The Astrophysical Journal, 955:94 (15pp), 2023 October 1 https://doi.org/10.3847/1538-4357/acec76
© 2023. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

The JWST Hubble Sequence: The Rest-frame Optical Evolution of Galaxy Structure at $1.5 < z < 6.5$

Leonardo Ferreira[1,2], Christopher J. Conselice[3], Elizaveta Sazonova[4], Fabricio Ferrari[5], Joseph Caruana[6,7], Clár-Bríd Tohill[1,8], Geferson Lucatelli[3], Nathan Adams[3], Dimitrios Irodotou[9], Madeline A. Marshall[10,11], Will J. Roper[12], Christopher C. Lovell[13], Aprajita Verma[14], Duncan Austin[3], James Trussler[3], and Stephen M. Wilkins[7,12]

[1] Department of Physics & Astronomy, University of Victoria, Finnerty Road, Victoria, British Columbia, V8P 1A1, Canada; lferreira@uvic.ca
[2] Centre for Astronomy and Particle Theory, University of Nottingham, Nottingham, UK
[3] Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester, UK
[4] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA
[5] Instituto de Matemática Estatística e Física, Universidade Federal do Rio Grande, Rio Grande, RS, Brazil
[6] Department of Physics, University of Malta, Msida MSD 2080, Malta
[7] Institute of Space Sciences and Astronomy, University of Malta, Msida MSD 2080, Malta
[8] Isaac Newton Group of Telescopes, 38700 Santa Cruz de La Palma, Canary Islands, Spain
[9] Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2, FI-00014, Helsinki, Finland
[10] National Research Council of Canada, Herzberg Astronomy & Astrophysics Research Centre, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada
[11] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[12] Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK
[13] Centre for Astrophysics Research, School of Physics, Engineering & Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK
[14] Sub-department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
Received 2022 November 2; revised 2023 July 30; accepted 2023 July 31; published 2023 September 22

Abstract

We present results on the morphological and structural evolution of a total of 3956 galaxies observed with JWST at $1.5 < z < 6.5$ in the JWST CEERS observations that overlap with the CANDELS EGS field. This is the biggest visually classified sample observed with JWST yet, ∼20 times larger than previous studies, and allows us to examine in detail how galaxy structure has changed over this critical epoch. All sources were classified by six individual classifiers using a simple classification scheme aimed at producing disk/spheroid/peculiar classifications, whereby we determine how the relative number of these morphologies has evolved since the Universe's first billion years. Additionally, we explore structural and quantitative morphology measurements using MORFOMETRYKA, and show that galaxies with $M_* > 10^9\ M_\odot$ at $z > 3$ are not dominated by irregular and peculiar structures, either visually or quantitatively, as previously thought. We find a strong dominance of morphologically selected disk galaxies up to $z = 6$ in this mass range. We also find that the stellar mass and star formation rate densities are dominated by disk galaxies up to $z \sim 6$, demonstrating that most stars in the Universe were likely formed in a disk galaxy. We compare our results to theory to show that the fraction of types we find is predicted by cosmological simulations, and that the Hubble Sequence was already in place as early as one billion years after the Big Bang. Additionally, we make our visual classifications public for the community.

Unified Astronomy Thesaurus concepts: Galaxy classification systems (582)

1. Introduction

Since the discovery of galaxies, a principal aim of their study has been to characterize their structures and morphologies. The very fact that galaxies appear to be extended, as opposed to point sources, already provides an elusive clue to their nature being different from that of the stars. In fact, it can be said that it was the extended nature of these objects that instigated the debate about whether they were external to our own Galaxy, a problem solved through obtaining distances to these systems (Reynolds 1920; Hubble 1926; Sandage 2005). Even before then, however, the fact that the structure of galaxies holds important information had been known since at least the time of Lord Rosse and his discovery of spiral structure in nearby massive galaxies such as M51 (Rosse 1850).

Since that time, galaxy structure, morphology, and how these properties evolve with time has remained a key aspect to understanding galaxy evolution (e.g., Conselice 2003, 2014; Lotz et al. 2004; Delgado-Serrano et al. 2010; Mortlock et al. 2013; Schawinski et al. 2014; Whitney et al. 2021; Ferreira et al. 2023). The resolved structure of distant galaxies, in particular with the advent of the Hubble Space Telescope, clearly revealed that faint distant galaxies were more peculiar and irregular, and few fit into the Hubble Sequence (e.g., Driver et al. 1995). Later, once redshifts became available, it became clear that galaxy structures evolve strongly, but systematically, with redshift, such that peculiar galaxies dominate the population at $z > 2.5$ (e.g., Conselice et al. 2008; Conselice 2014; Whitney et al. 2021). However, because of the limited red wavelengths of Hubble, we still have not yet been able to trace the rest-frame optical light of galaxies back to within the first few gigayears of the Big Bang. The F160W band on the Hubble Space Telescope (HST) can only probe rest-frame visible light up to $z \sim 2.8$, but JWST permits us to obtain the same type of data out to $z \sim 9$ with F444W (e.g.,

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Ferreira et al. 2023). Moreover, JWST's superior resolution and longer-wavelength filter set allows galaxy structure to be better measured than with the lower resolution of HST.

Observations of galaxy structure and morphology at $z > 3$ do show that in the rest-frame UV galaxies are peculiar and irregular (e.g., Conselice & Arnold 2009). Moreover, galaxies are often very clumpy at these redshifts, as seen with deep WFC3 data (e.g., Oesch et al. 2010). Furthermore, observations of pairs of galaxies show that the merger rate and fraction of galaxies up to $z \sim 6$ is very high, and therefore that galaxy structure should likewise be affected significantly (e.g., Duncan et al. 2019). At the same time, we know that the Hubble Sequence is already established at lower redshifts (e.g., Delgado-Serrano et al. 2010; Mortlock et al. 2013). However, whether the Hubble Sequence already existed in the earlier Universe remains an open question.

While earlier HST-based studies found that the dominating majority of galaxies at $z > 2$ are peculiar, recent JWST-based studies find a high number of regular disk galaxies at high redshift (e.g., Ferreira et al. 2023; Jacobs et al. 2023; Nelson et al. 2023; Robertson et al. 2023), consistent with an even earlier emergence of the Hubble Sequence. The nature of this evolution and its implications, however, are still unknown.

Quantitative measures of galaxy structure and morphology also present stringent constraints for numerical simulations to reproduce. In recent years, full hydrodynamic simulations (Schaye et al. 2015; Nelson et al. 2019; Lovell et al. 2021; Marshall et al. 2022) have enabled resolved morphologies to be predicted in a self-consistent manner, and recent novel simulation approaches allow these to be tested out to the highest redshifts (Roper et al. 2022). There are a number of difficulties when comparing morphologies between simulations and observations, but simple measures of the abundance of, e.g., disk and elliptical galaxies can provide hints as to the underlying mechanisms leading to morphological evolution. However, what we know from early JWST work is that the morphological and structural features of galaxies at $z > 1$ are much different than what was found with HST (e.g., Ferreira et al. 2020), and therefore a more thorough analysis is needed to address these fundamental problems.

Thus, in this paper we explore the morphological properties of 3956 galaxies observed with JWST through visual galaxy classifications and quantitative morphology, from $z = 1.5$ to 6. This sample is 20 times larger than any previous morphological and structural study using JWST. Among other things, we confirm that these early galaxies with masses $M_* \geqslant 10^9\ M_\odot$ have predominantly disk morphologies, and that the Hubble Sequence appears to already be established in some capacity as early as $z \sim 6$. However surprising, these results are still compatible with morphological fractions extracted from cosmological simulations of the distant Universe (Dekel et al. 2020; Park et al. 2022), as well as with the picture that galaxy disks can re-form after gas-rich major mergers (Hammer et al. 2005, 2009; Hopkins et al. 2009, 2010; Stewart et al. 2009; Puech et al. 2012; Sparre & Springel 2017; Peschken et al. 2020). In this paper we discuss these results and their implications for galaxy formation and evolution.

The paper is organized as follows. In Section 2 we describe the data products used, our reduction pipeline, the visual classification scheme adopted, as well as our methods of quantitative morphology measurement. Section 3 describes the results from our classification effort: quantitative morphology measurements including a discussion on the evolution of the Hubble Sequence from $z = 1.5$ to 6.5. We finish with a summary of our main results in Section 4.

2. Data and Methods

We use the public NIRCam JWST observations from the Cosmic Evolution Early Release Science Survey (CEERS; PI: Finkelstein, ID = 1345, Finkelstein et al. 2023), which overlap with the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) on the Extended Groth Strip field (EGS).

These data are reduced independently using a custom setup of the JWST pipeline version 1.6.2 using the on-flight calibration files available through the CRDS 0942; an extensive description is given in Section 2.1. A description of our sample selection based on the CANDELS catalogs is provided in Section 2.2.

Finally, we proceed with two different approaches to these data: first we perform visual classifications for all sources, which is described in detail in Section 2.3. Second, we perform quantitative morphology through MORFOMETRYKA (Ferrari et al. 2015), where we measure nonparametric morphology estimates such as concentration, asymmetry, smoothness (CAS), G–M_{20}, entropy, spirality, sizes, as well as light profile fitting, which is described in detail Section 2.4.

2.1. Data Reduction

We reprocess all of the uncalibrated lower-level JWST data products for this field following our modified version of the JWST official pipeline. This is similar to the process used in Adams et al. (2023) and Ferreira et al. (2023), but with minor updates and improvements, and can be summarized as follows. (1) We use version 1.6.2 of the pipeline with the Calibration Reference Data System (CRDS) version 0942, which was the most up-to-date version at the time of writing. Use of CRDS 0942 is essential for zero-point issues we discuss in Adams et al. (2022). (2) We apply the $1/f$ noise correction derived by Chris Willott to the resulting level 2 data of the JWST pipeline.[15] (3) We extract the sky subtraction step from stage 3 of the pipeline and run it independently on each NIRCam frame, allowing for quicker assessment of the background subtraction performance and fine-tuning. (4) We align calibrated imaging for each individual exposure to GAIA using `tweakreg`, part of the DrizzlePac Python package.[16] (5) We pixel-match the final mosaics with the use of `astropy reproject`.[17] The final resolution of the drizzled images is $0\farcs03$ pixel^{-1}. There is rapid development in the above procedure, and so we anticipate future studies to continue to make refinements to the JWST pipeline. Each one of the four June CEERS observations was processed into individual mosaics.

2.2. Sample Selection

We select 3956 sources in the range $1.5 < z < 6.5$ from the CANDELS EGS catalogs that overlap with the area covered by the initial CEERS observations. We take advantage of the robust photometric redshifts, star formation rates, and stellar masses already derived for CANDELS in previous works (Duncan et al. 2014, 2019; Whitney et al. 2021) to conduct this analysis. Neither is morphological information used for the

[15] https://github.com/chriswillott/jwst
[16] https://github.com/spacetelescope/drizzlepac
[17] https://reproject.readthedocs.io/en/stable/

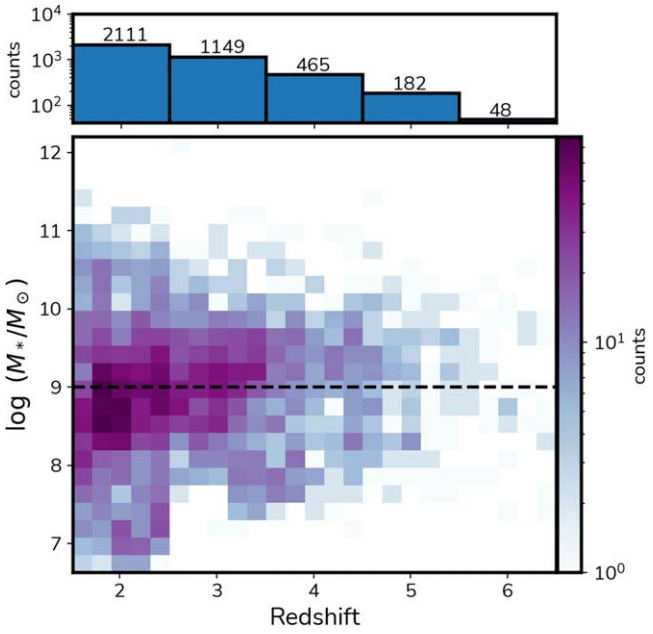

Figure 1. Redshift distribution of the sample (top). Numbers on top of each bar describe the galaxy counts in that redshift bin. Stellar mass–redshift distribution (bottom). The dashed line shows the stellar mass threshold at $\log(M_\odot/M_*) = 9$ used to divide the sample into two stellar mass bins. This threshold is close to the mean stellar mass of the whole distribution.

selection of sources nor are magnitude cuts employed, as we want to make sure that we include sources that might be faint in HST but bright in JWST observations. This is also the case for morphology: we are also interested in sources that can show dramatic changes in morphology between the two instruments.

In Figure 1 we show the redshift distribution of this selection in the top panel and the stellar mass–redshift distribution in the lower panel. The statistics of the sample drops off sharply at higher redshifts, especially for low-mass sources with $\log(M_\odot/M_*) < 9$, due to the detection limit of CANDELS EGS observations. This poses a challenge for a comparison of redshift evolution between low-mass galaxies in the redshift range probed here. However, we separate the sample into two mass bins: the low-mass bin with $\log(M_\odot/M_*) < 9$ and a high-mass bin with $\log(M_\odot/M_*) > 9$. For the high-mass bin the statistics are robust out to the highest redshifts probed, although still volume-limited.

The photometric redshifts that we use in this paper originate from the redshifts calculated in Duncan et al. (2019) for EGS. These are based on the original CANDELS+GOODS WFC3/ACS imaging and data, Spitzer/IRAC S-CANDELS (Ashby et al. 2015), and ground-based observations with CFHT (Stefanon et al. 2017). The overall method for this is described in detail in Duncan et al. (2019). We refer the reader to that paper for a detailed discussion on the mass completeness of the catalogs used here.

We investigate potential observational biases associated with high redshifts in an artificial redshift experiment described in Section 3.3.

2.3. Visual Classification

As a way to define the morphologies of the galaxies in our sample of 3956 sources, we construct a simple classification scheme that yields a large amount of information with a small number of classification questions, as opposed to having a very detailed subclassification scheme of structural subcomponents. The classification scheme is summarized in the flowchart in Figure 2. At high redshift, fine structural details are often difficult to recover and in general are ambiguous, hence these questions capture the overall appearance of the source.

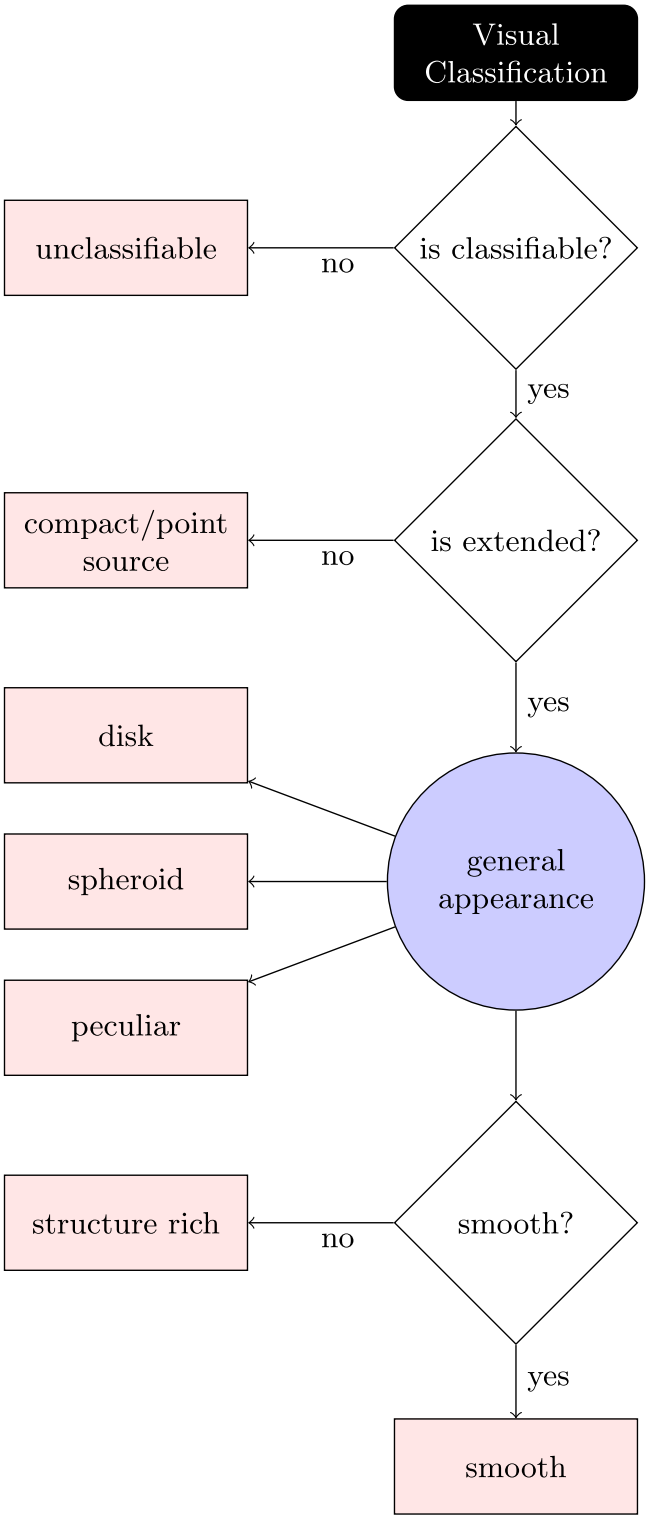

Figure 2. Flowchart of our visual classification process. The classification is based on four basic questions that can produce a simple disk/spheroid/peculiar/compact classification and additional flags regarding whether the source is smooth or structurally rich.

Our sample is the biggest visually classified sample observed with JWST yet, ∼20 times larger than what is reported in previous JWST morphology results (Ferreira et al. 2023; Jacobs et al. 2023; Nelson et al. 2023). A brief description of each possible resulting class is given below:

1. Unclassifiable: galaxies not clearly visible, too faint/noisy, and image issues such as artifacts and cosmic rays.
2. Point sources: sources that are smaller in angular size than the FWHM of the point-spread function (PSF_{FWHM}) or that present clear wings/spike patterns consistent with pointlike objects but no extended component.
3. Disks: galaxies with a resolved disk in the form of an outer area of lower surface brightness with a regularly increasing brightness toward the center of the galaxy. This classification does not depend on there being a feature present in the disk such as a spiral pattern or a bar, although one can be present.
4. Spheroid: resolved, symmetrically and centrally concentrated, with a smooth profile and round/elliptical in shape.
5. Peculiar: well-resolved galaxies with a morphology that is dominated by a disturbance or peculiarity, where the disturbance dominates any smooth components.

Each galaxy is further classified as smooth or structured, where structured galaxies have features standing out from the smooth stellar envelope, such as star formation clumps, tidal features, and merger signatures. Galaxies with distinct disk and bulge components were also classified as structured. Finally, the classifiers were able to provide additional notes on each source to aid in future analysis.

Ultimately, after all classifications were aggregated we determined the final class of each object as the the one receiving the majority of the votes, as discussed in detail in Section 2.3. In the cases where classifiers disagreed, we included an `ambiguous` class.

Based on this classification scheme (Figure 2), six authors of this study (C.B.T., C.C., E.S., J.C., G.L., L.F.) classified all the 3956 sources. This effort produced a robust catalog where every galaxy has all six classifications combined in classification fractions, one for each individual question present in the scheme (Figure 2).

To perform each classification, the classifiers were given access to a web application built with `flask`, `jinja`, and `bootstrap` specifically tailored for this task. The volunteers were presented with the rest-frame image in the filter that corresponds to the source redshift (minimizing for $\lambda_{rest}/(1+z)$), an RGB image (F277W+F356W+F444W) generated with TRILOGY (Coe et al. 2015), the PSF image of the respective filter together with the size of PSF_{FWHM}, and a questionnaire that reproduces Figure 2. Results are stored in a `MySQL` database, which is then reduced and aggregated with `pandas` (Reback 2022).

Each individual classifier's results are combined in a single table using the following criteria.

First, we define how many votes there are for each source—that is, the votes that are not considered to be `unclassifiable` and `point source`. Then, if at least 50% of all votes are assigned to any of these individual classes (i.e., three or more votes are exclusively in these categories), we consider the source to be unclassifiable or pointlike, by the labels `n/a` and `ps`, respectively. For each pointlike source, we compare its size to PSF_{FWHM}. If it is larger than PSF_{FWHM}, we change its classification to `spheroid`. The same is done the other way around: sources smaller than PSF_{FWHM} are changed programmatically to `ps`.

Second, for all the rest of the sources that have more than 50% of the votes in disk, spheroid, or peculiar categories, we average each individual classification decision in a class fraction. Hence, this class fraction is only based on the number of good votes (i.e., votes that are for classifiable and extended galaxies).

Third, to all galaxies that have a clear majority as $frac > 0.5$, we assign the given class as the final class. For all the remaining cases that the classifiers disagree on (e.g., two votes in each category), we define those sources to have an `ambiguous` class.

Finally, we include a structure index, `smooth_fraction`, that is independent of the general appearance, designed as a large umbrella to capture sources with rich structures, subcomponents, and merging features.

This framework enables sources that might be ambiguous between two classifiers to have a more robust classification. The class fractions can also be used to control the purity of the samples, as greater agreement will represent a less contaminated data set. We proceed, however, with the final classifications assigned by the majority of the classifiers.

2.4. MORFOMETRYKA

MORFOMETRYKA (Ferrari et al. 2015) performs several structural measurements on galaxy images in a straightforward, non-interactive way. It measures nonparametric morphometric quantities along with 1D and 2D single Sérsic model fitting (Sérsic 1963). The inputs are the galaxy and PSF images, from which it estimates the background with an iterative algorithm, segments the sources, and defines the target. If desired, it filters out external sources using GalClean[18] (Ferreira & Ferrari 2018). From the segmented region it calculates standard geometrical parameters (e.g., center, position angle, axial ratio) using image moments; it performs photometry by measuring fluxes in similar ellipses with the aforementioned parameters; point sources are masked with a sigma clipping criterion. From the luminosity growth curve it establishes the Petrosian radius and the Petrosian region (Petrosian 1976), inside which all the measurements will be made. The 1D Sérsic fit is performed on the luminosity profile and used as input for a 2D Sérsic fit done with the galaxy and the PSF images. Finally, it measures several morphometric parameters (concentrations; asymmetries; Gini coefficient (G); M_{20}; entropy, spirality, curvature, among others, Abraham et al. 1994; Bershady et al. 2000; Conselice et al. 2003, 2008; Lotz et al. 2004, 2008; Ferrari et al. 2015).

We run MORFOMETRYKA for all filters available but only report results for the band that matches closest the rest-frame optical of the source ($\lambda = 0.5$–0.7 μm).

3. Results

We report the morphology and structural evolution of the sample of 3956 galaxies (Section 2) based on visual classifications (Section 2.3) and in quantitative morphology measurements (Section 2.4). The aggregated classifications

[18] https://github.com/astroferreira/galclean

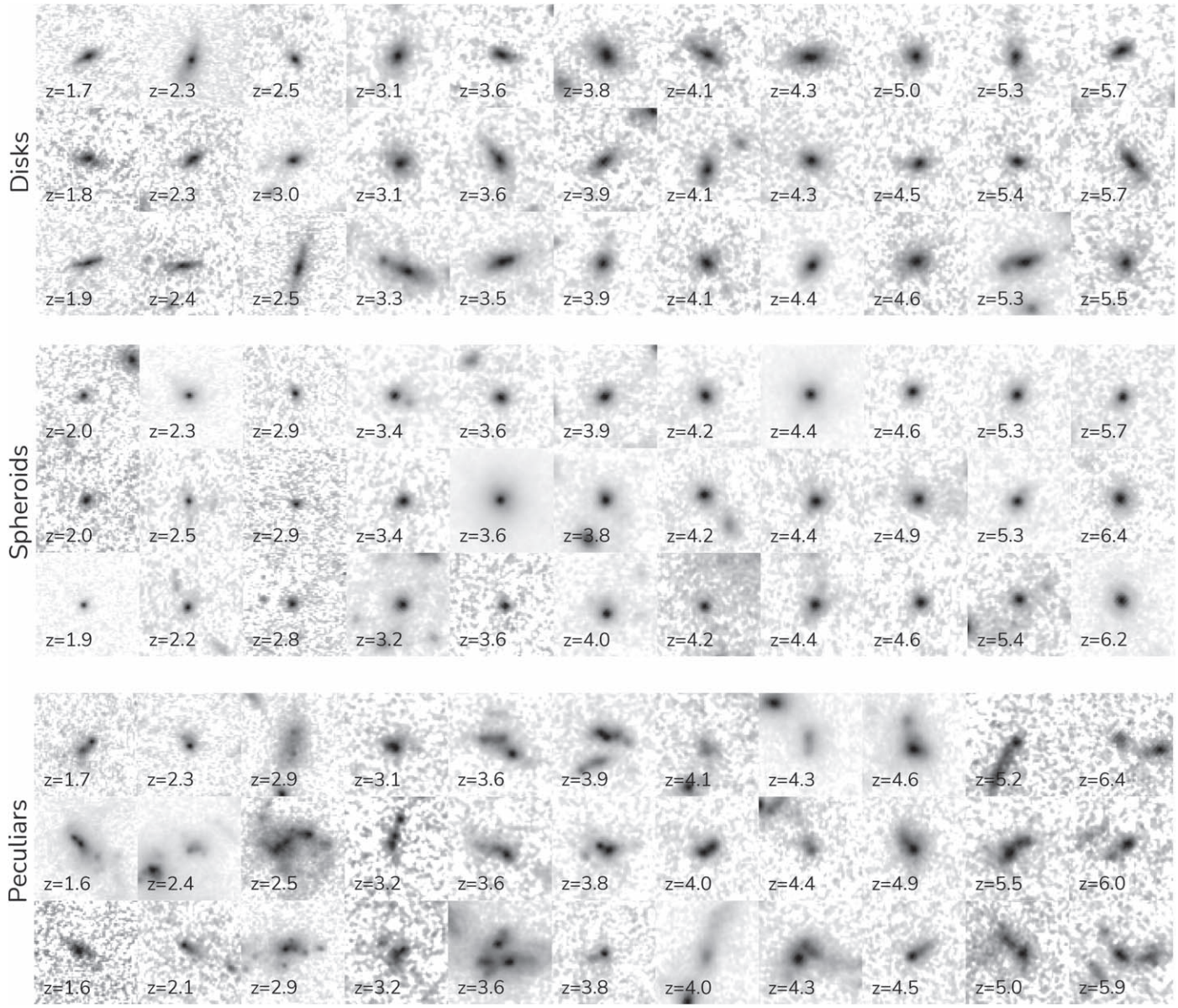

Figure 3. Rest-frame optical images for sources in our sample. The three panels show the three main classes `disks`, `spheroids`, and `peculiars`, respectively. Galaxies are ordered horizontally by redshift, from lowest on the left to highest on the right. Stamps are shown in square-root normalization. Redshifts are from Duncan et al. (2019) based on the CANDELS fields.

catalog based on six independent classifiers contains 1672 disks (∼42%), 553 spheroids (∼14%), 1100 peculiars (∼27%), 428 ambiguous sources (∼10%), 55 point sources (∼1%), and 148 unclassifiable sources (∼4%). Examples of each of these types are shown in Figure 3 in bins of increasing redshift. The full catalog is publicly available.[19]

These visual classifications are the basis for the discussion in this section. In Section 3.1 we detail the three base classes and the caveats from the visual classifications. We follow with a description of the quantitative morphologies of these sources and how they relate to the visual classifications in Section 3.2. We briefly discuss redshift biases in Section 3.3. We explore the evolution of the Hubble Sequence in Section 3.4, and the evolution of the contribution of each morphological class to star formation and stellar mass in Section 3.5. We compare these classifications with predictions from cosmological simulations in Section 3.6, and in Section 3.7 we briefly discuss the main differences between HST and JWST imaging that could explain some of the discrepancies from previous studies. In this section we only show the relative contributions of the three main classes (disks, spheroids, and peculiars), leaving the ambiguous, point sources, and unclassifiable sources out of the picture. At higher redshifts their fraction increases considerably, making it difficult for us to disentangle the contributions from the main classes. However, we discuss how they impact our results separately in Section 3.3.

3.1. Disks, Spheroids, and Peculiars

Figure 3 displays examples randomly drawn from the catalog for the three main morphological classes: disks, spheroids, and peculiars.

[19] Full catalog publicly available at https://github.com/astroferreira/CEERS_EPOCHS_MORPHO/.

The visual distinction between these classes is clear, with the disks often showing two structural components in the form of a concentrated bulge and a disky envelope, while the spheroids are mostly single-profile, centrally concentrated sources, with some exhibiting PSF-like structure due to the central concentration or emission from an active galactic nucleus. However, we note that for most cases, telling apart two types of light concentrations by eye is a difficult task, as sources at high redshifts do not show other clear, disk-like features such as spiral arms, bars, and rings, and overall display lower concentrations (Buitrago et al. 2008, 2013). For a better distinction between face-on disks and spheroids, a quantitative approach such as Sérsic fitting might be used alongside visual classification.

Figure 4 shows three indicators for visual distinctions between the overall sample of `disks` and `spheroids`. `spheroids` are more compact with lower effective radius, higher axis ratios, and lower information entropy (e.g., Shannon entropy) indicating lack of structure. The information entropy describes how the pixel values in an image are distributed: smooth galaxies will present low entropy while clumpy galaxies will have high entropy (see Ferrari et al. 2015 for details on how it is calculated). These distributions follow what is found for high-redshift spheroids in previous studies, in that they are round and smaller than disks (Buitrago et al. 2013). However, the axis ratios found here are at the high end, with a lack of spheroids with intermediate axis ratios $0.4 < b/a \sim 0.7$. Ultimately, some biases might be present, such as the elongation/axis ratio causing some contamination, and thus we possibly miss some elongated spheroids, but this is expected due to each classifier's subjective perspective on what defines these. We advise the user of the catalog to leverage the class fractions to control purity by only selecting sources with strong agreement.

The peculiars, on the other hand, vary wildly, from mild disturbances to clear signs of galaxy merging, often with a companion nearby. Additionally, some high-redshift disks and spheroids might end up being classified as peculiars due to more asymmetric/disturbed morphologies than low-redshift counterparts. However, as discussed in Section 3.2, for most peculiars, the quantitative morphology is consistent with disturbed morphologies.

3.2. Quantitative Morphology Evolution

Cross-examining the morphologies defined by eye using quantitative methods is essential for understanding how their appearance changes across cosmic time. We explore the visual morphologies with several quantitative morphology indicators, both nonparametric and parametric (Section 2.4).

Figure 5 shows the concentration (C) and asymmetry (A) plane based on MORFOMETRYKA measurements for four redshift bins. The mean values alongside the distribution's 15% and 85% percentiles for disks, spheroids, and peculiars are plotted as blue squares, red circles, and pink diamonds, respectively. Each class has its distributions positioned within the expected regions for high-redshift galaxies, with peculiars occupying the top of the diagram, the disks the central region, and the spheroids around the lower right, including a large overlap. The positions of each of the three classes remain fairly stable over all redshifts, but the spheroids have higher asymmetries and lower concentrations overall at higher redshift, with larger overlaps with the disks. Also displayed

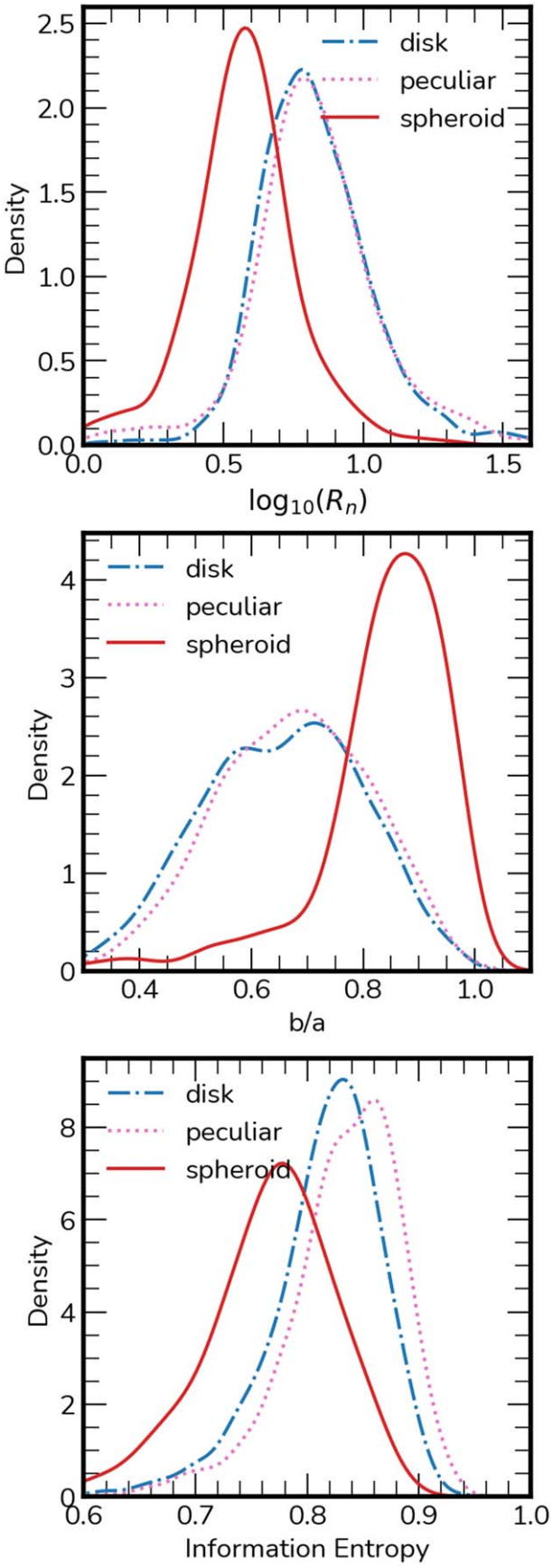

Figure 4. Effective radius (R_n) (top), axis ratio (b/a) (middle), and information entropy (bottom). We show key measurements that provide clues to differences most used by the classifiers for the spheroid and other classes. Spheroids are defined by their lack of structure, low elongation, and small sizes in general.

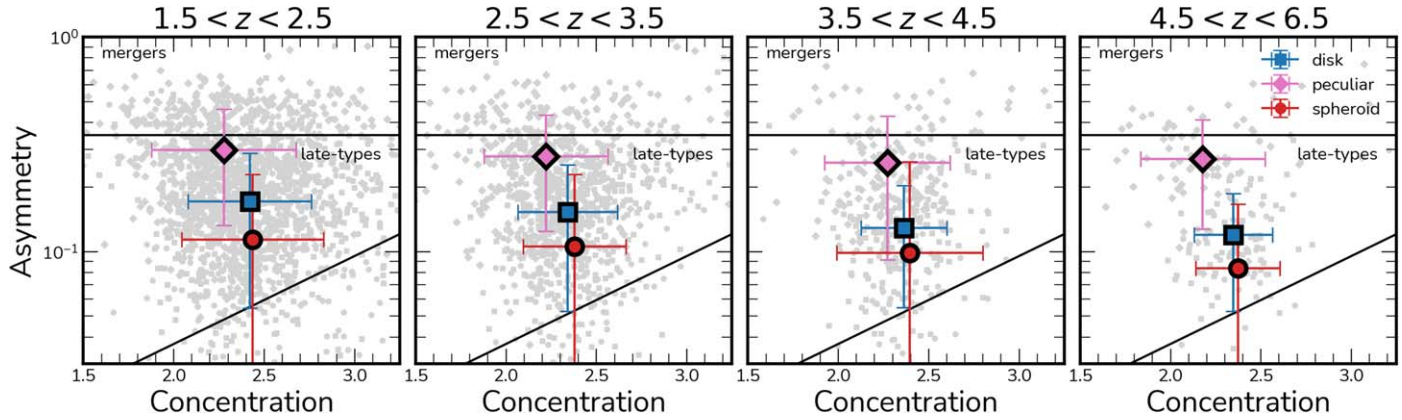

Figure 5. Diagrams showing the evolution of the concentration (C) and asymmetry (A) in four different redshift bins, one for each panel. Peculiars are shown as pink diamonds, disks as blue squares, and spheroids as red circles. The solid lines denote the merger selection threshold on the top, and the late-type/early-type separation on the lower diagonal line. Peculiars display high asymmetries when compared to other types. Disks display late-type-like morphology, while spheroids are regular at lower redshifts, located at the bottom right of the plots, but move toward the center with increasing asymmetry and decreasing concentration with redshift. Galaxies overall get less concentrated and more asymmetric with increasing redshift. However, at high redshifts, sources display higher concentrations and asymmetries when compared to simulations.

with solid lines is the merger criterion based on asymmetry, as

$$A > 0.35, \quad (1)$$

and the diagonal boundary between late types and intermediate types based on Bershady et al. (2000). We also explore the G–M_{20} plane (Lotz et al. 2004, 2008) but we do not find any clear separation between the types, apart from the distinction between sources that have close companions and those that are isolated, similar to what is reported in Rose et al. (2023).

The spirality index (Ferrari et al. 2015)—the standard deviation of the (r, ϕ) gradient map of the galaxy polar image, designed to measure the amount of nonradial structure in the galaxy—has proven to be very effective in discriminating different classes in this sample. If the galaxy is smooth, its polar image will consist of a single horizontal strip, which will imply a low value for σ_ψ. On the other hand, although we cannot resolve spiral arms in most cases, if the galaxy contains peripheral structures or has companions, the polar image will be irregular with a corresponding high σ_ψ. The ellipticity of a source would show as curved deviations on the lower horizontal band, which will increase σ_ψ, but not as strongly as structure on top of the central profile. This can help distinctions between spheroids and disks as the two will have slightly different mean spirality values. We point the reader to the MORFOMETRYKA paper (Ferrari et al. 2015) for a full description of the σ_ψ calculation.

In Figure 6 we show that combining σ_ψ with the asymmetry (A) warrants a reasonable quantitative separation of the overall classes, as the center of each class distribution is well separated in this plane, unlike in the C–A or G–M_{20} diagram. As an example of what is captured by the σ_ψ measurement, we show three galaxies in Figure 7, one for each class, with their respective polar coordinate image and the gradient lines that are used to compute σ_ψ. The distinction between spheroids and disks is subtle, but it is very powerful when large nonradial structure is present in the outskirts of the source.

Finally we explore the evolution of the Sérsic profiles (Sérsic 1963) through the redshift evolution of the Sérsic index for galaxies with $M_* > 10^9\ M_\odot$. In Figure 8 we report mean values together with the 15% and 85% percentile limits as error bars to represent the distributions in each redshift bin for each

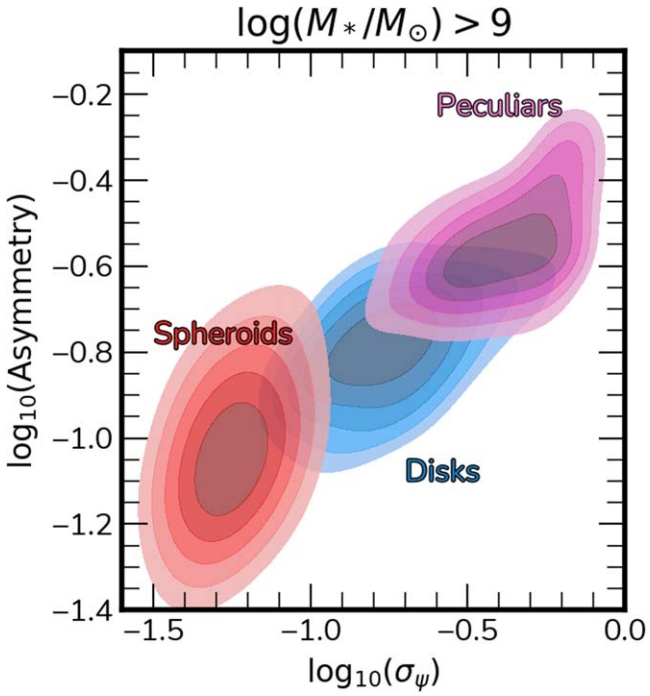

Figure 6. 2D distributions of log asymmetry (A) vs. log spirality (σ_ψ) for each class. Distributions of kernel density estimation are shown for the top 50% of each morphological class in five bins of 10% fractions of the distribution. The asymmetry and σ_ψ correlate strongly, but are independent measurements because each classification distribution has a different slope. Spheroids show high diversity in A and low diversity in σ_ψ, while the contrary is true for disks. These two measurements form a parameter space capable of separating the classes in this sample relatively well compared to C–A.

class. Disks and peculiars exhibit similar Sérsic profiles, with $n \approx 1.0$ at $z \sim 6$ to $n \approx 1.3$ at $z \sim 1.5$. The spheroids, on the other hand, show higher Sérsic indices at all redshifts, with $n \approx 1.8$ at $z \sim 6$ to $n \approx 2.5$ at $z \sim 1.5$. The distinction between the classes is clearer than what was reported in Ferreira et al. (2023) because the CANDELS overlap allows us to quickly select high-mass galaxies only. The slopes for each class are also different, with the spheroids increasing in Sérsic index more rapidly. The similarities among disks and peculiars

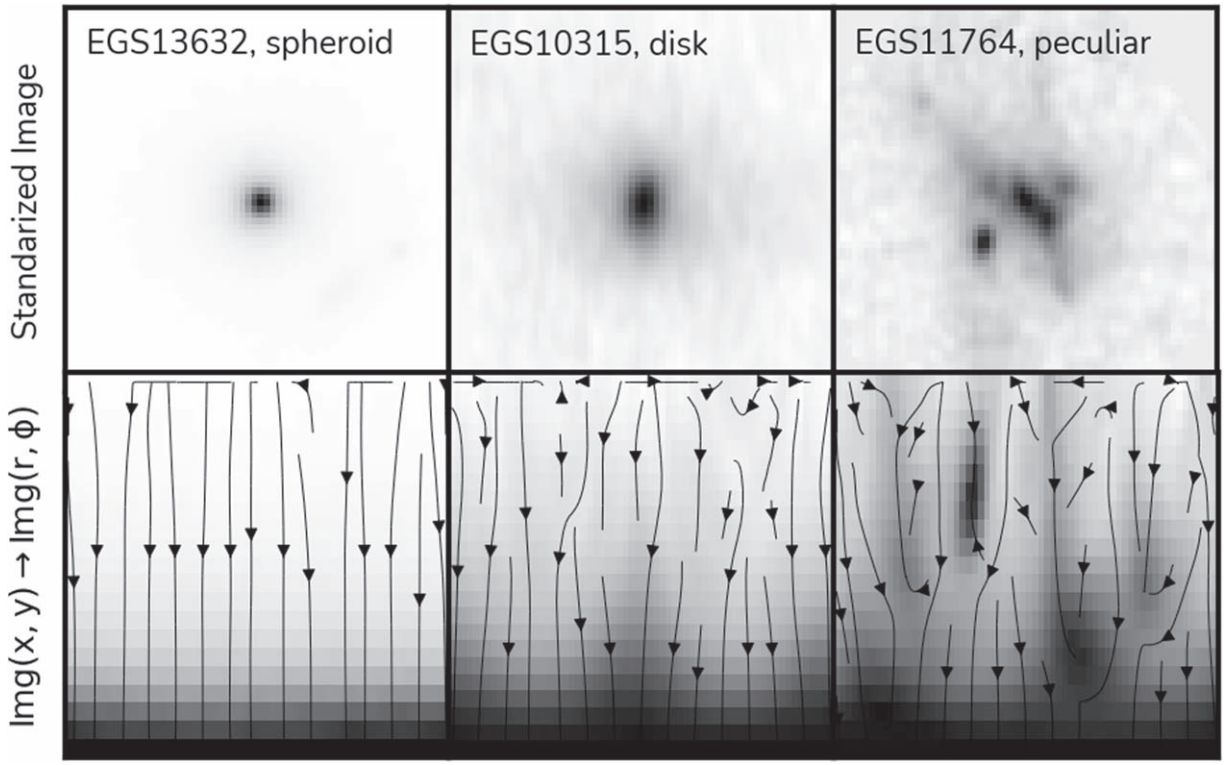

Figure 7. Three examples of the spirality σ_ψ measurement. The top row shows the standardized images of the sources ($q = 1$, position angle $= 0°$) while the bottom row displays the transformation to polar coordinates (r, ϕ) of the above image. Black lines display the gradient field of the image. The σ_ψ measurement is based on the standard deviation of these field lines. A round shape will show as a horizontal band in the bottom of the image; an elongated component will produce small bends in this distribution, both in the edges and at the center; nonradial asymmetric features will produce very irregular (r, ϕ) distributions, resulting in higher σ_ψ.

suggest that the majority of these disturbed and merging systems are still disk-dominated.

3.3. Redshift Effects

With increasing redshift, sources are more difficult to detect and to resolve due to cosmological and observational effects (Ferreira & Ferrari 2018). They suffer the so-called cosmological dimming, which can dim their surface brightness by a $(1+z)^4$ factor, which can be extreme for the high redshifts probed here. Additionally, other effects might be at play, such as distant galaxies being intrinsically smaller and in general more difficult to resolve. In this section we discuss the impact of cosmological effects on our classifications and correct for possible redshift effects in our class fractions.

To investigate redshift effects, we conducted a reclassification experiment. First, we randomly selected 40 galaxies for each main class (disk, peculiar, spheroid) in the lowest redshift bin $1.5 < z < 2.5$, and then simulated the observational effects in four redshift bins, to $z = 3$, 4, 5, 6, following the approach described in Whitney et al. (2021) and Tohill et al. (2021) and preserving the original process as closely as possible by following the rest-frame filters in each redshift bin. Then, L.F. and C.C. reclassified the simulated images using the same decision tree as outlined in Figure 2.

This experiment give us the information necessary to track how the original classifications from $1.5 < z < 2.5$ change with increasing redshift. To find corrections to the observed fractions from our original classifications, we apply the correction procedure described in Mortlock et al. (2013), where we compare the higher redshift classifications with the original classifications, computing corrections for the observed

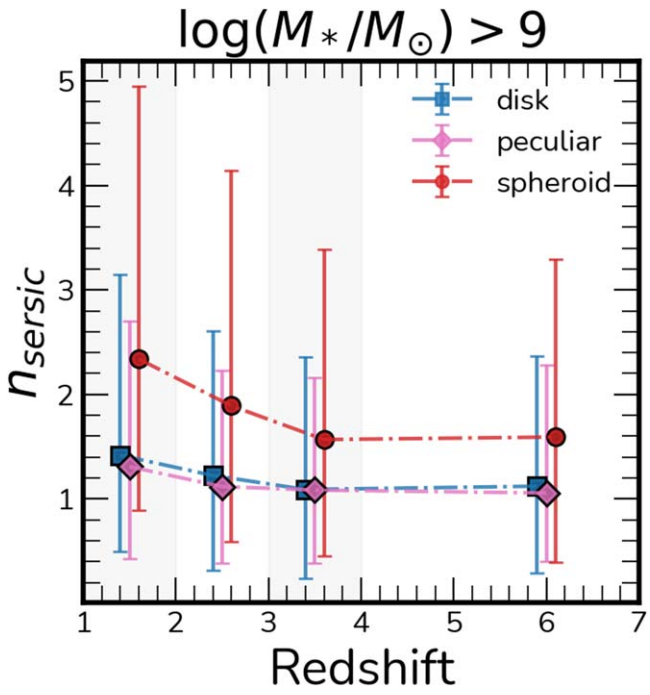

Figure 8. Redshift evolution of Sérsic index for each morphology class. Displayed as blue squares, red circles, and pink diamonds are the means for disks, spheroids, and peculiars, respectively. Error bars define the 15% and 85% percentiles of the distributions.

fractions. Two sets of corrections are devised, one for $\log(M_*/M_\odot) < 9$ and another for $\log(M_*/M_\odot) > 9$.

For low masses with $\log(M_*/M_\odot) < 9$, we find that the fraction of disks is overestimated by up to ∼20% in our

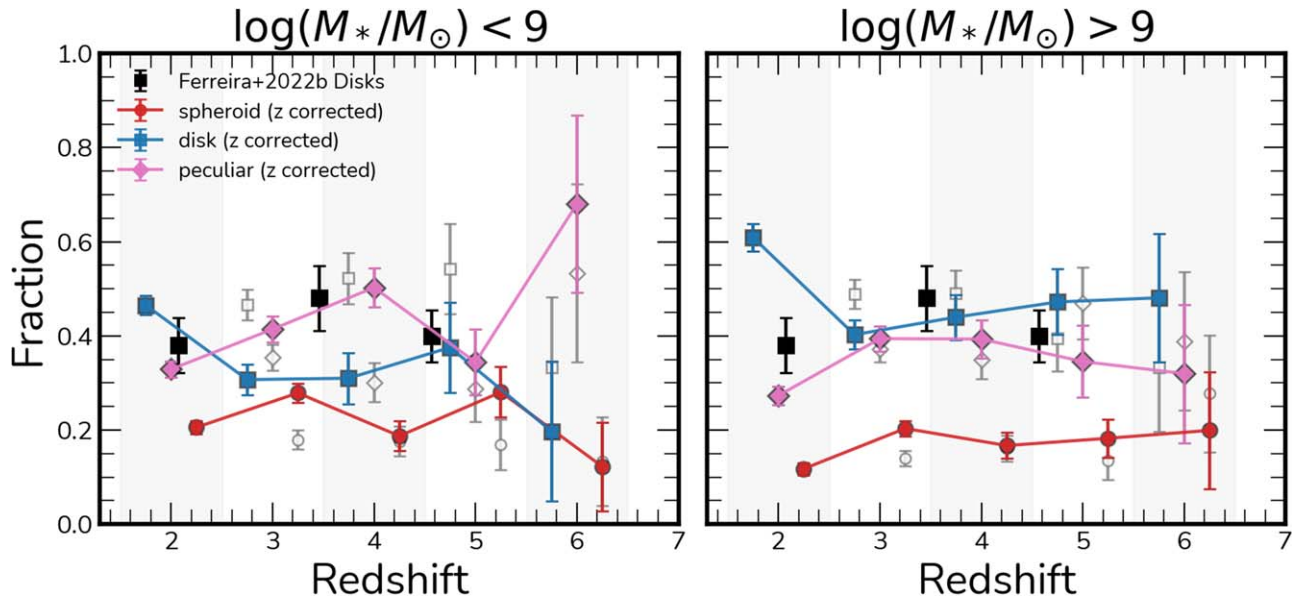

Figure 9. Morphology fraction vs. redshift. The evolution of morphology fraction with redshift for the main morphological classes of our classification framework (Figure 2) in two mass bins, $\log(M_*/M_\odot) < 9$ (left) and $\log(M_*/M_\odot) > 9$ (right). Disks, spheroids, and peculiars are shown as blue squares, red circles, and pink diamonds, respectively. The black squares show the disk fractions reported in Ferreira et al. (2023). The white markers show the class fractions before the redshift corrections.

original classifications, while peculiar morphologies are underestimated by a similar amount, with the spheroid population being fairly robust. This means that low-mass disks are more likely to be peculiar galaxies with unresolved structure. This effect is seen for all redshifts.

At higher masses with $\log(M_*/M_\odot) > 9$, the corrections are modest. For $2.5 < z < 4.5$, there is a ∼5% overestimation of disks that are spread over peculiars and spheroids. For $4.5 < z < 6.5$ the fraction of disks is underestimated by 15%, with most cases being disks classified as peculiars.

In addition to the effects of cross-class misclassifications, the rate of unclassified objects increases dramatically, especially at the highest redshifts and in the low-mass bin. For galaxies with $\log(M_*/M_\odot) < 9$, at $z = 3$%, 11% are unclassifiable or PSF-sized; similarly 23%, 48%, and 69% for $z = 4$, $z = 5$, and $z = 6$, respectively. For $\log(M_*/M_\odot) > 9$ this effect is not as significant, with 1%, 3%, 7%, and 11% for $z = 3$, 4, 5, and 6, respectively. This effect is a combination of sources extending over smaller angular sizes (as well as an increased effect from the PSFs) and sources being dimmer due to cosmological dimming. This demonstrates that the classifications for the low-mass regime can suffer from strong selection effects, especially at high z. On the other hand, the high-mass end of our sample is fairly robust against these effects.

In the following sections we will use these corrections to discuss the evolution of morphological fractions with redshift.

3.4. Evolution of the Hubble Sequence

One principal goal of looking at galaxy morphology and structure is to establish when and how the Hubble Sequence (Hubble 1926) emerges in the context of the hierarchical assembly of the Universe.

Here we report the redshift evolution of morphological classes that encompass the three main categories of the Hubble Sequence from $1.5 < z < 6$, from when the Universe was only ∼1 Gyr old up to ∼4.2 Gyr. In Figure 9 we show this evolution in two mass bins, with the left panel displaying sources with $M_* \leqslant 10^9\ M_\odot$ while the right panel shows this evolution for $M_* > 10^9\ M_\odot$. Spheroids are displayed as red circles, peculiars as pink diamonds, and disks as blue squares. The fractions shown are redshift-corrected fractions based on the analysis presented in Section 3.3, and the original fractions prior to these corrections are also shown as white markers. The fraction of disks from Ferreira et al. (2023) is shown as black squares for comparison.

For low masses ($M_* \leqslant 10^9\ M_\odot$) we find strong evolution with redshift, with the peculiars increasing dramatically from ∼30% at $z = 2$ to ∼70% at $z = 6$ while the fraction of disks decreases for all redshifts, and spheroids decrease from ∼20% to ∼10% at $z = 6$. The disk fractions in the range $2 < z < 5$ and the results previously reported in Ferreira et al. (2023) only agree on the uncorrected fractions. This suggests that these previous results are likely to be overestimated due to redshift effects in the lower mass range. However, it still shows a significant increase over disk fractions for the same redshift range from Mortlock et al. (2013).

In the high-mass case ($M_* > 10^9\ M_\odot$) we observe almost no evolution in the redshift range probed, with a slight offset from $z = 2$ to $z = 3$, but the fractions remain fairly constant up to $z = 6$. The disks stay at around 40%–60%, while peculiars correspond to 20%–40% of the overall fraction with spheroids steadily increasing from 10% to ∼20%.

We note, however, that within the error bars the fraction of high-mass disks in this work is consistent with our previous results in Ferreira et al. (2023).

3.4.1. Spheroid Evolution

Another remarkable aspect of the evolution of galaxy type fraction is that the spheroid fraction is roughly constant, even up to the highest redshifts, but lower than the level seen for local galaxies. We generally do not see an increase in the number of elliptical galaxies at lower redshifts, at least down to $z \sim 1.5$. We might have expected to see more spheroids at lower z if mergers are progressively transforming peculiars into

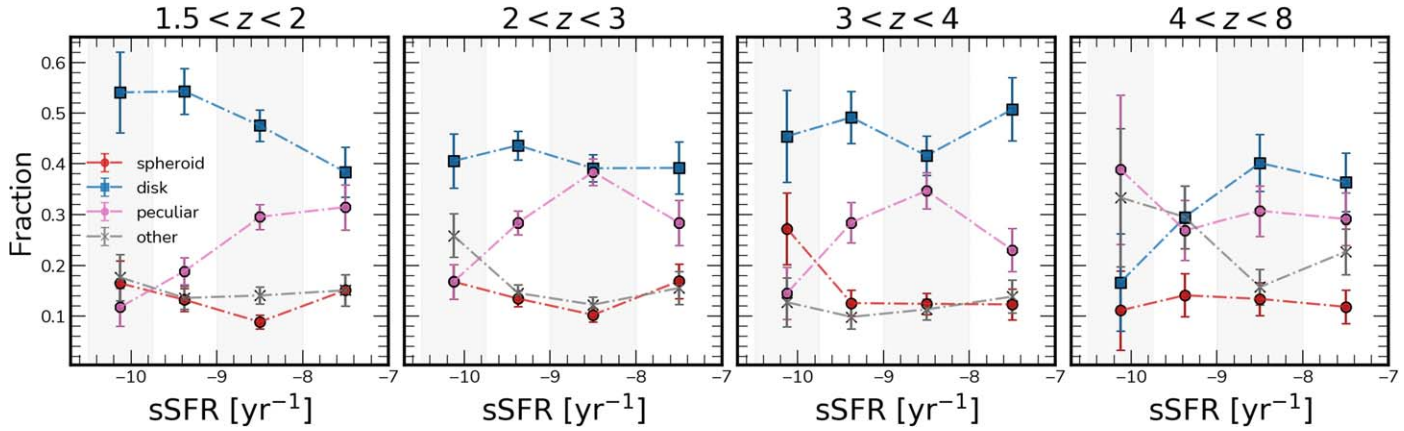

Figure 10. Morphology fractions vs. average specific star formation rate. Disks, spheroids, peculiars, and the other class are plotted as blue squares, red circles, pink diamonds, and gray crosses, respectively. Four redshift bins are shown. For $1.5 < z < 4.5$ the disk galaxies dominate the contribution, while for $z > 4.5$ peculiars contribute similarly or more than disks in each sSFR bin.

ellipticals. This indicates that elliptical galaxies in the mass ranges we probe must form morphologically relatively late in the history of the Universe. Despite this, there are clear indications that some ellipticals were morphologically present very early in the Universe's history. This may be a sign of different formation mechanisms at play for these spheroids at different epochs, with some forming through dissipative collapse, some from major mergers, and others through minor mergers.

These trends also suggest that the classic picture of morphology and structural evolution driven by merging might only be important for lower-mass galaxies, where the high-mass Universe can be described broadly by a consistent Hubble Sequence in the range $1.5 < z < 6$. We discuss this in more detail in the section below.

3.4.2. The Role of Mergers in Early Galaxy Assembly

Prior to JWST, the scientific consensus obtained through both observations and simulations was that in the early Universe, galaxies grow and evolve through hierarchical mergers. In simulations, this is seen in galaxy merger trees (e.g., Mo et al. 2010), as well as in the fact that the merger rate increases with redshift (e.g., Conselice et al. 2003; Rodriguez-Gomez et al. 2015; Duncan et al. 2019; Ferreira et al. 2020). Although observational studies of extremely high-redshift galaxies were previously limited, Mortlock et al. (2013) found an increase in the fraction of peculiar galaxies with redshift, supporting the idea that mergers were frequent.

However, the first JWST observations challenged this picture. Ferreira et al. (2023), Jacobs et al. (2023), Nelson et al. (2023), and Robertson et al. (2023) all found a higher fraction of high-redshift disks and a lower fraction of peculiar galaxies than expected. The initial results, although based on small data sets, suggested either that mergers are less frequent than we thought or that high-redshift disks tend to survive mergers and retain their disk morphology.

In this paper, for the first time, we have looked at a statistically large sample of high-redshift galaxies observed with JWST, split into two different mass bins, to trace the evolution of high- and low-mass galaxies.

As stated in Section 3.4, low-mass galaxies show a strong gradient in morphological type fractions. This suggests again that for low-mass galaxies, mergers do drive structural evolution. This result supports the merger paradigm in principle and is similar to the pattern seen with HST (e.g., Conselice et al. 2008). Our results are in fact consistent with a significant fraction of low-mass galaxies undergoing mergers in the early Universe, going through a peculiar phase, and then forming stable disks. Kinematic observations of these galaxies would go some way toward understanding whether this picture is correct.

In the high-mass bin, we see roughly constant class fractions for all redshifts. However, the fraction of high-mass disks is still significantly higher in our data than in HST-based analyses (e.g., Mortlock et al. 2013). At $z = 2$, ~60% of galaxies in our JWST observations have a disk morphology compared to ~10% in Mortlock et al. (2013). Therefore, our results still point at a tension between HST- and JWST-based morphology studies. Although our results are consistent with some of the evolution between $z = 1.5$ and $z = 6$ being merger-driven, the fraction of $z > 2$ disks is still too high. We will investigate in future papers the detailed merger histories of these galaxies based on the JWST data. Ultimately, we want a self-consistent observational picture for how galaxy formation is occurring.

Overall, we see that peculiar structure is more common among low-mass than high-mass galaxies, pointing toward mass assembly via mergers; however, the fraction of disk galaxies in all bins is still higher than what is observed by HST. Some implications of this result could be that either (1) galaxies grow by mergers as well as by another process, e.g., gas accretion (L'Huillier et al. 2012), or (2) mergers between high-mass galaxies do not destroy disks efficiently, allowing some galaxies to retain their disk morphology in the long term (Hopkins et al. 2009; Sparre & Springel 2017).

3.5. Star Formation and Evolution of Stellar Mass

One important goal of tracking morphologies across cosmic time is determining whether different morphologies contribute to the star formation and the stellar mass budget of the Universe differently. In the previous section we explored the fractional evolution with redshifts, concluding that disk galaxies dominate the overall fraction of morphologies at $1.5 < z < 6$ for high masses while peculiar galaxies dominate at the lower-mass end. In Figure 10 we show the class fractions in bins of specific star formation rate (sSFR) divided into four redshift bins. The overall fraction of galaxy types for each redshift

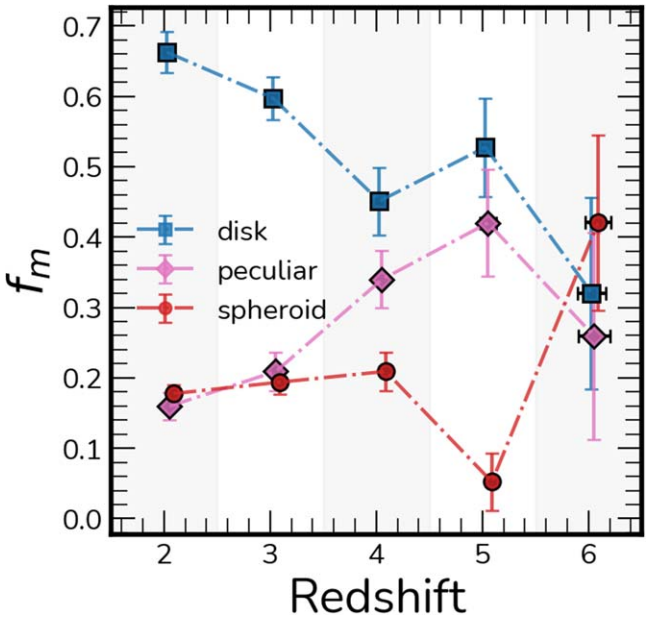

Figure 11. Fraction of the total stellar mass (f_m) in each morphology subsample vs. redshift. This represents the relative amount each class contributes to the total stellar mass at that redshift. Disks, spheroids, and peculiars are plotted as blue squares, red circles, and pink diamonds. The contribution from disks shows a trade-off with respect to peculiars and spheroids at higher redshifts. No stellar mass cut is applied: the whole sample is included.

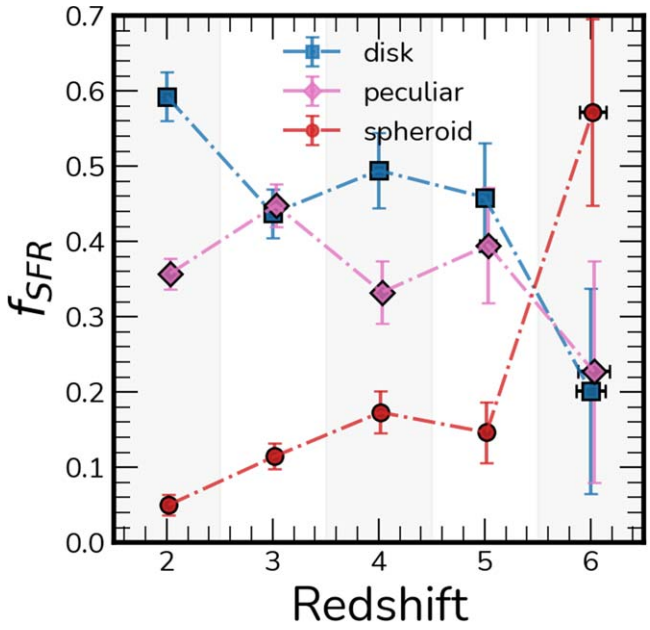

Figure 12. Fraction of SFR (f_{SFR}) in each morphology subsample vs. redshift. We show the contribution to total SFR of each redshift bin from each morphological class down to our mass limit. This represents how much each galaxy class contributes to the total SFR at that redshift. Disks, spheroids, and peculiars types are plotted in blue squares, red circles, and pink diamonds, respectively. No stellar mass cut is applied: the whole sample is included.

panel can be seen in Figure 9, while each bin in Figure 10 shows these fractions for a given sSFR bin. The disks dominate the overall contribution in sSFR, with spheroids showing a similar fraction to the overall fraction of spheroids in the sample. However, we see that in high-sSFR bins, the contribution of peculiar galaxies increases. The same trend is shown for all redshift bins, as the fraction of peculiars increases with increasing sSFR. For the highest redshift bin ($4.5 < z < 6.5$), peculiars display more or roughly the same contribution as disks. Spheroid fractions are slightly higher at lower sSFR. This suggests that peculiar galaxies are important sites of star formation at all times in the Universe, and especially at higher redshifts.

In Figure 11 we show the contribution of each morphological class to the total stellar mass in each redshift bin. This shows that most of the mass in the very early Universe was located in spheroid and peculiar galaxies, while a clear trend with redshift is evident for disk galaxies, such that for $z < 3$ most of the mass of the sample is distributed among disk galaxies. More massive overall individually, the spheroid galaxies hold just a small fraction of the total stellar mass in this sample ($f_m \sim 15\%$) as it is greatly outweighed by the amount of mass in galaxies with disk and peculiar morphologies, with the exception of the highest redshift bin. This is in contrast to what is found in the Hubble Deep Field (Conselice et al. 2005) due to the very different morphology fractions reported. It thus appears that for the bulk of the history of the Universe, most stellar mass in the Universe has been located in galaxies with disk-like morphologies.

We plot in Figure 12 the fraction of star formation that exists in different galaxy types. As can be seen, the disk galaxies dominate the star formation rate in the Universe up to at least $z \sim 5$, with similar contributions from peculiars close behind. What this implies is that, for our sample, the most likely galaxy type in which stars form is disk galaxies. However, this is still less than 50% of the star formation until $z < 2.5$ when the contribution of disks reaches as high as ~60%. As can be seen, a significant amount of star formation also occurs in peculiar galaxies, which are probably mergers, but these are not the dominant locations in which stars are formed.

3.6. Predictions from Simulations

We now look at numerical simulation results for morphological evolution over a similar redshift range. Simulations that resolve galaxies self-consistently typically model mass elements either on a grid or as particles. Particle-based decomposition methods (e.g., Abadi et al. 2003; Crain et al. 2010; Thob et al. 2019; Irodotou & Thomas 2021; Zana et al. 2022) have been used extensively in order to split galaxies into different morphological classes and facilitate a comparison between observed and simulated galactic properties (Tissera et al. 2012; Pillepich et al. 2015; Irodotou et al. 2019; Monachesi et al. 2019; Trayford et al. 2019; Rodriguez-Gomez et al. 2022). However, the true morphology of a system may not always be accurately captured, as particle-based methods can be sensitive to small perturbations in the distribution of particles, which become progressively more significant at lower stellar masses as these galaxies are resolved with fewer particles.

In this work, to ensure that both galaxies and their components are sufficiently resolved, and thus a particle-based decomposition is applicable, we use central and satellite galaxies from the EAGLE (Crain et al. 2015; Schaye et al. 2015, for $1.5 < z < 4$) and FLARES (Lovell et al. 2021, for $5 < z < 8$) simulations with stellar masses $\log(M_*/M_\odot) > 10$ (i.e., even for a galaxy with a bulge-to-total mass ratio $B/T \sim 0.2$, the bulge is resolved with more than a thousand

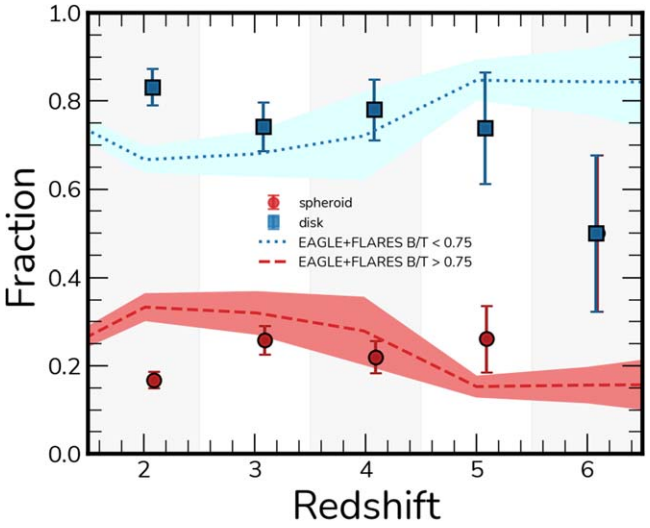

Figure 13. Morphology fractions compared to B/T morphological type selection in EAGLE and FLARES for massive galaxies ($M_* \geqslant 10^{10}\ M_\odot$). The relative fractions of disks and spheroids are shown as blue squares and red circles, respectively. The blue dotted line shows mean values for galaxies with $B/T < 0.75$ in EAGLE and FLARES. The dashed line shows galaxies with $B/T > 0.75$. Shaded regions represents $\pm 2\sigma$ for each distribution. The visual classification fractions only account for disks and spheroids to allow a more direct comparison with the two thresholds in B/T in the simulations.

particles with a mass of a few $\times 10^6\ M_\odot$ each). We use the method developed in Irodotou & Thomas (2021) to decompose galaxies by first creating a Mollweide projection of the angular momentum map of each galaxy's stellar particles. Then, stellar particles are assigned to a disk or spheroid component based on their angular separation from the densest grid cell. This allows us to calculate B/T ratios and use these to split galaxies into two morphological classes: (i) spheroids with $B/T > 0.75$, (ii) spirals with $B/T < 0.75$. The aforementioned B/T limits were calibrated at $z \sim 0$ in order for the EAGLE galaxies to match the morphological classes in the Conselice (2006) sample. In Figure 13 we show the comparison of these fractions with the relative fraction of disks and spheroids from the visual classifications for high-mass galaxies with $M_* \geqslant 10^{10}\ M_\odot$, ignoring the peculiars, as we do not have a direct way to classify peculiars in the simulation data set. The trends between simulations and visual classifications agree for $z > 3$, with the exception of a single anomalous visual classification redshift bin showing similar fractions of disks and spheroids for $5 < z < 6$. Fractions for $z < 3$ overall disagree, with an excess of 10% of disks in the visual classifications. It is worth noting, however, that in this redshift range (i.e., $1.5 < z < 3$), Lagos et al. (2018) showed that the fraction of dry major mergers in the EAGLE volume increases. Since this type of merger can efficiently reduce the angular momentum of the remnant, this will translate to a negative correlation between our B/T values and redshift, as also seen in, e.g., Figure 4 of Clauwens et al. (2018) for $M_* \geqslant 10^{10.5}\ M_\odot$.

Unfortunately, there is currently no way of separating potential peculiars from the whole population of low-B/T and high-B/T galaxies from the simulation without a different selection approach. A fraction of these two types would indeed show peculiar morphologies, lowering the overall fractions of both disks and spheroids. However, here we also do not include the peculiars from our sample and instead use only the disks and spheroids. We refer the reader to the findings of Park et al. (2022), which reports similar fractions for the Horizon simulations, with the $z \geqslant 5$ fraction of disks dominating at around 75% by using quantitative morphology indicators as a proxy for morphology. In their case, peculiars are included and consistent with this overall picture.

3.7. HST versus JWST

As discussed in Section 3.4, there is a stark difference between morphological classifications derived with HST and JWST observations. The new JWST data are challenging our understanding of galaxy evolution and structure formation in the early Universe by revealing the resolved optical morphologies of high-redshift galaxies for the first time. Here, at the end of this paper, we discuss some differences between HST and JWST observations and provide an analysis of what contributes to this discrepancy. This is meant to facilitate discussion about why our results are so different from previous work and to point the way toward understanding how to carry out future JWST work on galaxy structure.

A comparison for a select number of galaxies in our sample between the NIRCam stamps and the HST ACS and WFC3 stamps is shown in Figure 14. The classification label shown is derived from our JWST classifications. Many galaxies show very clear structures in NIRCam but ambiguous morphologies in HST. In some cases, such as EGS 23205, only the central component is clearly seen in HST, while a disk, spiral arms, and a bar pop up in the longer-wavelength bands. In a few cases, such as EGS 22543, the source is barely detected in the WFC3 and SW NIRCam images, while a clear disk is visible for the LW NIRCam stamps.

A detailed comparison with the CANDELS classifications (Kartaltepe et al. 2015) is beyond the scope of this paper because our classifications do not align perfectly with the scheme defined in Kartaltepe et al. (2015). However, we briefly discuss the modes for which the discrepancies between classifications based on HST imaging and NIRCam can be explained. First, many of the fine structures, such as bars and spirals, are hard to resolve at high redshifts due to the WFC3 pixel scale, and can be mistaken for merging signatures or disturbances, such as in the case of EGS 14565. This shows clear spiral structure in JWST but not as clear in HST, and could be mistaken for a merger in HST. Second, we find the wavelength coverage to be critical, as many of the galaxies at high redshift in HST are probed in the blue side of the optical, and are prone to absorption from dust, giving rise to asymmetric-looking structures. Moreover, the bluer bands probe the youngest stars, which have more irregular spatial distributions and trace sites of ongoing star formation but not the underlying mass distribution. Galaxies such as EGS 522543 and EGS 16559 are good examples of this. This is also expected to be an issue for galaxies at $z > 7$ in NIRCam images, as we start having the same issue as WFC3 for $2 < z < 3$.

Future morphology classification studies, with more detailed descriptions and covering larger data sets, such as the complete cycle of CEERS, PRIMER, COSMOS-WEB, JADES, and NGDEEP, together with the scope of citizen science projects such as Galaxy Zoo (Lintott et al. 2011), will enable a detailed discussion on the main differences between HST and JWST morphology in the overlap region $2 < z < 3$.

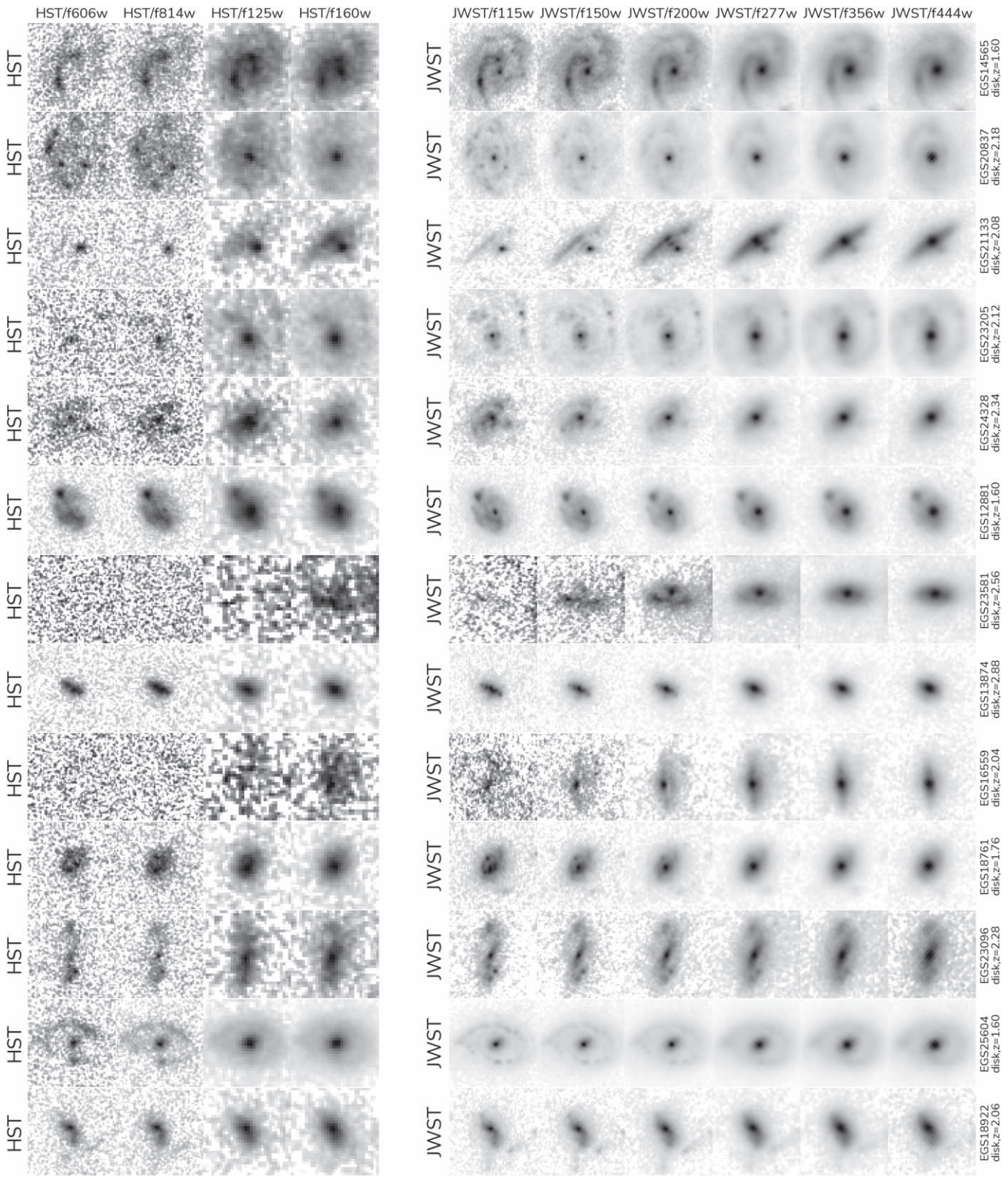

Figure 14. A comparison of HST vs. JWST. We show 13 galaxies in our sample that have observations in the four main CANDELS filters (left panel) and SW and LW filters in JWST (right panel). Faint features in CANDELS are generally very clear in JWST. In some cases only the central core of the galaxy is visible with the HST imaging. The classification label shown is derived from our JWST classifications.

4. Summary and Conclusions

We present results on the rest-frame optical morphologies and the structural evolution of JWST-observed galaxies at $z = 1.5$–6 in a statistically significant sample of 3956 galaxies for the first time, using both visual classifications and quantitative morphology. We focus on galaxies observed by the CEERS program that overlap with the CANDELS fields,

enabling us to use robust measurements of redshifts, stellar masses, and star formation rates available in CANDELS.

Our major findings are:

1. After accounting for redshift effects, distant galaxies at $z > 1.5$ display surprisingly regular disk morphologies at early times when compared to HST. However, for galaxies with $M_* < 10^9\ M_\odot$, tremendous evolution is observed in the fractions of disks and peculiars. The fraction of disks decreases from ∼40% at $z \sim 2$ to ∼20% at $z \sim 6$, while peculiars increase in a similar proportion. Spheroids remain roughly constant over time. This suggests that the role of mergers might be more important for less massive galaxies.
2. For galaxies with higher masses $M_* > 10^9\ M_\odot$, the fractions of disks/spheroids/peculiars appear to be roughly constant at $1.5 < z < 6.5$, showing that the Hubble Sequence was already in place as early as one billion years after the Big Bang.
3. Nonparametric morphology measurements agree well with visual classifications. However, a large overlap exists between classes in the usual CAS, etc. planes. We find that the spirality index (σ_ψ) when combined with the asymmetry (A) makes a powerful diagnostic to separate disks/spheroids/peculiars.
4. Comparisons with B/T studies from EAGLE and FLARES show that quantitative structures at high redshifts agree well with simulations and are not unexpected from a theoretical standpoint, even if discrepant with previous morphological studies with HST.
5. Galaxies with disk morphologies dominate both the low-sSFR and high-sSFR populations, fairly outnumbering spheroids. However, the peculiar contribution to the sSFR budget increases with increasing sSFR and redshift, such that at the highest redshifts, the majority of the highly star-forming galaxy population has disturbed/peculiar morphologies.
6. The contribution to the total stellar mass of galaxies at high redshift is dominated by disk galaxies at $z < 4$, with similar contributions from peculiars at $z > 4$. However, still most of the stellar mass in the Universe at $z < 3$ is located in disk galaxies. We also find that disks and peculiar galaxies contribute similarly to the star formation rate at $z < 5$, suggesting that most stars in the Universe were formed in a galaxy with a disk or peculiar morphology.
7. We report clear examples of galaxies whose morphologies are hidden in HST imaging, but become clear and unambiguous in the NIRCam observations. Spirals and bars are better resolved and clear in the LW NIRCam filters.

In addition to the morphology study presented in this paper, we release the first version of our catalog of aggregated classifications to the community. Our goal is that this large sample of visually classified galaxies will serve as a base for early studies on morphology and structure, and will help the community develop methods and tools to tackle scheduled observations of larger areas such as COSMOS-WEB, while data releases from large citizen science classification projects, such as Galaxy Zoo, are not available. These classifications can be used, for example, as an early training data set for deep learning methods or as a transfer learning sample for already established models.

Acknowledgments

We thank the anonymous referee for invaluable feedback that improved the paper greatly. We thank Anthony Holloway, Sotirios Sanidas, and Phil Perry for critical and timely help with computer infrastructure that made this work possible. We acknowledge support from the ERC Advanced Investigator Grant EPOCHS (788113), as well as a studentship from STFC. L.F. acknowledges financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES) in the form of a PhD studentship. D.I. acknowledges support by the European Research Council via ERC Consolidator Grant KETJU (Co. 818930). C.C.L. acknowledges support from the Royal Society under grant RGF/EA/181016. C.T. acknowledges funding from the Science and Technology Facilities Council (STFC). This work is based on observations made with the NASA/ESA Hubble Space Telescope (HST) and NASA/ESA/CSA James Webb Space Telescope (JWST) obtained from the `Mikulski Archive for Space Telescopes` (`MAST`) at the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST, and NAS 5-26555 for HST. The specific observations analyzed can be accessed via doi:10.17909/xm8m-tt59.

This research made use of the following Python libraries: ASTROPY (Astropy Collaboration et al. 2022); MORFOMETRYKA (Ferrari et al. 2015); PANDAS (Reback 2022); MATPLOTLIB (Hunter 2007); PHOTUTILS (Bradley et al. 2020).

ORCID iDs

Leonardo Ferreira https://orcid.org/0000-0002-8919-079X
Christopher J. Conselice https://orcid.org/0000-0003-1949-7638
Elizaveta Sazonova https://orcid.org/0000-0001-6245-5121
Fabricio Ferrari https://orcid.org/0000-0002-0056-1970
Joseph Caruana https://orcid.org/0000-0002-6089-0768
Clár-Bríd Tohill https://orcid.org/0000-0003-2527-0819
Geferson Lucatelli https://orcid.org/0000-0002-2410-1776
Nathan Adams https://orcid.org/0000-0003-4875-6272
Dimitrios Irodotou https://orcid.org/0000-0003-2946-8080
Madeline A. Marshall https://orcid.org/0000-0001-6434-7845
Will J. Roper https://orcid.org/0000-0002-3257-8806
Christopher C. Lovell https://orcid.org/0000-0001-7964-5933
Aprajita Verma https://orcid.org/0000-0002-0730-0781
Duncan Austin https://orcid.org/0000-0003-0519-9445
James Trussler https://orcid.org/0000-0002-9081-2111
Stephen M. Wilkins https://orcid.org/0000-0003-3903-6935

References

Abadi, M. G., Navarro, J. F., Steinmetz, M., & Eke, V. R. 2003, ApJ, 591, 499
Abraham, R. G., Valdes, F., Yee, H. K. C., & van den Bergh, S. 1994, ApJ, 432, 75
Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, ApJS, 218, 33
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645

Bradley, L., Sipőcz, B., Robitaille, T., et al. 2020, astropy/photutils: v1.0.0, Zenodo, doi:10.5281/zenodo.4044744
Buitrago, F., Trujillo, I., Conselice, C. J., et al. 2008, ApJL, 687, L61
Buitrago, F., Trujillo, I., Conselice, C. J., & Häußler, B. 2013, MNRAS, 428, 1460
Clauwens, B., Schaye, J., Franx, M., & Bower, R. G. 2018, MNRAS, 478, 3994
Coe, D., Bradley, L., & Zitrin, A. 2015, ApJ, 800, 84
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J. 2006, ApJ, 638, 686
Conselice, C. J. 2014, ARA&A, 52, 291
Conselice, C. J., & Arnold, J. 2009, MNRAS, 397, 208
Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
Conselice, C. J., Blackburne, J. A., & Papovich, C. 2005, ApJ, 620, 564
Conselice, C. J., Rajgor, S., & Myers, R. 2008, MNRAS, 386, 909
Crain, R. A., McCarthy, I. G., Frenk, C. S., Theuns, T., & Schaye, J. 2010, MNRAS, 407, 1403
Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937
Dekel, A., Ginzburg, O., Jiang, F., et al. 2020, MNRAS, 493, 4126
Delgado-Serrano, R., Hammer, F., Yang, Y. B., et al. 2010, A&A, 509, A78
Driver, S. P., Windhorst, R. A., & Griffiths, R. E. 1995, ApJ, 453, 48
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, ApJ, 876, 110
Ferrari, F., de Carvalho, R. R., & Trevisan, M. 2015, ApJ, 814, 55
Ferreira, L., Adams, N., Conselice, C. J., et al. 2023, ApJL, 938, L2
Ferreira, L., Conselice, C. J., Duncan, K., et al. 2020, ApJ, 895, 115
Ferreira, L., & Ferrari, F. 2018, MNRAS, 473, 2701
Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJL, 946, L13
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Hammer, F., Flores, H., Elbaz, D., et al. 2005, A&A, 430, 115
Hammer, F., Flores, H., Puech, M., et al. 2009, A&A, 507, 1313
Hopkins, P. F., Bundy, K., Croton, D., et al. 2010, ApJ, 715, 202
Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009, ApJ, 691, 1168
Hubble, E. P. 1926, ApJ, 64, 321
Hunter, J. D. 2007, CSE, 9, 90
Irodotou, D., & Thomas, P. A. 2021, MNRAS, 501, 2182
Irodotou, D., Thomas, P. A., Henriques, B. M., Sargent, M. T., & Hislop, J. M. 2019, MNRAS, 489, 3609
Jacobs, C., Glazebrook, K., Calabrò, A., et al. 2023, ApJL, 948, L13
Kartaltepe, J. S., Mozena, M., Kocevski, D., McIntosh, D. H., & Lotz, J. 2015, ApJS, 221, 11
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Lagos, C. d. P., Stevens, A. R. H., Bower, R. G., et al. 2018, MNRAS, 473, 4956
L'Huillier, B., Combes, F., & Semelin, B. 2012, A&A, 544, A68
Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177
Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
Lovell, C. C., Vijayan, A. P., Thomas, P. A., et al. 2021, MNRAS, 500, 2127
Marshall, M. A., Watts, K., Wilkins, S., et al. 2022, MNRAS, 516, 1047
Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution (Cambridge: Cambridge Univ. Press)
Monachesi, A., Gómez, F. A., Grand, R. J. J., et al. 2019, MNRAS, 485, 2589
Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185
Nelson, D., Springel, V., Pillepich, A., et al. 2019, ComAC, 6, 2
Nelson, E. J., Suess, K. A., Bezanson, R., et al. 2023, ApJL, 948, L18
Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2010, ApJL, 709, L21
Park, C., Lee, J., Kim, J., et al. 2022, ApJ, 937, 15
Peschken, N., Łokas, E. L., & Athanassoula, E. 2020, MNRAS, 493, 1375
Petrosian, V. 1976, ApJL, 210, L53
Pillepich, A., Madau, P., & Mayer, L. 2015, ApJ, 799, 184
Puech, M., Hammer, F., Hopkins, P. F., et al. 2012, ApJ, 753, 128
Reback, J. 2022, pandas-dev/pandas: Pandas, v1.4.2, Zenodo, doi:10.5281/zenodo.3509134
Reynolds, J. H. 1920, MNRAS, 81, 129
Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, ApJL, 942, L42
Rodriguez-Gomez, V., Genel, S., Fall, S. M., et al. 2022, MNRAS, 512, 5978
Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, MNRAS, 449, 49
Roper, W. J., Lovell, C. C., Vijayan, A. P., et al. 2022, MNRAS, 514, 1921
Rose, C., Kartaltepe, J. S., Snyder, G. F., et al. 2023, ApJ, 942, 54
Rosse, T. E. O. 1850, RSPT, 140, 499
Sandage, A. 2005, ARA&A, 43, 581
Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Sérsic, J. L. 1963, BAAA, 6, 41
Sparre, M., & Springel, V. 2017, MNRAS, 470, 3946
Stefanon, M., Yan, H., Mobasher, B., et al. 2017, ApJS, 229, 32
Stewart, K. R., Bullock, J. S., Wechsler, R. H., & Maller, A. H. 2009, ApJ, 702, 307
Thob, A. C. R., Crain, R. A., McCarthy, I. G., et al. 2019, MNRAS, 485, 972
Tissera, P. B., White, S. D. M., & Scannapieco, C. 2012, MNRAS, 420, 255
Tohill, C., Ferreira, L., Conselice, C. J., Bamford, S. P., & Ferrari, F. 2021, ApJ, 916, 4
Trayford, J. W., Frenk, C. S., Theuns, T., Schaye, J., & Correa, C. 2019, MNRAS, 483, 744
Whitney, A., Ferreira, L., Conselice, C. J., & Duncan, K. 2021, ApJ, 919, 139
Zana, T., Lupi, A., Bonetti, M., et al. 2022, MNRAS, 515, 1524