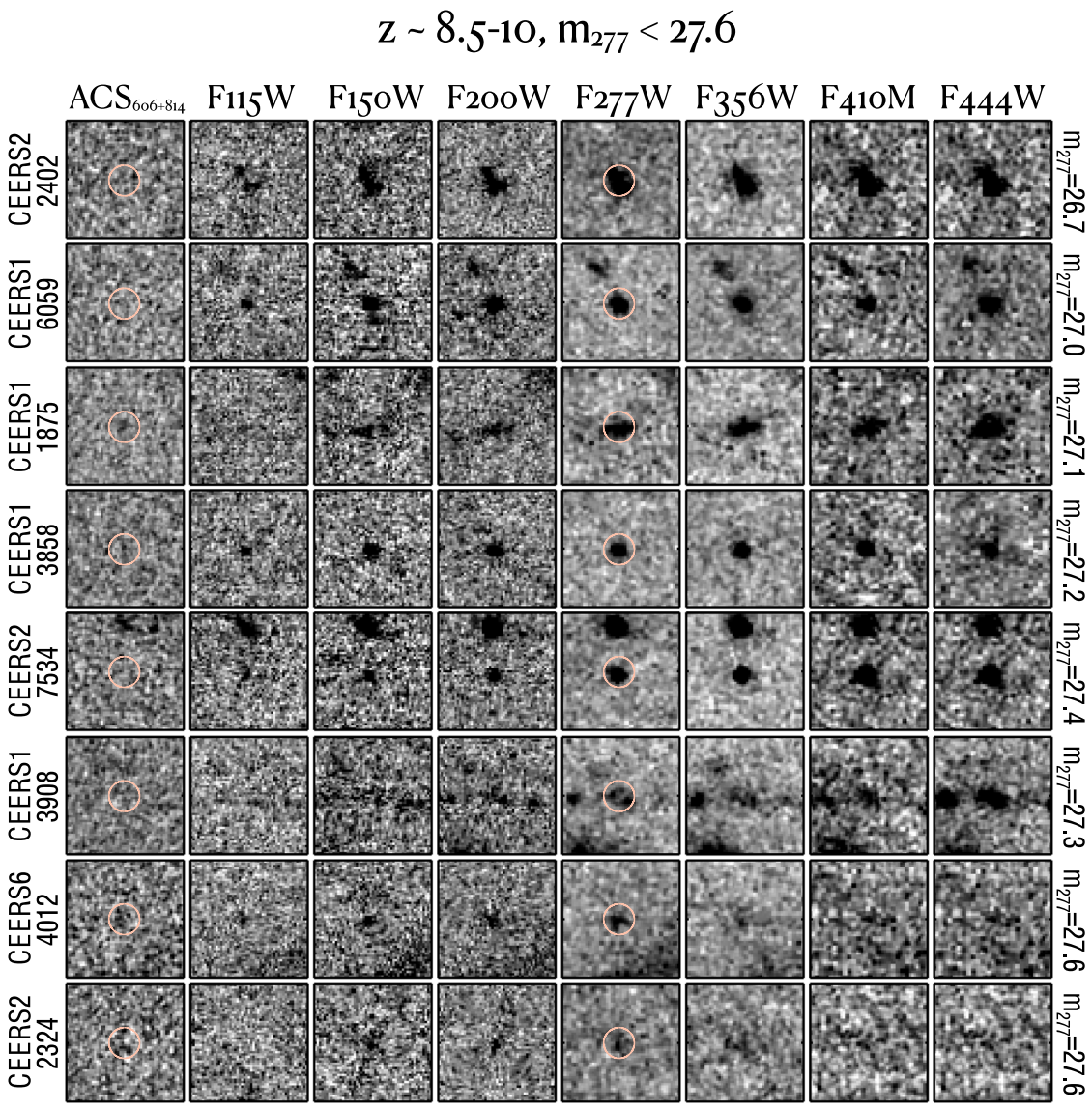

Figure 16. Similar to Figure 5, showing the eight candidates with best-fit photometric redshifts of $z \sim 9$ and an F277W magnitude brighter than 27.6.

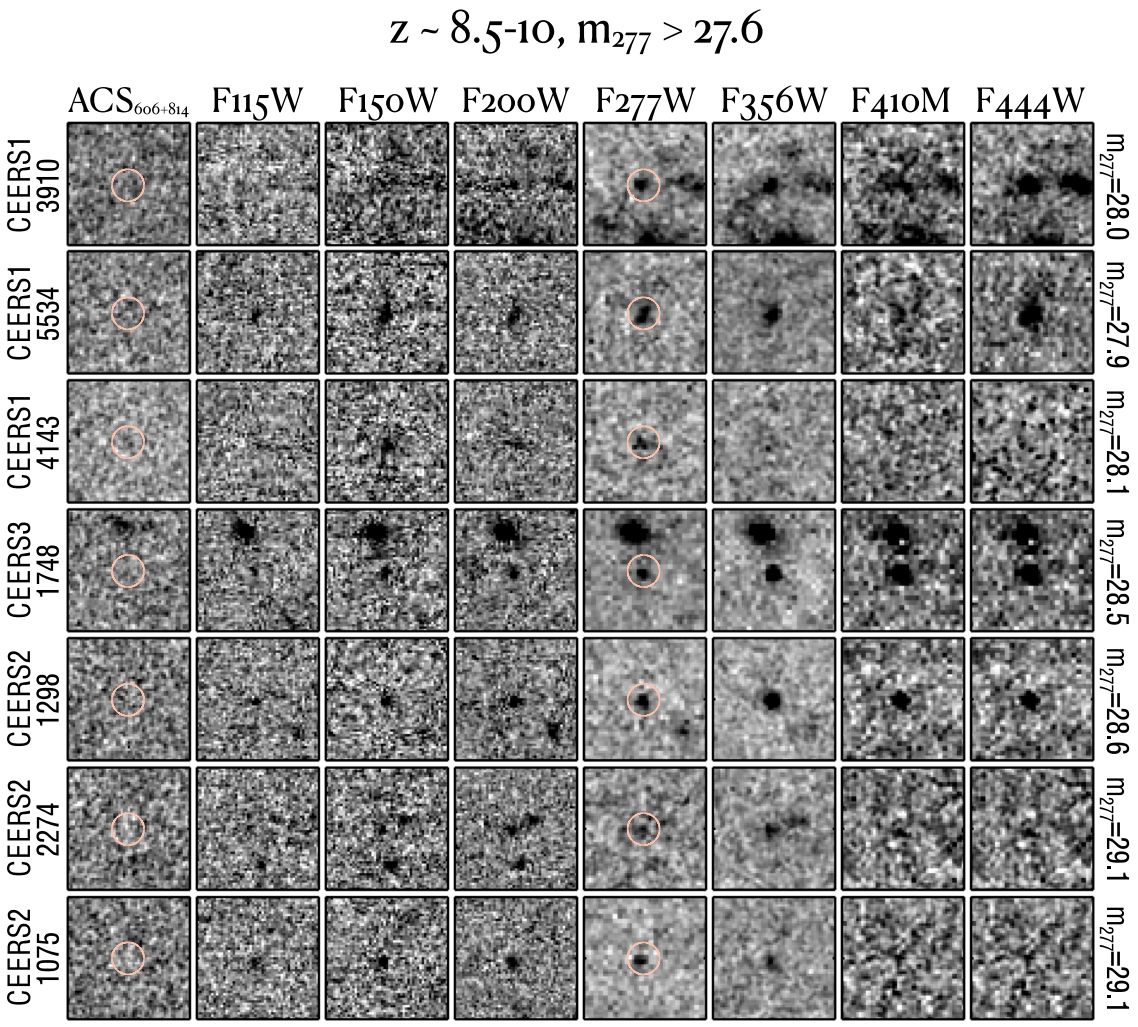

Figure 17. Similar to Figure 5, showing the nine candidates with best-fit photometric redshifts of $z \sim 9$ and an F277W magnitude fainter than 27.6.