



# Groups and Protocluster Candidates in the CLAUDS and HSC-SSP Joint Deep Surveys

Qingyang Li<sup>1</sup>, Xiaohu Yang<sup>1,2</sup>, Chengze Liu<sup>1</sup>, Yipeng Jing<sup>1,2</sup>, Min He<sup>1</sup>, Jia-Sheng Huang<sup>3</sup>, Y. Sophia Dai<sup>3</sup>, Marcin Sawicki<sup>4</sup>, Stephane Arnouts<sup>5</sup>, Stephen Gwyn<sup>6</sup>, Thibaud Moutard<sup>5</sup>, H. J. Mo<sup>7</sup>, Kai Wang<sup>7,8</sup>, Antonios Katsianis<sup>1</sup>, Weiguang Cui<sup>9,10</sup>, Jiaxin Han<sup>1</sup>, I-Non Chiu<sup>2</sup>, Yizhou Gu<sup>1</sup>, and Haojie Xu<sup>1</sup>

<sup>1</sup> Department of Astronomy, School of Physics and Astronomy, and Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China; [qingyli@sjtu.edu.cn](mailto:qingyli@sjtu.edu.cn), [xyang@sjtu.edu.cn](mailto:xyang@sjtu.edu.cn)

<sup>2</sup> Tsung-Dao Lee Institute and Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

<sup>3</sup> Chinese Academy of Sciences South America Center for Astronomy (CASSACA), National Astronomical Observatories of China, Chinese Academy of Sciences, 20A Datun Road, Beijing 100012, People's Republic of China

<sup>4</sup> Department of Astronomy and Physics and Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, NS, B3H 3C3, Canada

<sup>5</sup> CNRS, CNES, LAM, Aix Marseille Université, 38 rue F. Joliot-Curie, F-13388, Marseille, France

<sup>6</sup> NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada

<sup>7</sup> Department of Astronomy, University of Massachusetts Amherst, MA 01003, USA

<sup>8</sup> Department of Astronomy, Tsinghua University, Beijing 100084, People's Republic of China

<sup>9</sup> Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

<sup>10</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK

Received 2022 April 3; revised 2022 May 4; accepted 2022 May 9; published 2022 June 28

## Abstract

Using the extended halo-based group finder developed by Yang et al., which is able to deal with galaxies via spectroscopic and photometric redshifts simultaneously, we construct galaxy group and candidate protocluster catalogs in a wide redshift range ( $0 < z < 6$ ) from the joint CFHT Large Area *U*-band Deep Survey and Hyper Suprime-Cam Subaru Strategic Program deep data set. Based on a selection of 5,607,052 galaxies with *i*-band magnitude  $m_i < 26$  and a sky coverage of  $34.41 \text{ deg}^2$ , we identify a total of 2,232,134 groups, of which 402,947 groups have at least three member galaxies. We have visually checked and discussed the general properties of these richest groups at redshift  $z > 2.0$ . By checking the galaxy number distributions within a  $5\text{--}7 h^{-1} \text{ Mpc}$  projected separation and a redshift difference  $\Delta z \leq 0.1$  around those richest groups at redshift  $z > 2$ , we identify lists of 761, 343, and 43 protocluster candidates in the redshift bins  $2 \leq z < 3$ ,  $3 \leq z < 4$ , and  $z \geq 4$ , respectively. In general, these catalogs of galaxy groups and protocluster candidates will provide useful environmental information in probing galaxy evolution along cosmic time.

*Unified Astronomy Thesaurus concepts:* Dark matter (353); Dark matter distribution (356); Large-scale structure of the universe (902); Galaxies (573); Galaxy groups (597); Galaxy clusters (584); Galaxy dark matter halos (1880); Protoclusters (1297); Galaxy evolution (594)

## 1. Introduction

Over the past two decades, great achievements have been made in building galaxy–halo connections, which have enabled us to better understand galaxy formation processes, to infer cosmological parameters, and to probe the properties and distribution of dark matter (see Wechsler & Tinker 2018 for a recent review). Other than the theoretical approaches to modeling galaxy–halo connections, which range from empirical models, such as halo occupation models and conditional luminosity functions (e.g., Jing et al. 1998; Peacock & Smith 2000; Yang et al. 2003), to physical models, such as semi-analytical models or hydrodynamical simulations (e.g., Kauffmann et al. 1993; Springel 2005; Cui et al. 2012; Vogelsberger et al. 2014; Schaye et al. 2015; Cui et al. 2022), there is also a direct way of studying the galaxy–halo connection, by using galaxy groups/clusters, which are defined as sets of galaxies that reside in the same dark matter halos.

Relatively *complete* galaxy group and cluster catalogs have been successfully constructed from various large galaxy surveys, especially at low redshift, where extensive spectroscopic data are

obtained below a (shallow) limiting magnitude, e.g., the 2° Field Galaxy Redshift Survey (2dFGRS; e.g., Merchán & Zandivarez 2002; Eke et al. 2004; Yang et al. 2005b; Tago et al. 2006; Einasto et al. 2007), the Two Micron All Sky Redshift Survey (e.g., Crook et al. 2007; Díaz-Giménez & Zandivarez 2015; Lu et al. 2016; Lim et al. 2017), and, most notably, the Sloan Digital Sky Survey (SDSS), with a friends-of-friends algorithm (e.g., Goto 2005; Merchán & Zandivarez 2005; Berlind et al. 2006; Tempel et al. 2017), with a C4 algorithm (e.g., Miller et al. 2005), and with the halo-based group finder developed in Yang et al. (2005b; e.g., Weinmann et al. 2006; Yang et al. 2007, 2012; Duarte & Mamon 2015; Rodríguez & Merchán 2020). Among these group finders, the halo-based group finder established in Yang et al. (2005b, 2007) has the particular advantage of linking galaxies to their common dark matter halos (e.g., Campbell et al. 2015; Lu et al. 2015; Wang et al. 2020; Tinker 2021). Thus, such constructed group catalogs can be used to study the properties of galaxies as a function of their halo and group properties, and to probe how the member galaxies evolve within different environments (e.g., Yang et al. 2005a; Collister & Lahav 2005; van den Bosch et al. 2005; Robotham et al. 2006; Weinmann et al. 2006; Zandivarez et al. 2006; Wang et al. 2018). Furthermore, these groups associated with dark matter halos can also be used to trace the large-scale structure (LSS) of



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

the universe (e.g., Yang et al. 2005c, 2005d; Coil et al. 2006; Yang et al. 2006).

On the other hand, for deeper surveys that have the aim of probing galaxy properties and their evolutions at high redshifts, the resulting *complete* group or cluster catalogs are still quite limited. Nevertheless, group or cluster catalogs from small-area surveys have been obtained, e.g., from the high-redshift CNOC2 survey (Carlberg et al. 1999), the DEEP2 survey (Gerke et al. 2005), zCOSMOS (Wang et al. 2020), or from photometric galaxy samples, e.g., using red-sequence cluster finders (Koester et al. 2007; Rykoff et al. 2016) or other techniques (Mehmood et al. 2016; Abdullah et al. 2018; Banerjee et al. 2018). Group and cluster catalogs can also be extracted from weak lensing (Miyazaki et al. 2018), X-ray surveys, and Sunyaev–Zeldovich effects (e.g., Hasselfield et al. 2013; Planck Collaboration et al. 2016; Bleem et al. 2020; Hilton et al. 2021). Of the above studies, most of them focus on extracting the most prominent cluster structures in the universe, but lack appropriate assessments of the overall completeness of these clusters. Very interestingly, in a recent study, Yang et al. (2021) extended the halo-based group finder of Yang et al. (2005b) and Yang et al. (2007), so that it could deal with galaxies with spectroscopic and photometric redshifts simultaneously. This new version of the group finder was successfully applied to the DESI image legacy surveys, where complete group catalogs ranging from low-mass isolated galaxies (halos) to massive clusters in the redshift range  $0 < z < 1.0$  with a sky coverage of 18,000 square degrees were constructed.

Due to the lack of observational data, galaxy groups and clusters at redshifts beyond  $z \sim 2$  are rarely studied. Most of the studies have focused on the so-called protocluster population (see Kravtsov & Borgani 2012 and Overzier 2016 for recent reviews). The discovery of protoclusters in observations usually relies on the overdensity of star-forming regions. These regions are common at high redshift and accompanied by intrinsically high luminosity. Different sources are used to trace the star-forming areas, including  $H\alpha$  emitters (Cooke et al. 2014; Katsianis et al. 2017; Darvish et al. 2020; Koyama et al. 2021; Shi et al. 2021),  $Ly\alpha$  emitters (Venemans et al. 2007; Chiang et al. 2015; Cai et al. 2016, 2017, 2019; Jiang et al. 2018; Hu et al. 2021), Lyman-break galaxies (Miley et al. 2004; Toshikawa et al. 2018), submillimeter galaxies (Beuther et al. 2007; Negrello et al. 2017; Cheng et al. 2019), and active galactic nuclei, such as high- $z$  radio galaxies (Galametz et al. 2012; Wylezalek et al. 2013) and quasistellar objects (QSOs; Capak et al. 2011). The distribution of gas has also been applied to find protoclusters (e.g., Miller et al. 2018; Oteo et al. 2018). However, the protoclusters identified with overdensity methods may not necessarily be linked with the most massive halos at those redshifts, and thus not necessarily able to form massive galaxy clusters at  $z = 0$  (Cui et al. 2020). In addition, the Planck all-sky survey provides a large sample of protocluster candidates selected by their dust emission excess in the 545 GHz band (Planck Collaboration et al. 2016). The identification of protoclusters has so far been challenging, due to the low number density and the faintness of distant galaxies (Mulder et al. 2015). Over the past few years, only a few protoclusters have been confirmed through multiwavelength and spectroscopic analysis (e.g., Diener et al. 2015; Wang et al. 2016; Lemaux et al. 2018; Polletta et al. 2021).

On the galaxy observation side, the recently completed Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP;

Aihara et al. 2018) on the Subaru telescope, using the HSC imager (Miyazaki et al. 2018), reaches  $m_i \sim 27.1$  ( $5\sigma$  in  $2''$  apertures) at deep fields. Although a few of the group catalogs using HSC data have been constructed with different methods, such as red-sequence galaxies (Oguri et al. 2017) or weak-lensing techniques (Miyazaki et al. 2018; Hamana et al. 2020; Oguri et al. 2021), and a system of galaxy protoclusters at  $z \sim 4$  has been searched using a sample of  $g$ -dropout galaxies selected from the wide fields (Toshikawa et al. 2018), only a few group or cluster catalogs have been constructed for the deep fields in particular (e.g., the updated data version of Oguri et al. 2017). Ando et al. (2022) searched for cores of protoclusters at  $1 < z < 1.5$ , using photometric data from the HSC-SSP wide and deep fields. In addition to the *grizy* five-band photometries, the  $U$  band contained in the CFHT Large Area  $U$ -band Deep Survey (CLAUDS; Sawicki et al. 2019) allows the bracketing of the Balmer and 4000 Å breaks at intermediate redshift, which improves the performance of the photometric redshift obtained from spectral energy distribution (SED) fitting (e.g., Connolly et al. 1995; Sawicki et al. 1997, 2019).

In this study, we set out to search for groups and protocluster candidates from the joint CLAUDS and HSC-SSP deep data (Sawicki et al. 2019), by adopting the extended halo-based group/cluster finder developed by Yang et al. (2021), paying particular attention to groups/protoclusters at redshifts beyond  $z \sim 2$ . It is also worth noting that the CLAUDS and HSC-SSP deep surveys completely cover the sky areas used to study galaxy evolution in the science themes of the Prime Focus Spectrograph (PFS; Takada et al. 2014). It would be intriguing as well as useful to explore the galaxy properties at high redshifts in combination with our group and protocluster candidate catalogs, together with the galaxies to be observed by the PFS survey.

This paper is organized as follows. In Section 2, we first introduce the data set and the conditions for selecting galaxies, before checking the performance of the photometric redshift with respect to spectroscopic redshift and the distribution of luminosity functions for the selected sample galaxies. The extended version of the halo-based group finder and some basic information about the group catalog constructed from the galaxy catalog are described in Section 3. We discuss the properties of the groups at different redshifts in Section 4. In Section 5, we calculate the number density around the richest groups/clusters at different redshifts, and provide a list of protocluster candidates. Finally, we make our conclusions in Section 6. Throughout the paper, we adopt a  $\Lambda$  cold dark matter (CDM) cosmology, with parameters that are consistent with the Planck 2018 results (Planck Collaboration et al. 2020):  $\Omega_m = 0.315$ ,  $\Omega_\Lambda = 0.685$ ,  $n_s = 0.965$ ,  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.674$ , and  $\sigma_8 = 0.811$ .

## 2. Galaxy Samples

In this section, we describe the data sets used in this study and the criteria for selecting the galaxy samples. We assess the performance of the photometric redshifts and measure the galaxy luminosity functions to evaluate our galaxy samples.

### 2.1. The Photometric Surveys

We use the joint CLAUDS and HSC-SSP data set (Sawicki et al. 2019), which has been applied for studies including UV and  $U$ -band luminosity functions (Moutard et al. 2020) and

**Table 1**  
The Selection Criteria Imposed on the Galaxy Sample Using Only HSC-SSP PDR2

Conditions	True or False	Descriptions
<i>isprimary</i>	True	Identify a single version of an astrophysical object
$g r i z y\_inputcount\_value \geq 2$	False	The number of images contributing at the center
$g r i z y\_mask\_pdr2\_bright\_objectcenter$	False	The source center is close to bright object pixels
$i\_extendedness\_value = 1$	True	Extended object
$g r i z y\_pixelflags\_edge$	False	Too close to an image boundary
$g r i z y\_pixelflags\_interpolatedcenter^*$	False	Interpolated pixel in the source center
$g r i z y\_pixelflags\_saturatedcenter$	False	Saturated pixel in the source center
$g r i z y\_pixelflags\_crcenter^*$	False	Cosmic ray in the source center
$g r i z y\_pixelflags\_bad$	False	Bad pixel in the source center

**Note.** These conditions are referring Oguri et al. (2017) and Mandelbaum et al. (2018). The flags marked \* are only True for the E-COSMOS field.

source classification (Golob et al. 2021). This joint data set is an SExtractor-based multiband catalog, as described in Sawicki et al. (2019). The detection to an object uses the signal-to-noise ( $\Sigma$ SNR) image, which is constructed from all available CLAUDS  $u/u^*$  and HSC-SSP *grizy* images. Once objects are detected by the SExtractor software (Bertin & Arnouts 1996) in the  $\Sigma$ SNR image, the multiband catalog is then created by running SExtractor in dual-image mode, with various measurements recorded for each object, such as positions, fluxes (in Kron, isophotal, and fixed-radius circular apertures), fiducial radii, and ellipticities. Here, we only give a brief description of the combination of the two data sets, while more details can be found in Sawicki et al. (2019) and Moutard et al. (2020).

The CLAUDS and HSC-SSP data set contains  $U + grizy$  six-band data, distributed in four roughly equal-sized ( $\sim 4\text{--}6\text{ deg}^2$ ) fields: E-COSMOS, XMM-LSS, ELAIS-N1, and DEEP2-3. CLAUDS provides the  $U$ -band data with a median depth of  $U_{AB} = 27.1$  ( $5\sigma$  in  $2''$  apertures), covering a total of  $18.60\text{ deg}^2$  in the HSC-SSP deep layer, and a  $1.36\text{ deg}^2$  subarea reaching a depth of  $U_{AB} = 27.7$  within the ultra-deep layer (Sawicki et al. 2019). CLAUDS uses two  $U$ -band filters: the new  $u$  filter is applied in the ELAIS-N2 and DEEP2-3 fields, while the older  $u^*$  filter is adopted in XMM-LSS. The E-COSMOS field uses both the  $u$  and  $u^*$  filters in the central region, and only the  $u$  filter in other areas. The median seeing in the entire deep field of CLAUDS at the  $U$  band is  $0''.92$ .

The HSC-SSP data contains five *grizy* wave bands, with depths of  $g_{AB} \sim 27.3$ ,  $r_{AB} \sim 26.9$ ,  $i_{AB} \sim 26.7$ ,  $z_{AB} \sim 26.3$ , and  $y_{AB} \sim 25.3$  ( $5\sigma$  in  $2''$  apertures) in the deep and ultra-deep regions (Aihara et al. 2019), respectively. The average seeing in the  $i$  band is the best of the five wave bands, reaching  $\sim 0''.62$ . The HSC-SSP data set in total has 14,789,205 objects over  $34.41\text{ deg}^2$ , within which  $18.60\text{ deg}^2$  have CLAUDS  $U$ -band observations. We note that our data set uses an updated version of the HSC-SSP sample, which is based on the second public data release (PDR2; Aihara et al. 2019). This data version increases the galaxy number, but not significantly in the deep and ultra-deep fields. In the following, we describe the procedures used to select our galaxy samples from these data sets.

## 2.2. Galaxy Selection

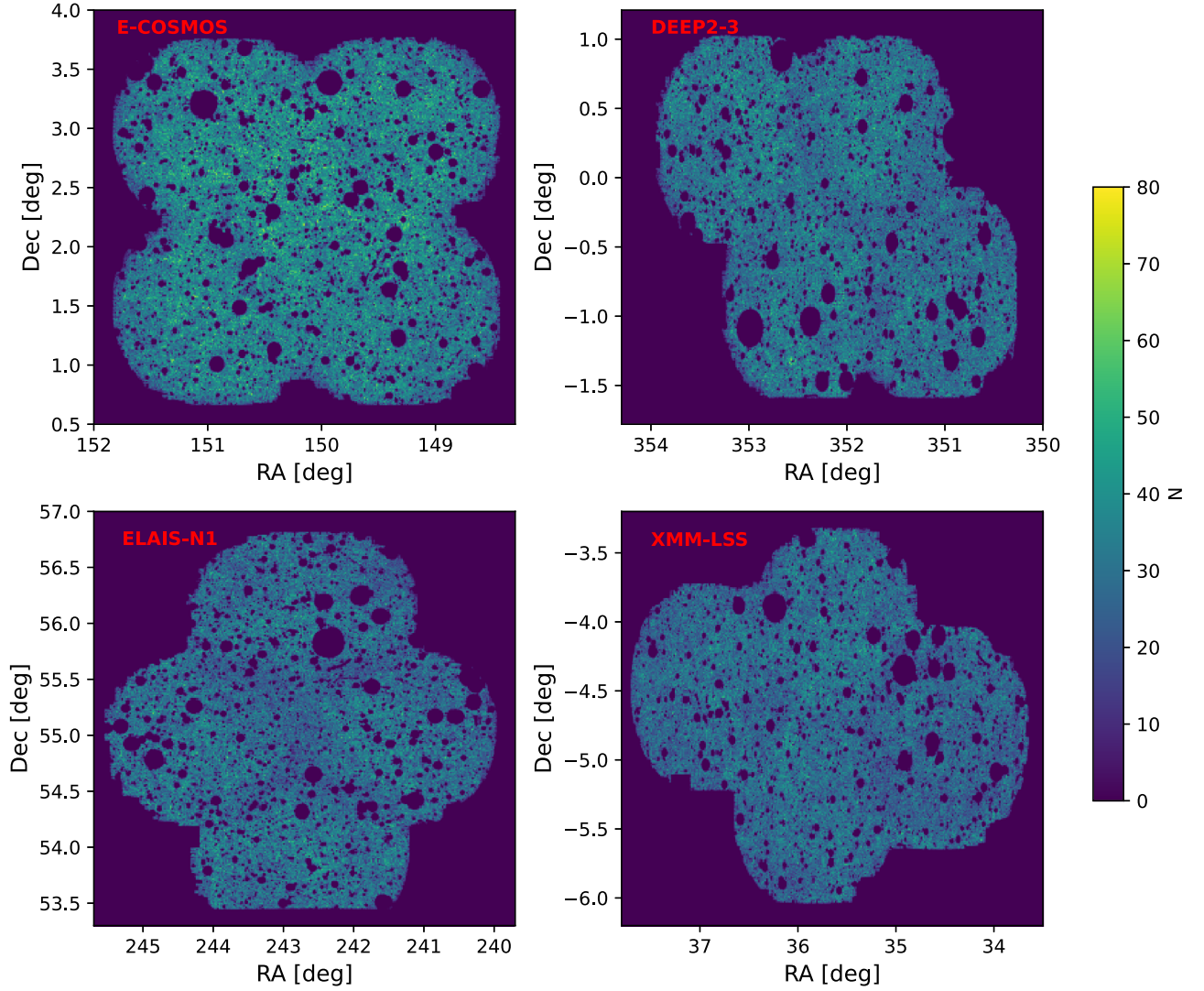
Our galaxy sample selection begins with the rejection of objects with mask flags, to avoid the light influence from bright stars. While the objects in the data set have been divided into stars, galaxies, and QSOs using the gradient boosted trees method (Golob et al. 2021), we only select the objects

classified as galaxies. We also exclude galaxies with *grizy* apparent magnitude less than 0, to avoid spurious photometric redshift determinations. As the  $i$  band is priority observed, and has the smallest seeing of the HSC-SSP observations, we choose to use galaxies observed in the  $i$  band with an apparent magnitude limit  $m_i < 26$  and a magnitude error  $\sigma_i < 1$ . We discard two small but bad areas at the edges of ELAIS-N1 and XMM-LSS. With these initial cuts for the CLAUDS and HSC-SSP data set, we have 7,752,546 galaxies spanning an area of  $33.97\text{ deg}^2$ .

We then try to exclude the unreliable galaxies, due to the pollution of bad pixels or problems still existing in the image processing. However, the original data set provided in Sawicki et al. (2019) does not give such information about these galaxies. We thus construct a new independent galaxy sample based on the HSC-SSP PDR2 database, and apply several pixel flags of images to exclude these unreliable sources. The chosen flags refer to the criteria adopted in Oguri et al. (2017) and Mandelbaum et al. (2018). We first require the flag *isprimary* to be *True*, to identify a single version of each astrophysical object. We throw out sources close to bright object pixels at *grizy* bands. The object type is determined as a galaxy by setting the star-galaxy separation parameter  $i\_extendedness\_value = 1$  at the  $i$  band. The number of visits to each object is indicated with the *inputcount* parameter, where we choose  $inputcount \geq 2$  for all *grizy* bands. Moreover, we also discard the possibly polluted galaxies with these impositions—objects too close to an image boundary or those that have interpolated, saturated, bad, or cosmic hit pixels in the source center at any broad bands. Because the ultra-deep fields undergo 100 or more visits, any one of which could be affected by a cosmic ray, this results in a substantial chance of the object being excluded, i.e., there is a lower source density in the ultra-deep fields than in the deep fields, after excluding the influenced sources. We thus do not apply the interpolated and cosmic ray flags for the E-COSMOS field, which suffers from this effect heavily, though the effects of cosmic rays are minor on coadds (Aihara et al. 2022). After applying the selection criteria as listed in Table 1, we obtain a galaxy checking sample with a total of 11,177,216 sources.

Finally, we match the initially selected CLAUDS and HSC-SSP galaxy sample with this checking sample obtained from HSC-SSP PDR2 by asking the agreement of their coordinates to be less than  $1''$ . By selecting galaxies with photometric redshift  $0 < z_{photo} < 6$ , we finally obtain a galaxy catalog with 5,607,052 galaxies. The distributions of the matched galaxies in the final sample over the four separated fields are shown in



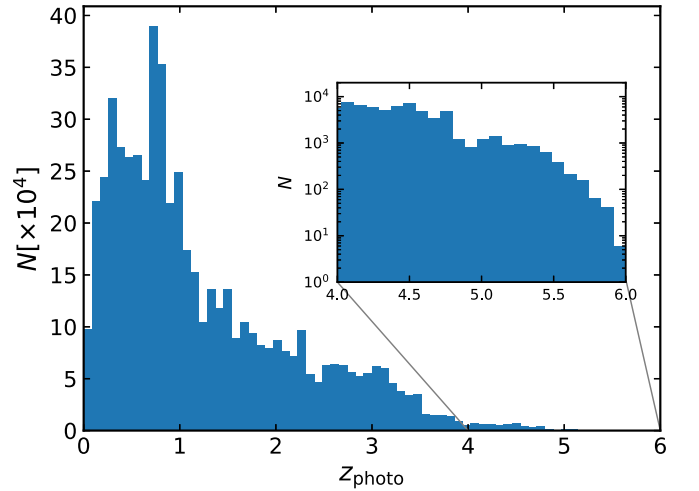


**Figure 1.** The sky coverage of the galaxies, distributed over four separated fields: E-COSMOS, DEEP2-3, ELAIS-N1, and XMM-LSS. The galaxy number count in each pixel with an area of about  $1.4 \times 10^{-4} \text{ deg}^2$  is coded with the color bar. The empty circles inside the coverage of the galaxies correspond to the masked areas.

the different panels in Figure 1. We present the redshift distribution of selected galaxies in Figure 2.

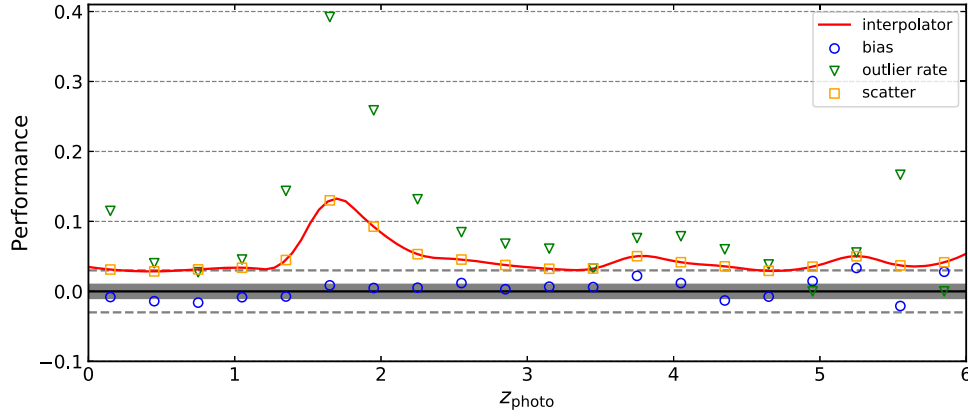
### 2.3. Assessing the Photometric Redshift Quality

The photometric redshift,  $z_{\text{photo}}$ , in our galaxy sample is calculated using a color space nearest-neighbor machine-learning technique (hereafter, kNN; Sawicki et al. 2019) combined with the template-fitting code LE PHARE (Arnouts et al. 2002; Ilbert et al. 2006). Specifically, the kNN method uses the 30-band COSMOS photometric redshifts from Laigle et al. (2016) as a training set. The 50 nearest neighbors around each object are determined by kNN in color space. Then, each object is fitted with a weighted Gaussian kernel density estimator with the weighted redshifts of these neighbors. This method obtains redshifts with a low scatter and bias, on average, but suffers from more outliers. The photometric redshifts from LE PHARE are computed using the template library of Coupon et al. (2015), and with a consideration of the four extinction laws as described in Ilbert et al. (2006). The final photometric redshifts are obtained by combing the outputs



**Figure 2.** The number distribution of the galaxies as a function of photometric redshift in our final sample. The inserted panel shows the number distribution of the galaxies at  $4 < z_{\text{photo}} < 6$  in particular.





**Figure 3.** Assessing the quality of the photometric redshifts with respect to the spectroscopic redshifts as a function of photometric redshift. The blue circles, green triangles, and orange squares represent the bias, scatter, and outlier rate, respectively. The gray shadows and dashed lines show the  $\pm 0.01$  and  $\pm 0.03$  regions, respectively. The red line is the second-order spline interpolation for the scatter parameter.

from the kNN method and LE PHARE, i.e., the photometric redshift values of the outliers in the kNN photoz catalog were replaced with the values from LE PHARE. The details of the computation are introduced in Moutard et al. (2020).

As the group finder uses both the value and the error of the photometric redshifts (see Section 3.1 for details), we investigate the quality of the photometric redshifts by comparing with the subsample of high-quality spectroscopic redshifts,  $z_{\text{spec}}$ . We use quantities including bias, scatter, and outlier rate to assess the quality of the photometric redshifts. The bias is defined as the median value of  $(z_{\text{photo}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ . The scatter is estimated using the normalized median absolute deviation:  $\text{median}(|z_{\text{photo}} - z_{\text{spec}}|/(1 + z_{\text{spec}}))/0.6745$ . The outlier rate is the fraction of galaxies with  $|z_{\text{photo}} - z_{\text{spec}}|/(1 + z_{\text{spec}}) > 0.15$  in each photometric redshift bin.

Based on the spectroscopic redshifts of 65,135 galaxies from a compilation of surveys (Lilly et al. 2007; Bradshaw et al. 2013; Le Fèvre et al. 2013; McLure et al. 2013; Comparat et al. 2015; Kriek et al. 2015; Silverman et al. 2015; Masters et al. 2017; Tasca et al. 2017; Scodreggio et al. 2018) in the CLAUDS and HSC-SSP sample, we assess the photometric redshift quality in our galaxy sample. The performance as a function of photometric redshift is shown in Figure 3. The bias is constrained around  $\pm 0.01$  at  $z_{\text{photo}} < 3.5$  and within 0.03 at higher redshifts. The scatter remains level at 3% across whole redshift range, except it reaches  $\sim 12\%$  at  $z \sim 1.8$ , where the outlier rate is as high as about 40%. Thus, in general, the groups that we extracted at redshift  $z \sim 1.8$  should be less reliable than those at other redshift ranges. We use a second-order spline interpolation method to fit the scatter as a function of redshift, which is shown as the red line in Figure 3. We use this fitting result to describe the photometric redshift error of each galaxy for our group finder. If a galaxy has a spectroscopic redshift, we replace the photometric redshift with the spectroscopic redshift, and set the redshift error as 0.0001. As tested in Yang et al. (2021), galaxy groups, especially massive ones, can be reliably detected for galaxy samples with a photometric redshift error at the 3% level. We note that once the PFS starts its operation, we will keep updating the photometric redshifts with spectroscopic redshifts, and hence improve the resulting group catalogs. We expect the completeness and purity, and, more importantly, the redshift accuracy of the groups in the updated versions to be improved.

#### 2.4. Galaxy Luminosity Functions

As the halo-based group finder uses the group/galaxy luminosity as a proxy for halo mass estimates, it is important to check if there are spuriously bright galaxies due to redshift errors, etc. The absolute magnitude of each galaxy is calculated with its apparent magnitude and redshift, following the formula

$$M_i - 5 \log h = m_i - 5 \log D_L(z) - 25, \quad (1)$$

where  $D_L(z)$  is the luminosity distance in units of  $h^{-1}\text{Mpc}$ . We obtain the luminosity of each galaxy  $L$  using the formula

$$\log(L/h^{-2}L_\odot) = 0.4 \times (4.52 - M_i), \quad (2)$$

where 4.52 is the  $i$ -band absolute magnitude of the Sun.

The galaxy luminosity function measures the comoving number density of the galaxies as a function of luminosity, which is one of the most essential tools for characterizing the galaxy population. As our galaxy sample covers a large redshift range  $0 < z < 6$ , we divide the galaxies into 15 redshift bins, each with a bin width  $\Delta z = 0.4$ . We calculate the galaxy luminosity functions at  $i$  band in all redshift bins. Note that more comprehensive studies of the luminosity functions based on the CLAUDS and HSC data set have been carried out at UV and  $U$  bands (Ono et al. 2018; Moutard et al. 2020; Harikane et al. 2022; C. Liu et al. 2022, in preparation). Following Yang et al. (2021), here we use the  $V_{\text{max}}$  method to calculate the luminosity function,

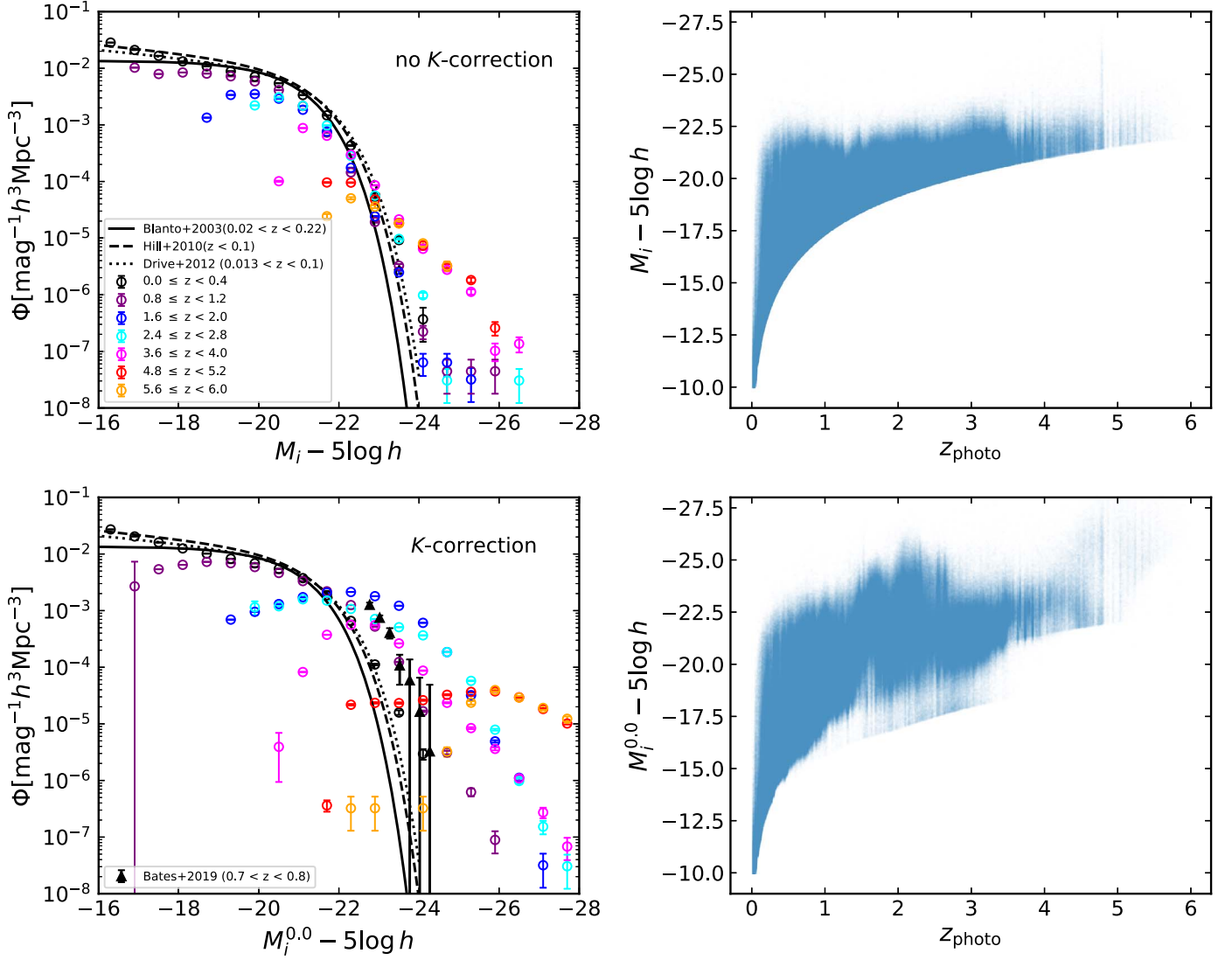
$$\phi(M_i) d\log M_i = \sum_j 1/V_{\text{max}}, \quad (3)$$

where the summation is performed for all the galaxies in a given redshift bin. The  $V_{\text{max}}$  is the comoving volume within the given redshift bin, which is computed according to the maximum redshift,  $z_{\text{max}}$ , where the apparent magnitude of galaxies can be observed:

$$V_{\text{max}} = V(\min[z_{\text{max}}, z_{\text{bin,up}}]) - V(z_{\text{bin,low}}), \quad (4)$$

where  $z_{\text{bin,up}}$  and  $z_{\text{bin,low}}$  are the lower and upper limits of the corresponding redshift bin, respectively.

In Figure 4, we present the luminosity functions and magnitude–redshift distribution of our sample galaxies selected from different redshift bins at  $i$  band. We specifically compare the results for galaxies with and without  $K$ -correction in calculating the absolute magnitudes of the galaxies. Based on



**Figure 4.** Left panels: galaxy luminosity functions in different redshift bins. The results for galaxies with or without  $K$ -correction are shown in the lower and upper panels, respectively. The errors are derived from the square root of the sum of the weights squared (the Poisson error). Seven redshift bins with 0.4 intervals are shown with different colors, as labeled in the legend in the top panel. Comparisons with other observations at low redshifts in the  $i$  band include Blanton et al. (2003; solid black lines), Hill et al. (2010; dashed black lines), and Driver et al. (2012; dotted black lines). The black triangles show a relatively higher-redshift observation result with  $K$ -correction from Bates et al. (2019). Right panels: the absolute magnitude vs. photometric redshift distribution of a subset of randomly selected sample galaxies. The results for the absolute magnitudes with or without  $K$ -correction are shown in the lower and upper panels, respectively.

the HSC-SSP *grizy* five-band magnitudes and the CLAUDS  $U$ -band magnitude (if available), we calculate  $K$ -correction in the  $i$  band to redshift  $z \sim 0$  using the “Kcorrect” model of Blanton & Roweis (2007). In the top left panel of Figure 4, we present the galaxy luminosity functions without consideration of  $K$ -correction at seven redshift bins, as indicated. The distribution of the low-redshift galaxies shows a profile consistent with other observations at  $i$  band (Blanton et al. 2003; Hill et al. 2010; Driver et al. 2012), as shown with the black lines, while at high redshifts, the galaxy luminosity functions show some significant enhancements at the bright ends. Shown in the lower left panel of Figure 4 are the luminosity functions that are  $K$ -corrected to redshift  $z=0$ . We include a comparison with a relatively high-redshift ( $0.7 < z < 0.8$ ) observation from Bates et al. (2019), shown with black triangles, which matches well with our results after taking  $K$ -correction into account. There is a strong evolution in the luminosity functions at  $i$  band with  $K$ -correction taken into account (Mobasher et al. 1996). Overall,

the luminosity function with  $K$ -correction in the lowest-redshift bin is similar to that without the consideration of  $K$ -correction. However, in the higher-redshift bins, the luminosity functions show extraordinary enhanced bumps at the bright ends. The overall behavior deviates from the typical Schechter functional form, and there are galaxies with  $M_i < -28$ .

The differences can also clearly be seen in the magnitude–redshift distributions of the galaxies, as shown in the right panels of Figure 4. For galaxies at  $z > 3.5$ , there appears to be a gap for the faint galaxies, as the  $K$ -correction model we used is most applicable to  $z \sim 2$  (Blanton & Roweis 2007). The galaxies at different redshifts have different  $K$ -correction scatter ranges.

Note that in this paper, we are using galaxy luminosity as proxy for halo mass estimation, not trying to provide a coherent and accurate galaxy luminosity function measurement. In addition, as we have separated the galaxies into small redshift bins, without  $K$ -correction only means that we are measuring

the galaxy luminosity functions in the *observed* frame around the median redshifts of the galaxies in these redshift bins. Furthermore, the derived halo mass does not strongly depend on the luminosity function, as we will use an abundance matching method to obtain the halo mass (see the further details in next section). Given these situations, we decide to use the galaxy absolute magnitudes without  $K$ -correction in this study.

### 3. The Method and Basic Quantities

In this section, we first give a brief description of the method that we used to extract the groups from our galaxy sample. After the construction of the group catalog, we present some basic quantities of it.

#### 3.1. The Halo-based Group Finder

The group finder that we used in this study is a halo-based method developed in Yang et al. (2005b, 2007). The galaxy–dark matter connection has been extensively studied in theories (e.g., halo occupation distribution, stellar-to-halo mass ratios) that provide the foundation for this group finder. The first version of the group finder was only applicable to galaxies with spectroscopic redshifts, and hence it was applied to 2dFGRS and SDSS data to search for groups at low redshifts (Yang et al. 2005b, 2007). Recently, the group finder has been improved, to make it applicable to both spectroscopic and photometric redshifts (Yang et al. 2021). The performance of the new version of the group finder has been tested against mock galaxy redshift samples. It turns out that the group finder is stable and reliable in constructing group catalogs for galaxy samples with photometric redshift error  $\sim 3\%$ . This extended version will allow us to probe group contents over large redshift ranges using photometric redshift galaxy data. Here, we give a brief description of the main steps in this method (further details can be found in Yang et al. 2021).

The group finder starts from the assumption that each galaxy is a group candidate. To alleviate the impact of galaxy incompleteness due to the magnitude limit cut, the possible impacts of galaxy luminosity evolution, and the effect of not taking  $K$ -correction into account, here we separate our sample galaxies into 15 redshift bins, each with an interval of  $\Delta z = 0.4$ . We measure the total luminosity of each group by summing up the luminosity of all the member galaxies, and compute the cumulative group luminosity functions in each redshift bin. Meanwhile, the cumulative halo mass function is obtained from the analytic model prediction by Sheth et al. (2001), corresponding to the median redshifts of the groups in each bin. We determine the mass-to-light ratios of the groups in each redshift bin with cumulative halo mass functions and group luminosity functions, using the abundance matching method (Yang et al. 2007).

Then, each tentative group is assigned a halo mass  $M_L$ , based on the upper mass-to-light ratio, using interpolation techniques. With the halo mass, each group can have a halo radius and velocity dispersion along the line of sight. The halo radius is defined as 180 times the average matter density of the universe, expressed as

$$r_{180} = 0.781 h^{-1} \text{Mpc} \left( \frac{M_L}{\Omega_m 10^{14} h^{-1} M_\odot} \right)^{1/3} (1 + z_{\text{group}})^{-1}, \quad (5)$$

where  $z_{\text{group}}$  is the redshift of the group center. The line-of-sight velocity dispersion of a dark matter halo is obtained using the

fitting function of van den Bosch et al. (2004), with slight modifications to make it suitable to  $\Lambda$ CDM cosmology, by using other  $\Omega_m$  values:

$$\sigma_{180} = 632 \text{s}^{-1} \text{km} \left( \frac{M_L \Omega_m}{10^{14} h^{-1} M_\odot} \right)^{0.3224}. \quad (6)$$

The group membership updates begin from the most massive one, by taking the luminosity-weighted group center as the halo center and assuming that the distribution of the member galaxies in phase-space follows that of the dark matter particles. The probability of a galaxy being a member galaxy can be written as

$$P_M(R, \Delta z) = \frac{H_0}{c} \frac{\Sigma(R)}{\bar{\rho}} p(\Delta z), \quad (7)$$

where  $R$  is the projected distance from the group center,  $\Delta z = z - z_{\text{group}}$ ,  $c$  is the velocity of light,  $\Sigma(R)$  is the projected surface density for a Navarro–Frenk–White halo (Navarro et al. 1997), and  $p(\Delta z)$  is a Gaussian function form to describe the redshift distribution of the galaxies within the halo (see the details in Yang et al. 2021). Here, we have  $\sigma = \max(\sigma_{180}, c\sigma_{\text{photo}})$ , where  $\sigma_{\text{photo}}$  is the typical photometric redshift error, as described by the solid line shown in Figure 2. Note that in our galaxy sample, if a galaxy has a spectroscopic redshift, we assign it with a  $\sigma_{\text{photo}} = 0.0001$  value. For this galaxy, the  $\sigma = \sigma_{180}$  value will automatically be used, while for the majority of galaxies with only photometric redshifts,  $\sigma = c\sigma_{\text{photo}}$ .

Next, we assign the galaxies to candidate groups with a judgment between  $P_M(R, \Delta z)$  and  $B\sigma_{180}/\sigma$ . If  $P_M(R, \Delta z) \geq B\sigma_{180}/\sigma$ , the galaxy will be assigned to the group. Here, the background value  $B$ , independent of halo mass, perceptively quantifies the threshold of the redshift space density contrast of the groups (Yang et al. 2005b). We adopt the theoretically gauged parameter 10 as the  $B$  value during the group finding. Decreasing this background value may slightly increase the richness of the groups. The ratio  $\sigma_{180}/\sigma$  is used to account for the decrease in the density contrast caused by the photometric redshift error.

After assigning all the galaxies to groups, we update the group centers and luminosities, recalculate the halo information, then find the member galaxies again. The iterations stop until there are no more changes to group memberships. Finally, we start from the beginning again to perform another iteration, aimed at the convergence of the mass-to-light ratios, which normally need three to four iterations.

#### 3.2. Survey Edge Effect

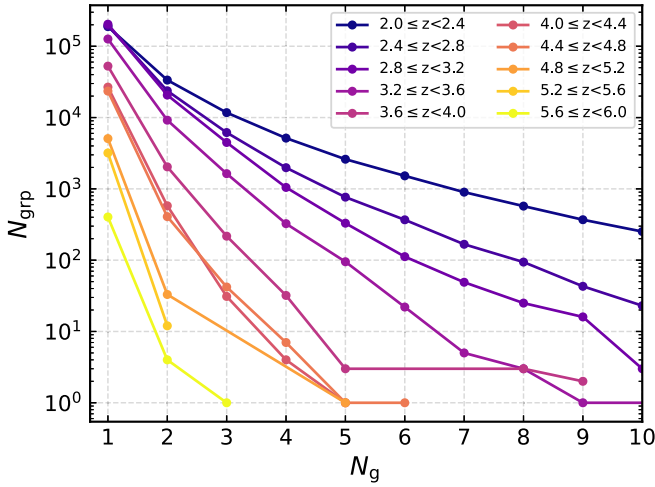
As we can see in Figure 1, the survey geometries of our galaxy catalogs are quite complicated. We follow Yang et al. (2007) to provide a parameter,  $f_{\text{edge}}$ , to quantify the survey edge effect of the groups. For this purpose, we randomly distribute 200 points within the radius ( $r_{180}$ ) of each halo. Then, we remove those random points that are outside the survey region, according to the masks of the CLAUDS and HSC-SSP data set. For each group, we calculate the number of remaining points,  $N_{\text{remain}}$ , and define the fraction  $f_{\text{edge}} \equiv N_{\text{remain}}/200$  to gauge the volume of the group that lies within the survey edges. About 84% of the groups in our catalog have an  $f_{\text{edge}}$  value larger than 0.9. For the groups with  $M_h > 10^{13} h^{-1} M_\odot$ ,  $\sim 72\%$  of the groups have  $f_{\text{edge}} > 0.9$ . This



**Table 2**Number, Percentage, and Median Log Halo Mass of the Groups, with Different Numbers of Member Galaxies ( $N_g$ ) and within Different Redshift Ranges

Redshift (1)	$N_g \geq 1$ $N_{\text{grp}} (\log M_h)$ (2)	$N_g \geq 2$ $N_{\text{grp}} (\%, \log M_h)$ (3)	$N_g \geq 3$ $N_{\text{grp}} (\%, \log M_h)$ (4)	$N_g \geq 5$ $N_{\text{grp}} (\%, \log M_h)$ (5)	$N_g \geq 10$ $N_{\text{grp}} (\%, \log M_h)$ (6)
$0 < z < 6$	2,232,134 (11.86)	669,209 (30.0, 12.20)	402,947 (18.1, 12.32)	204,099 (9.1, 12.50)	68,711 (3.1, 12.79)
$z \geq 1$	1,617,926 (11.87)	379,346 (23.4, 12.27)	178,326 (11.0, 12.49)	65,034 (4.0, 12.77)	13,576 (0.8, 13.15)
$z \geq 2$	955,893 (11.82)	132,022 (13.8, 12.22)	41,815 (4.4, 12.50)	8,993 (0.9, 12.83)	919 (0.1, 13.25)
$z \geq 3$	374,688 (11.78)	28,284 (7.5, 12.15)	5,403 (1.4, 12.40)	404 (0.1, 12.79)	5 (0.001, 13.30)
$z \geq 4$	59,997 (11.87)	1,125 (1.9, 12.22)	89 (0.1, 12.55)	4 (0.006, 13.00)	0
$z \geq 5$	6,553 (11.95)	32 (0.5, 12.20)	1 (0.01, 12.78)	0	0

**Note.** Column (1) lists the different redshift bins. Columns (2), (3), (4), (5), and (6) list the number of groups,  $N_{\text{grp}}$ , with at least 1, 2, 3, 5, and 10 members, respectively. The values for the percentage and log median halo mass in each redshift and richness bin are listed in the parentheses.



**Figure 5.** The distribution of the number of groups  $N_{\text{grp}}$  as a function of the number of member galaxies  $N_g$ . The different colors represent the different redshift ranges, as shown in the top right legend.

parameter is provided in our catalog in the event that a study needs to consider this edge effect.

### 3.3. Basic Quantities

After applying the extended halo-based group finder to the selected CLAUDS and HSC-SSP galaxy sample, we obtain a total of 2,232,134 groups. There are 402,947 and 68,711 groups containing at least three and 10 member galaxies, respectively. As our group catalog covers a wide redshift range, it would be interesting to see those numbers and masses in different redshift ranges. We list the number of groups, the percentages of the groups with respect to the total population, and the median halo masses of the groups, with the different numbers of member galaxies and redshift ranges, in Table 2. More explicitly, we present the number of groups as a function of the number of member galaxies in different redshift bins in Figure 5 for groups with redshift  $z \geq 2$ . As redshift increases, the number of member galaxies decreases dramatically. Indeed, at  $z \geq 5$ , most groups are isolated galaxies, and 31 groups have only two member galaxies, with one group having three members.

As an illustration, we show in Figure 6 the projected distributions of the groups selected from a small sky coverage area. The y-axis and x-axis represent the transverse distance in the RA direction from the field center and the line-of-sight distance, respectively. It is quite obvious that the richness of the

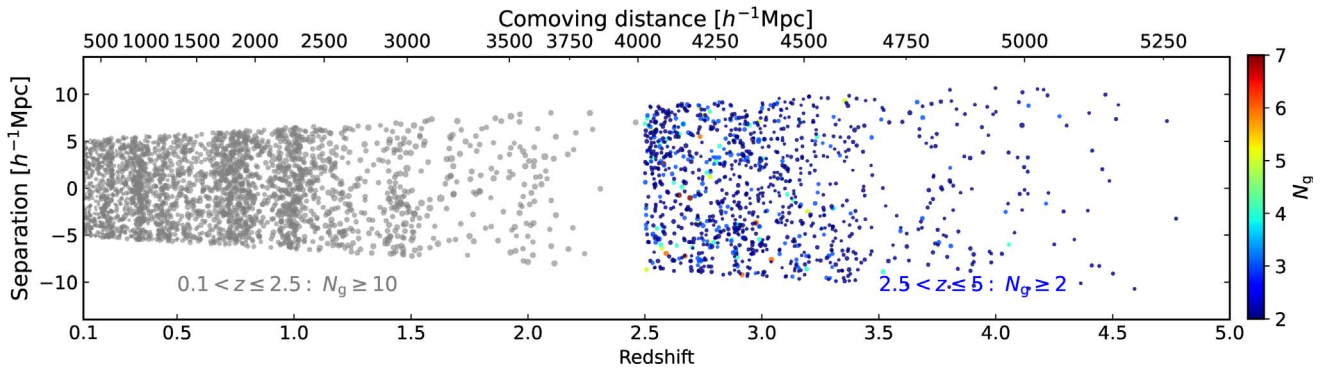
groups decreases with the increase in redshift. Most of the groups with at least 10 members are below redshift  $z = 2.5$ .

We note that in our galaxy sample, the photometric redshifts may be contaminated by starlight, etc., in certain bands. In particular, we examine the distribution of the groups at  $z \geq 4$  in the images from HSC-SSP PDR2. Most groups are well established, as clearly shown in the images, and present a peak in galaxy numbers around the group center (e.g., Figure 9). However, we also find that some member galaxies of the groups in the images suffer from contaminations, such as scattered light, satellite trails, and the long wings of bright stars (Aihara et al. 2018, 2019), even though the processing pipeline (Bosch et al. 2018) used in HSC-SSP PDR2 has improved a lot. We believe that the groups determined in these cases are not reliable. Although we have excluded some polluted sources by setting the selection flags, a few galaxies located in the contaminated areas are hard to classify and were left in our sample, occupying a small fraction of the total galaxy sample. About 17% (2/12)<sup>11</sup> and 21% (7/34) of the groups with at least two member galaxies at redshifts  $5.2 < z < 5.6$  and  $4.8 < z < 5.2$ , respectively, suffer from this problem. This problem is especially severe for the groups with at least three members at redshift  $4.4 < z < 4.8$ : 71% (36/51) of the groups are located in the contaminated areas (see also the top left panel of Figure 4 for the enhanced number of very bright galaxies at redshift  $z \sim 4.8$ , a hint of such contamination). Thus, it is better to visually inspect the images before studying these individual groups, especially at the particular redshift bin  $4.4 < z < 4.8$ . Besides, at  $4.0 < z < 4.4$ , only four cases in a sample of 36 groups with at least three members suffer from this problem. In general, by investigating the redshift distribution of the contaminated galaxies in the images, we find that most galaxies are at  $z \gtrsim 4.4$ . We suppose that these contaminated galaxies are prone to be assigned high redshifts during their redshift SED fitting. Nevertheless, the current galaxy sample is already one of the best deep photometric redshift catalogs that we currently have, and PFS observations will provide massive spectroscopic redshifts in these regions in the near future. These contaminated groups are excluded from the following analysis and plots.

### 4. Properties of Galaxy Groups at Different Redshifts

We set out to investigate the properties of the groups, paying special attention to the evolutionary trend of the groups as a function of redshift.

<sup>11</sup> The number of groups containing contaminated members divided by the total number of such groups.



**Figure 6.** The projected distribution of a selection of groups in a transverse vs. line-of-sight direction plane. The groups with at least 10 member galaxies at  $0.1 < z \leq 2.5$  are shown with gray dots, while those with at least two member galaxies at  $2.5 < z \leq 5$  are color coded by their number of member galaxies. The size of the dot is proportional to the log mass of the group.

#### 4.1. Galaxy Pairs at $z \sim 5$

We start our analysis from high redshift. As shown in Figure 5, and listed in Table 2, the vast majority of our groups at  $z > 5$  contain only one member galaxy. There are in total only a few tens of galaxy pairs. Comparing to the rich systems at low redshifts, the poor systems found at such high redshift generally have two origins: one is the observation selection effect and the other is the growth of halos. We carry out a rough investigation into the observational selection effect, i.e., that faint galaxies may not be observed at high redshifts, on our determination of the high-redshift group members. We select the 100 richest groups at  $z \sim 1$ , with a median value of 197 members. Then, we move these groups to higher redshifts and estimate the number of remaining group members according to the magnitude limit, without a consideration of the halo evolution. At  $z = 5.4$ , we find that most groups possess about two members, which is roughly consistent with our expectation. However, taking into account the strong evolution of the galaxy luminosity functions at redshift  $\gtrsim 3.6$ , those richest systems should contain more bright galaxies than the ones that have been observed, indicating quite a strong growth or accretion of other member galaxies at a later stage.

We present four of these groups at  $z > 5$  in Figure 7. The images are taken from the HSC-SSP PDR2 deep observations at the  $i$  band. We mark the galaxies around the group center in our galaxy sample with solid circles color coded by  $z_{\text{photo}}$ , with their values being written to the upper right of their positions. We use red squares to indicate the member galaxies. The center and halo size (with radius  $r_{180}$ ) of each group are presented with a red star and dashed magenta circle, respectively. The information about the group, including its redshift  $z$ , number of member galaxies  $N_g$ , halo radius  $r_{180}$ , and mass  $M_h$ , is given in the top right corner of each panel.

As the group masses are estimated using the halo abundance matching method (see Yang et al. 2012 for an illustration of the halo mass functions at different redshifts), the masses of these groups are around  $\sim 10^{12} h^{-1} M_\odot$ , with a mean comoving radius of  $\sim 0.28 h^{-1} \text{Mpc}$ . By checking the neighboring galaxies, we find that the member galaxies are well determined by our group finder, i.e., no galaxies are seen with similar redshifts near the groups that might be missed.

Comparing to the halo radius, the member galaxies here seem to have relatively large separations, meaning that, at such high redshift, galaxy or halo major mergers are not frequent. In

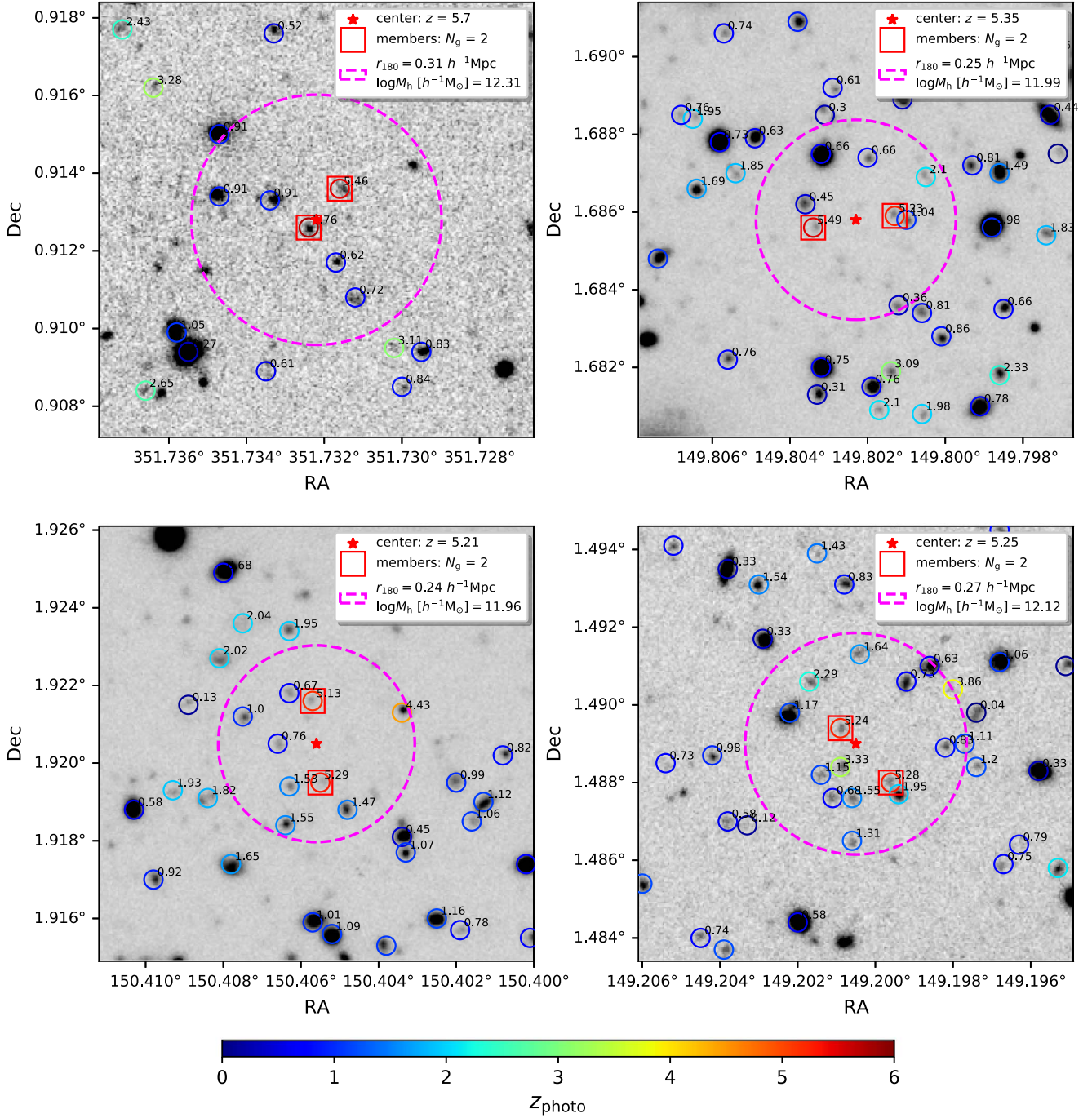
addition, the galaxy pairs in most cases do not show significant differences, i.e., the distinction between one bright central galaxy (BCG) and another faint satellite galaxy (FSG). Quite interestingly, evidence has shown that proto-BCGs at  $z \sim 1.6$  have formed at high redshift through equal-mass mergers of massive galaxies (see, e.g., Sawicki et al. 2020). Theoretical models (e.g., De Lucia & Blaizot 2007; Contini et al. 2015) also predict such merger formation scenarios. Thus, we speculate that the galaxy pairs at  $z \sim 5$  might be the progenitors of those equal-mass mergers, whose later descendants are ultramassive quiescent galaxies at lower redshifts (Sawicki et al. 2020).

#### 4.2. Triple-galaxy Systems at $z \sim 4$

As redshift decreases to  $z \sim 4$ , we have more galaxies in our sample, but the numbers of galaxies determined as members of groups are still small. Apart from four groups with at least five members, the most massive groups at this redshift range roughly have three members. Similar to Figure 7, we show in Figure 8 the galaxy distributions around four typical groups with three members at  $z \sim 4$ . These groups with halo mass around  $\sim 10^{12.3} h^{-1} M_\odot$  have a mean comoving radius of  $\sim 0.32 h^{-1} \text{Mpc}$ .

Here again, compared to the member galaxies enclosed in the magenta dashed circles, there are no obvious galaxies at similar redshifts being missed by the group finder. Comparing to those galaxy pairs at redshift  $z \sim 5$ , the triple-galaxy systems here have the following features:

1. Of the three galaxies, a pair of galaxies seem to be very close, with the third being somewhat further away. Within the close pairs, there is one galaxy with a relatively higher luminosity (a deeper grayscale compared with the other galaxies), i.e., the distinction between BCG and FSG starts to be significant. This might be due to the different star formation efficiencies associated with the halo potential, gas reservoir, and cooling rate.
2. Associating these triple systems with those galaxy pairs with similar abundance at higher redshift whose member galaxies are quite similar, this might indicate that the faint galaxy (subhalo) in the close pair might have been accreted quite early in the halo, but has not been observed at higher redshift due to the flux limit.



**Figure 7.** Four cases of groups at  $z > 5$  in the HSC-SSP PDR2 deep  $i$ -band images. The galaxies around the group center in our galaxy sample are marked with solid circles color coded by  $z_{\text{photo}}$ , with their values being written to the upper right of their positions, while the member galaxies are indicated by red squares. The centers and halo radii of the groups are presented with red stars and dashed magenta circles, respectively. Note that the center here is the luminosity-weighted center, as defined in the text. The information about the group, including its redshift  $z$ , number of member galaxies  $N_g$ , halo radius  $r_{180}$ , and mass  $M_h$ , is given in the top right corner of each panel.

3. The faint galaxy in a close pair is about to—but not yet—merge into the brighter galaxy around this redshift. Since it is much fainter than the primary galaxy, this might correspond to a minor merger.

#### 4.3. Distribution of Member Galaxies at $z \sim 3$

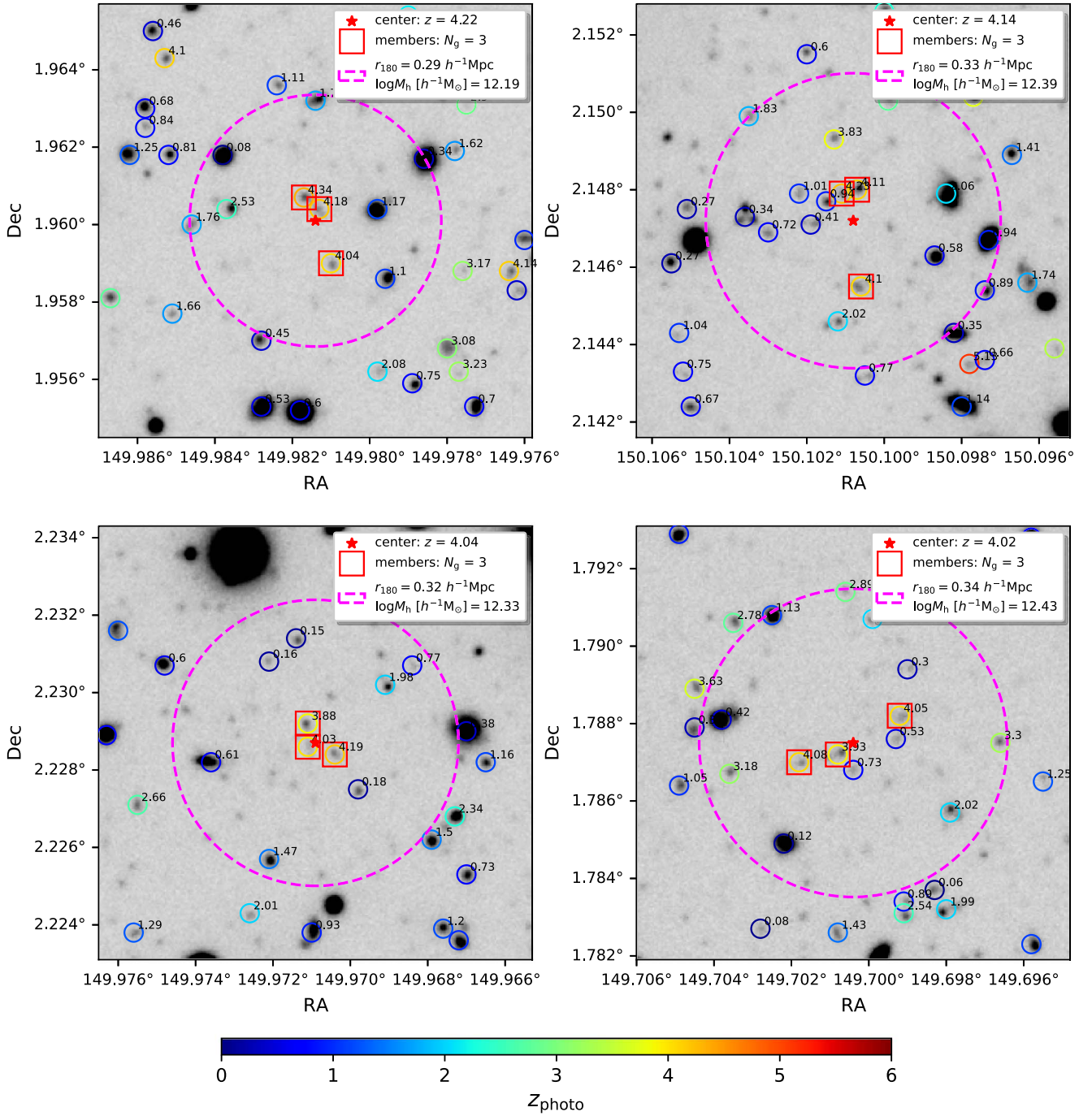
When redshift decreases to  $z \sim 3$ , the groups possess more member galaxies. Overall, at  $z \geq 3$ , there are more than 400 groups with at least five member galaxies. There are about 144

groups with at least six members. In Figure 9, we present four cases of galaxy distributions around groups at  $z \sim 3$ . Three of them have six member galaxies and one has seven members. The mass of the groups is in the range of  $10^{12.59} - 10^{12.74} h^{-1} M_\odot$ , with a mean virial radius ( $r_{180}$ ) of  $\sim 0.42 h^{-1} \text{Mpc}$ .

Comparing to the triple-galaxy systems at redshift  $z \sim 4$ , the galaxy groups here have the following features:

1. There are one or two prominent brightest galaxies in each group. These are quite distinct from the other galaxies,





**Figure 8.** Similar to Figure 7, but for groups at  $z \sim 4$ . Here we show groups with three member galaxies.

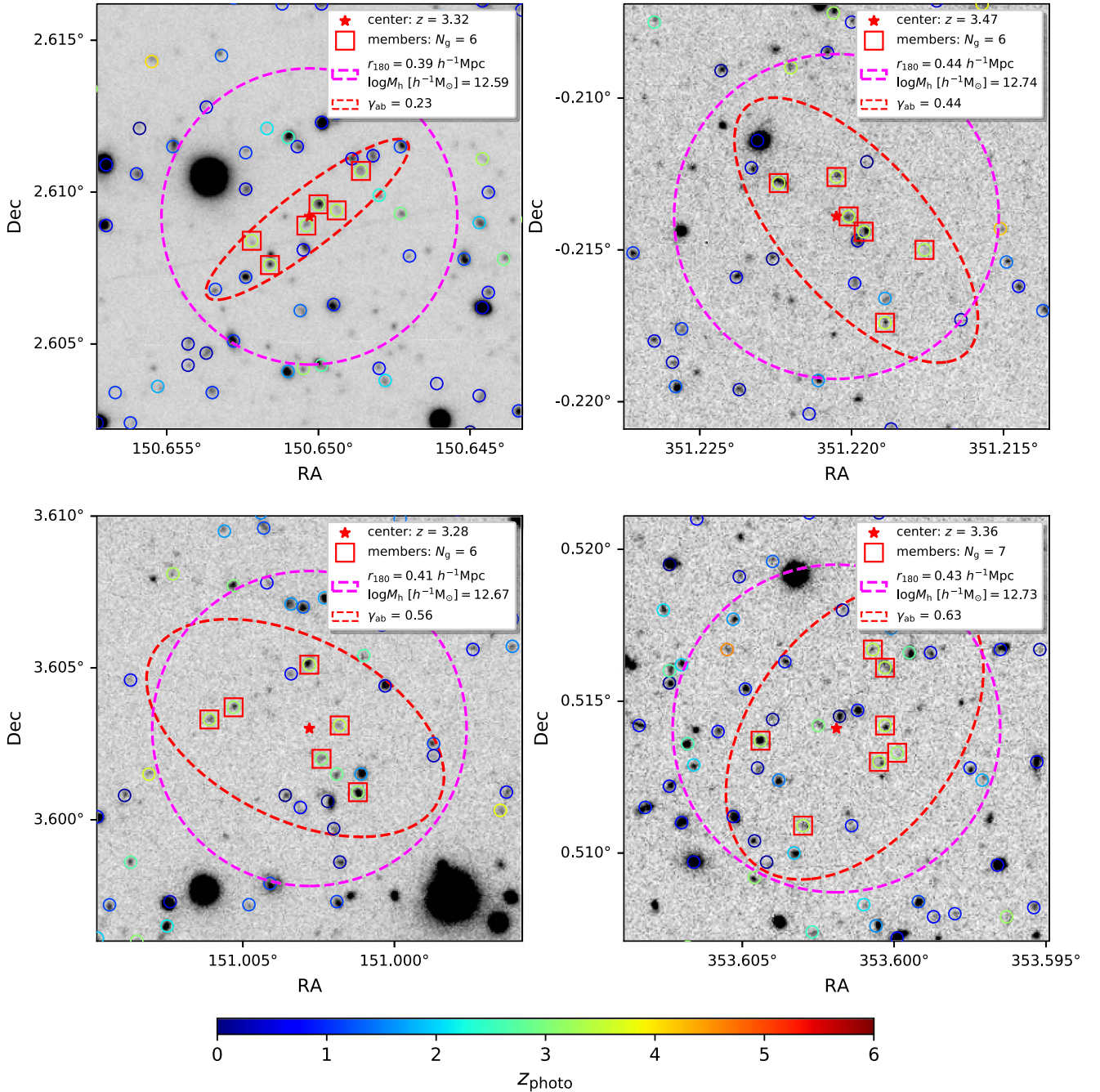
which might indicate that these galaxies, or at least some of them, may have devoured their closest neighboring FSGs and thus become the dominant BCGs.

2. In the cases where there are two prominent brightest galaxies in a group, they usually have quite a large separation, with or without their associated faint galaxy accompanying them. Such a feature may indicate that it is a relatively newly merged system.

The one or two galaxies that are prominently brighter than the other members in a group can be related to the “luminosity gap” first described by Ostriker & Tremaine (1975). As dynamical friction is the strongest for the most massive galaxy

in a cluster, the massive galaxies tend to sink into the centers of the halos more quickly, which means that they merge fairly rapidly, before the lower-mass galaxies join them. This results in a galaxy population within the halo that contains one or two very massive galaxies, plus all the lower-mass galaxies. Thus, the most massive galaxy looks like an outlier compared to the rest of the population.

In addition to these case-by-case investigations, we also statistically quantify the distribution of the member galaxies in the groups, since we have a relatively sufficient number of member galaxies. We try to explore the shape of the groups as indicated by the positions of their member galaxies. Theoretically, the environments of the halos are usually divided into



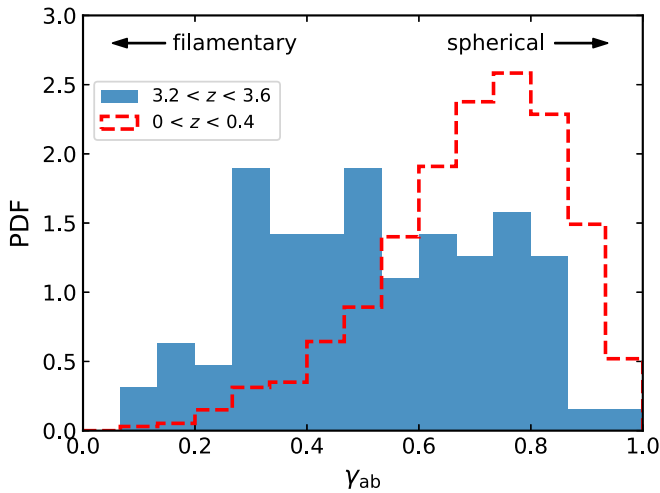
**Figure 9.** Similar to Figure 7, but for groups at  $z \sim 3$ . The red dashed ellipse indicates the  $2\sigma$  coordinate distribution of the member galaxies. The ratio between the minor and the major axes of the red ellipse is written in the right corner of the legend. For clarity, we omit the redshift values indicated in the previous figures.

four classes: voids, filaments, sheets, and clusters (referred as the cosmic web), where the filaments are generally considered to form earlier (e.g., Bond et al. 2010; Cautun et al. 2014). At a given redshift, the massive halos are preferentially located in the cluster environments, while the low-mass halos tend to be located in the filament environments (e.g., Yang et al. 2017). In cases where the shapes of the groups are correlated with their surrounding environments, we would expect the groups with the same halo mass at different redshifts to display different shapes.

Here, we directly use the coordinates of the member galaxies of the groups to quantify their shapes. More explicitly, we take the decl. and R.A. relative to the group center as two

parameters, where R.A. is multiplied by  $\cos(\text{decl.})$  to correct the projection effect at high latitudes. Then, we calculate the  $2\sigma$  scatters of these two parameters ( $\Delta \text{decl.}$ ,  $\Delta \text{R.A.} \cos(\text{decl.})$ ) among the member galaxies, with the assumption of a two-dimensional Gaussian distribution, which can be shown as an ellipse. We adopt the ratio between the minor and major axes of the scatter ellipse,  $\gamma_{ab}$ , to indicate the distribution of the member galaxies or the shape of the groups. In the case of a filamentary distribution,  $\gamma_{ab}$  tends to approach zero. In Figure 9, we show the  $2\sigma$  ellipses with red dashed curves centered on the group center, and mark the  $\gamma_{ab}$  value on the legend. The four cases with  $\gamma_{ab}$  vary from 0.23 to 0.63, representing elongated to relatively spherical shapes of groups at  $z \sim 3$ .





**Figure 10.** The probability density function (PDF) of the coordinate ratio,  $\gamma_{ab}$ . The blue filled and red dashed histograms, respectively, represent the groups at  $3.2 < z < 3.6$  and at  $0 < z < 0.4$  with only the five brightest member galaxies. The groups in the low-redshift bin are selected according to the halo masses of those groups at  $3.2 < z < 3.6$ . We only consider the five brightest member galaxies during the calculations.

Statistically, we show in Figure 10 the number of the  $\gamma_{ab}$  distribution for groups with  $N_g = 5$  members at redshift  $3.2 < z < 3.6$ , and compare them to a reference group sample at  $0 < z < 0.4$  with at least five members. Here, the reference sample is generated by selecting groups at redshift  $0 < z < 0.4$ , according to the halo masses in the primary group sample at redshift  $3.2 < z < 3.6$ . Then, for each reference group, we only keep the five brightest member galaxies for our consideration. The numbers of groups at  $3.2 < z < 3.6$  and  $0 < z < 0.4$  are 95 and 3983, respectively. In the high-redshift sample,  $\gamma_{ab}$  mainly distributes between 0.3 and 0.8, without a clear peak, while the  $\gamma_{ab}$  for the groups at redshift  $0 < z < 0.4$  are concentrated at  $\sim 0.75$ . By comparing the distribution of  $\gamma_{ab}$  in the low- and high-redshift samples, we infer that the groups at early times are prone to form with a filamentary shape, whereas the groups at late times mainly have a more spherical shape. This is consistent with the formation of the LSS under the hierarchical framework that filaments are formed earlier than clusters, where the galaxies are distributed more spherically than those in the filaments.

#### 4.4. Rich Groups at $z \sim 2$

As redshift decreases to  $z \sim 2$ , groups possess more member galaxies due to the increasing cosmic star formation rate (Katsianis et al. 2021) and merger rate (Ventou et al. 2017) at this period. There are 914 groups at redshift  $z \sim 2$  with at least 10 members. These groups have a median halo mass of  $1.8 \times 10^{13} h^{-1} M_\odot$ . The radius of the groups ( $r_{180}$ ) can extend to  $\sim 1 h^{-1} \text{Mpc}$ .

In Figure 11, we show four rich groups at  $z \sim 2$  with member galaxies  $N_g \sim 20$ . Overall, the member galaxies are distributed more isotropically compared with the groups at  $z \sim 3$  (Figure 8). In addition, quite different to those groups at redshift  $z \gtrsim 3$  where there are almost no galaxies at similar redshift close to their halo boundaries (radius), here we start to see some galaxies with similar redshifts approaching the host groups. These coeval galaxies that frequently appear around the boundaries of the groups will significantly contribute to the growth of the groups, i.e., by forming clusters, at a later stage.

The evolution of the major merger rate is popularly studied in simulations and observations, despite there still being debates at high redshifts. The major merger rate investigated in some simulations can increasingly be extended to  $z \gtrsim 3$  (e.g., Stewart et al. 2008; Lagos et al. 2018), whereas a few simulations show a steady profile at  $z \sim 2-3$  (e.g., Kaviraj et al. 2015; Qu et al. 2017; Snyder et al. 2017), which is consistent with the results investigated by means of photometric and flux-ratio-selected galaxy pairs in observations (e.g., Bluck et al. 2009; Man et al. 2012; Duncan et al. 2019; Ventou et al. 2019). As we have obtained a rather uniform flux-limited galaxy group sample at high redshift, we can use the separation between the member galaxies, as well as the stellar mass growth of the member galaxies, to probe the merger rates of the galaxies in observations, and compare them with theoretical model predictions (e.g., Jiang et al. 2008; O’Leary et al. 2021).

#### 4.5. Groups/Clusters at Lower Redshifts

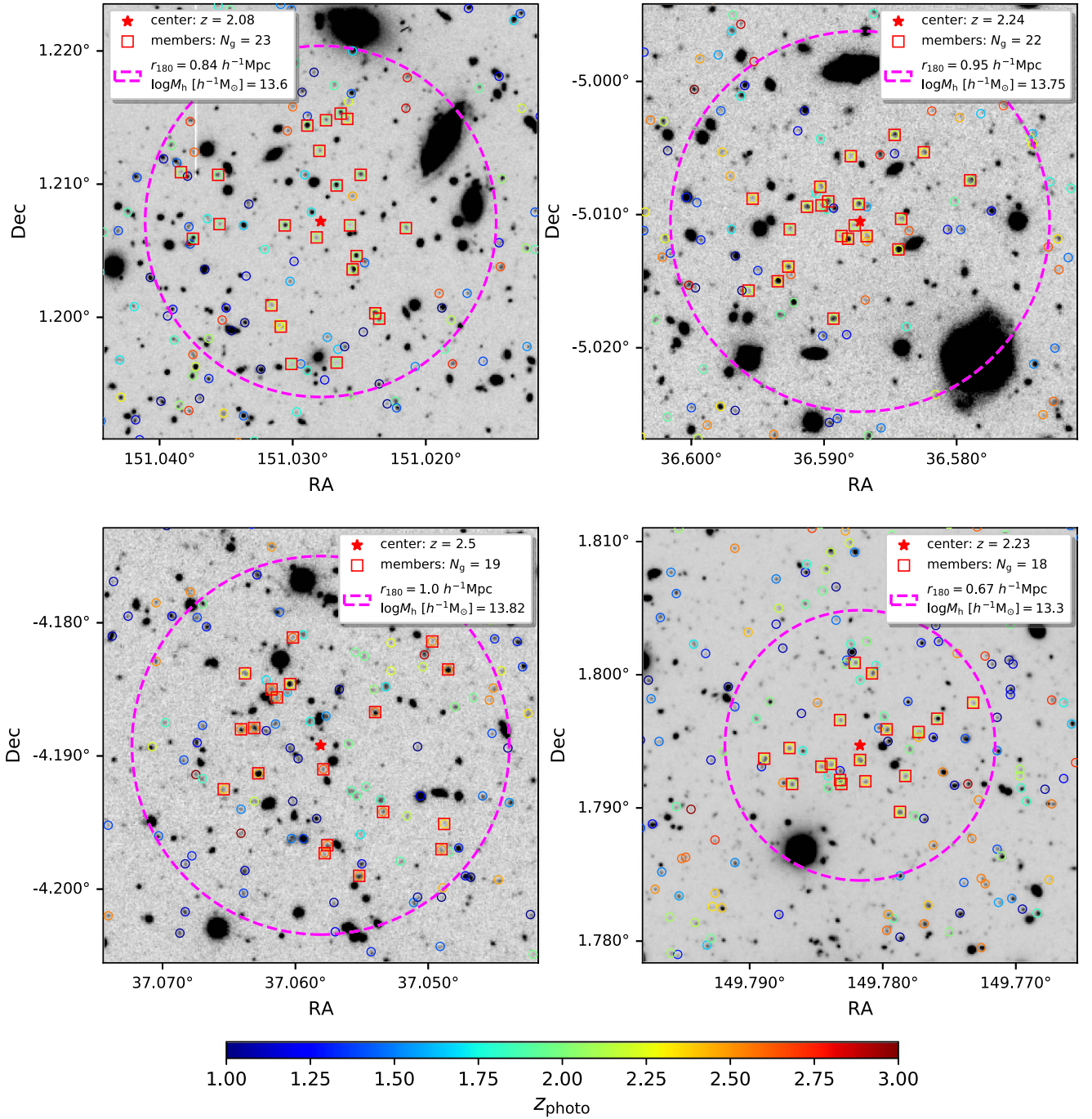
At lower redshifts, there are a total of 68,711 groups with at least 10 members in our group catalog. Here, we do not perform further visual inspections of these individual groups/clusters, but try to evaluate our low-redshift groups by comparing them with other group/cluster data sets. Fruitful numbers of groups/clusters at low redshifts have been found with different methods. Here, we compare our data with two sets of them.

We use our galaxy groups to match the groups determined in Oguri et al. (2017), who applied the CAMIRA (Cluster-finding Algorithm based on Multi-band Identification of Red-sequence galaxies) algorithm (Oguri 2014) to the HSC Wide S16A data set. Here, we consider their version with updated data, which used the HSC-SSP PDR2 with photometric redshifts. In total, there are 197 groups at  $0.1 < z < 1.2$  located in the same fields as ours. As the number density of their groups is roughly above the theoretical curve of  $10^{14} h^{-1} M_\odot$  halos, according to the halo mass function (Oguri et al. 2017), they are mostly clusters with masses larger than  $10^{14} h^{-1} M_\odot$ . Before we proceed to match these clusters with our group/cluster sample, we check whether the BCGs (also determined as the centers) in their cluster sample exist in our galaxy sample. Requiring the photometric difference to be less than 0.2, we have a total number of 140 clusters remaining for our cross-check. Afterward, we search our group catalog around each of their clusters within a projected  $2 h^{-1} \text{Mpc}$  comoving radius and a redshift difference  $\Delta z < 0.2$ . We take the richest group that fulfills these criteria as the matched counterpart. In Figure 12, we show the number distribution of the halo masses in our matched sample. The matched groups are massive, with a median of  $10^{14.08} h^{-1} M_\odot$ , which is close to the induced mass of the groups in Oguri et al. (2017). The lowest group mass in our matched sample is about  $10^{13.0} h^{-1} M_\odot$ , which is still quite massive. Note that here we are using different photometric redshift sources and group detection techniques to the ones used in Oguri et al. (2017), which is the main cause of the differences.

#### 5. Protocluster Candidates

As we outlined in Table 2, there are 914 groups at redshift  $2 \leq z < 3$  with at least 10 members. These groups have a median halo mass of  $1.8 \times 10^{13} h^{-1} M_\odot$ . At redshift  $3 \leq z < 4$ , there are 400 groups with at least five members and a median halo mass of  $6.1 \times 10^{12} h^{-1} M_\odot$ . At redshift  $z \geq 4$ , there are 89





**Figure 11.** Similar to Figure 7, but for groups at  $z \sim 2$ . Note that we only show the galaxies at  $1 \leq z \leq 3$ , marked with the same color bar range.

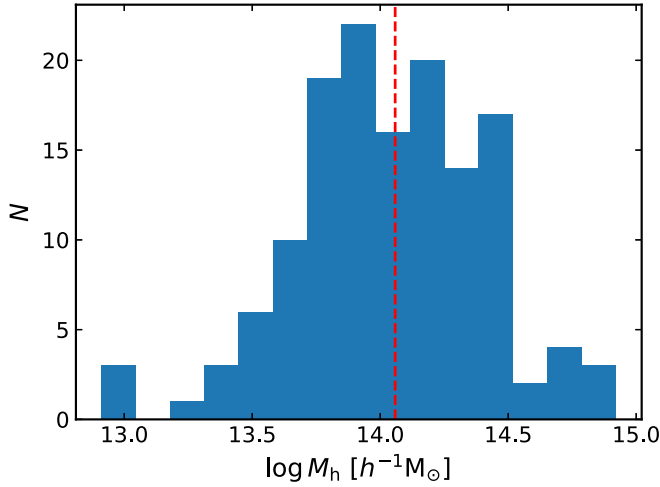
groups with at least three members and a median halo mass of  $3.5 \times 10^{12} h^{-1} M_{\odot}$ . According to the abundance of these groups, and the theoretical median mass growth history of dark matter halos (e.g., Zhao et al. 2009; Yang et al. 2012), these groups should on average be able to grow into clusters with mass  $\gtrsim 10^{14} h^{-1} M_{\odot}$  at redshift  $z = 0$ . Individually, however, they might not all grow into clusters (see the discussion in Cui et al. 2020). In this section, we set out to assess the probability of these group systems growing into clusters. We refer to those with high possibilities as protocluster candidates.

For convenience, we separate the groups used to search for protocluster candidates into three samples, according to their redshifts and number of members: (1)  $2 \leq z < 3$  and

$N_g \geq 10$ ; (2)  $3 \leq z < 4$  and  $N_g \geq 5$ ; and (3)  $z \geq 4$  and  $N_g \geq 3$ , which are referred to as samples S1, S2, and S3, respectively. We exclude groups that suffer significantly from survey edge effects, including the bright star mask and other contaminations (see Section. 3.3). Finally, we obtain 761, 343, and 43 groups in the S1, S2, and S3 samples, respectively.

### 5.1. Assessment Indicators

In general, galaxy clusters in the local universe are embedded in a virialized dark matter halo with a mass greater than  $10^{14} h^{-1} M_{\odot}$ . Estimating the predicated halo mass at  $z = 0$  of a high-redshift overdensity is a common way of judging whether the discovered structure is a protocluster or not (e.g.,

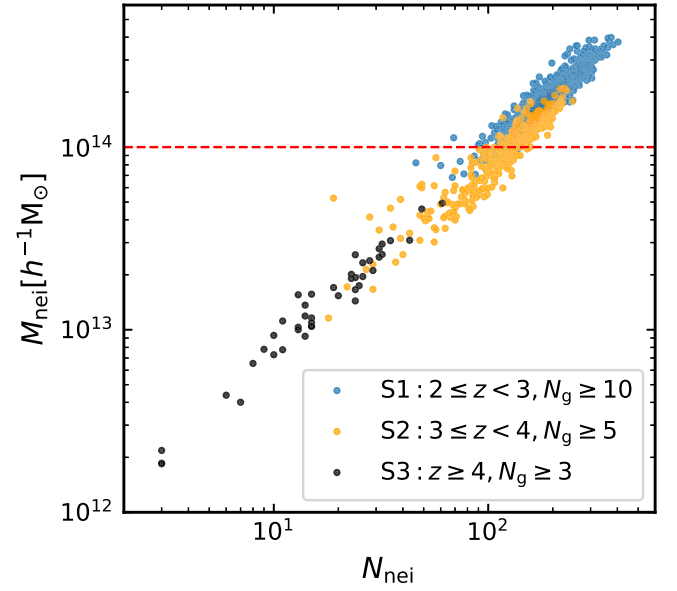


**Figure 12.** Number distribution for the halo mass of our groups matched to those from Oguri et al. (2017). The red line marks the median halo mass,  $\sim 10^{14.08} h^{-1} M_{\odot}$ .

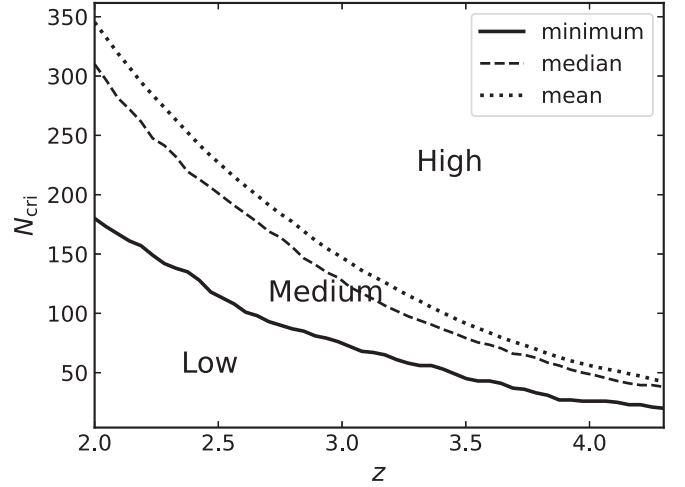
Chiang et al. 2013; Cheema et al. 2020; Polletta et al. 2021). However, the methods used to estimate the  $z=0$  halo mass usually rely on overdensity and accurate redshift (volume) measurements (e.g., Steidel et al. 1998; Cucciati et al. 2014) or on clustering measurements (e.g., Cheema et al. 2020). For our sample, the groups are determined by their photometric redshifts, and their neighboring overdensity may be significantly affected by the photoz quality. Thus, rather than using overdensity, here we introduce a new set of indicators, based on the distribution of neighboring galaxies and groups, to quantify the possibility of being a protocluster candidate.

The galaxies that we link to each group are supposed to be mostly located within the halo’s virial radius ( $r_{180}$ ). Those systems in our S1, S2, and S3 samples can be regarded as the cores of the protoclusters, as also shown in Ando et al. (2022), so we check the available matter (or galaxies) surrounding them. If the mass and/or number of galaxies within and surrounding the group are sufficient to build a redshift  $z=0$  cluster with mass  $\gtrsim 10^{14} h^{-1} M_{\odot}$ , we regard the group as a protocluster candidate. As the forming area for a protocluster is much larger than its already formed group (halo) region at high redshift, we search the galaxy and group distributions around each of the candidate groups within a projected radius and redshift difference  $\Delta z \leq 0.1$ , i.e., the neighboring criteria. The chosen radius criteria for S1, S2, and S3 are  $\sim 5$ , 6, and  $7 h^{-1} \text{Mpc}$ , respectively, which roughly correspond to the effective radii of “Virgo”-type protoclusters at different redshifts, as defined in Chiang et al. (2013; see also Figure 3 in their paper). The chosen redshift boundary roughly corresponds to a  $3\sigma$  scatter of photometric redshift, as shown in Figure 3. For convenience, we call the galaxies and groups around the candidate groups that fulfill the neighboring criteria (within the chosen radius and redshift) “neighboring galaxies and groups” in the following analysis.

The first indicator that we set out to use is the total halo mass of the neighboring groups,  $M_{\text{nei}}$ . Because of the photometric redshift error, as well as the velocity dispersion of the member galaxies, some neighboring groups only contribute a fraction of their member galaxies according to our neighboring criteria, so we calculate  $M_{\text{nei}}$  by taking into account the luminosity fraction of the satisfied member galaxies in their groups. If the  $M_{\text{nei}}$  thus



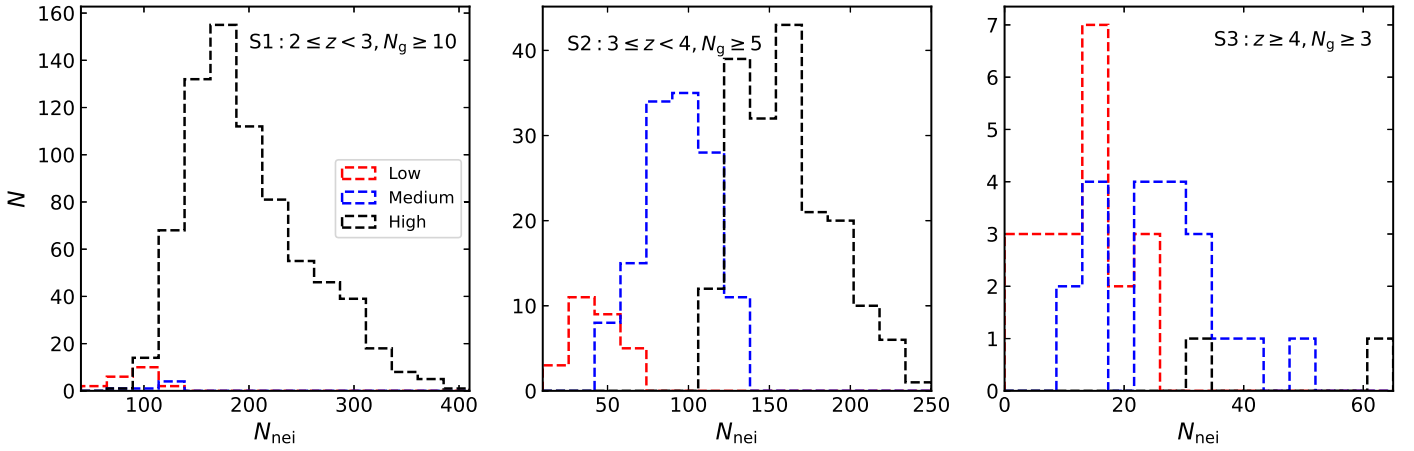
**Figure 13.** The total halo mass of the neighboring groups vs. the total number of neighboring galaxies. The neighboring galaxies or groups are searched for within the projected 5, 6, and  $7 h^{-1} \text{Mpc}$  comoving radii for the S1, S2, and S3 samples, respectively. The groups in these three samples are shown with blue, orange, and black dots, respectively. The red dashed line marks the  $M_{\text{nei}}$  criterion ( $10^{14} h^{-1} M_{\odot}$ ) for a group with neighboring galaxies to be considered as a protocluster candidate.



**Figure 14.** The redshift evolution of the number of neighboring galaxies criteria,  $N_{\text{cri}}$ .  $N_{\text{cri}}$  is calculated based on the number of low-redshift massive clusters. The solid, dashed, and dotted lines, respectively, represent the minimum, median, and mean number of neighboring galaxies criteria. The protocluster candidates below the minimum  $N_{\text{cri}}$ , between the minimum  $N_{\text{cri}}$  and mean  $N_{\text{cri}}$ , and above the mean  $N_{\text{cri}}$ , respectively, are regarded as having low, medium, and high probabilities of growing into clusters, respectively.

calculated is larger than  $10^{14} h^{-1} M_{\odot}$ , we can in general directly regard this group as a protocluster candidate. However, since we are using a flux-limited galaxy sample, we can only detect groups whose member galaxies can pass this limit, which results in a halo mass limit when calculating the  $M_{\text{nei}}$ . It is not straightforward to use  $M_{\text{nei}}$  as a criterion for assessing the probability of a group being a protocluster candidate, especially at very high redshifts.

The next indicator that we use is the total number of neighboring galaxies,  $N_{\text{nei}}$ . Overall, this parameter reflects the available neighboring galaxies that can be accreted to the target



**Figure 15.** The number distribution of the protocloud candidates as a function of the number of neighboring galaxies,  $N_{\text{nei}}$ . The left, middle, and right panels show the protocloud candidates in the S1, S2, and S3 samples, respectively. The red, blue, and black dashed lines represent the protocloud candidates with low, medium, and high probabilities of growing into clusters, respectively.

group, supporting its growth toward a cluster at a later time. Note that here we do not intend to provide the probabilities of being protocloud candidates for all groups, but rather to provide the number of neighboring protocloud candidates in our catalog based on a small selection of the richest groups at different redshift bins. Since the halo masses are estimated from the ranking of the total group luminosity, there should be some intrinsic correlation between  $M_{\text{nei}}$  and  $N_{\text{nei}}$ . As an illustration, we show in Figure 13 the distributions of our rich group systems on an  $M_{\text{nei}}$  versus  $N_{\text{nei}}$  plane. They actually show a strong and continuous linear correlation across different redshift ranges, which indicates that our defined quantity  $N_{\text{nei}}$  can be regarded as a substitution for halo mass estimation. Most groups in S1 can be directly determined as protocloud candidates, according to the values of their  $M_{\text{nei}}$ , whereas none of clusters can be regarded as protocloud candidates when only judged by  $M_{\text{nei}}$  in S3. Nevertheless, even if the  $M_{\text{nei}}$  is smaller than  $10^{14} h^{-1} M_{\odot}$ , it does not necessarily mean that the protocloud candidate cannot form a cluster at a later time, since our calculation of  $M_{\text{nei}}$  suffers from halo mass incompleteness, as these high-redshift halos are still forming with the process of accumulating their mass. In the next subsection, we will provide details of using  $N_{\text{nei}}$  as the assessment indicator by properly taking into account the survey magnitude limit and halo evolution effects.

### 5.2. Finding the Protocloud Candidates

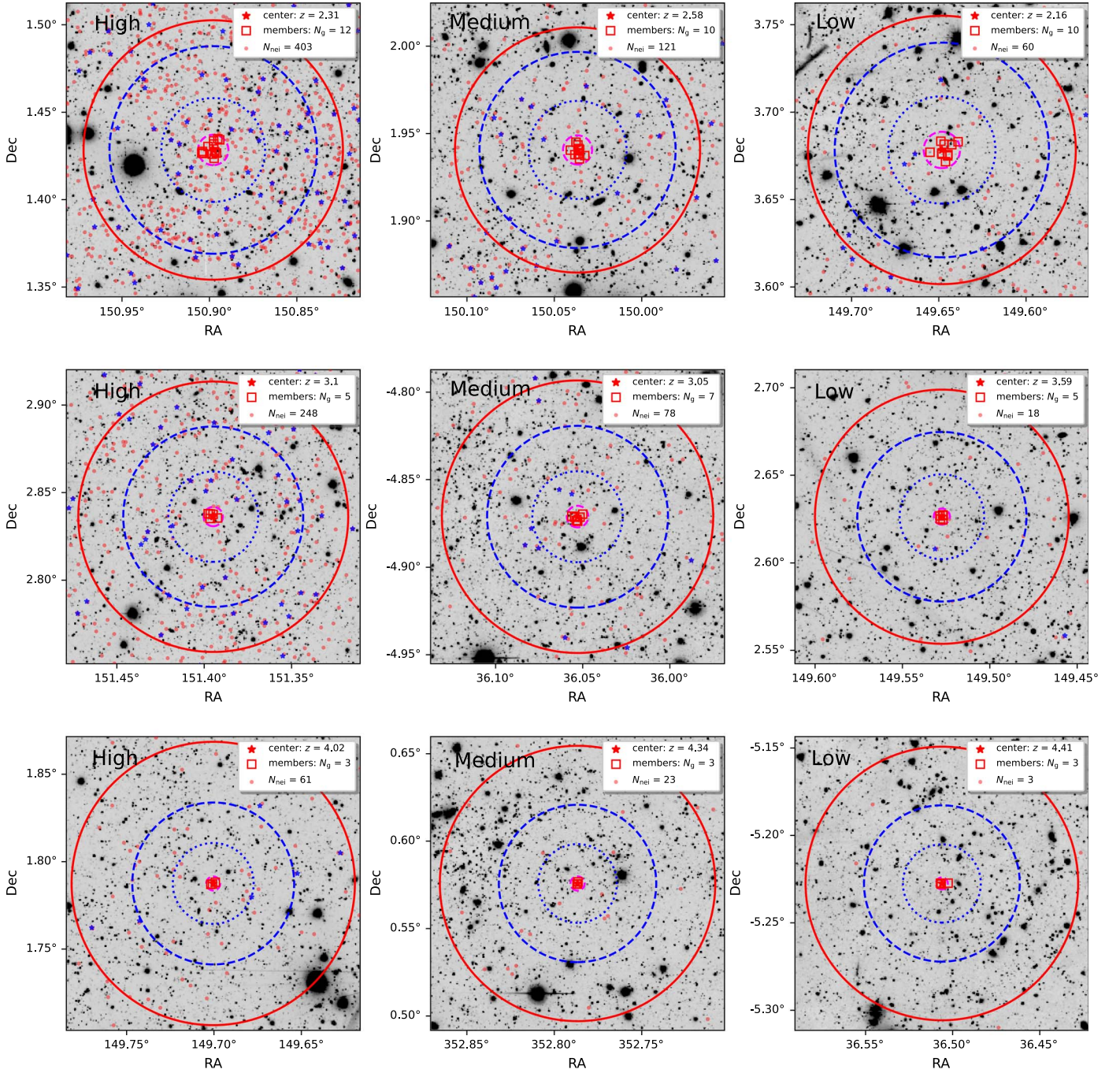
In this subsection, we judge whether the target group is a protocloud or not according to the  $N_{\text{cri}}(z)$  criterion, which can be obtained in the following way. We first select clusters at redshift  $z \leq 0.3$  with mass  $\gtrsim 10^{14} h^{-1} M_{\odot}$ . We obtain 36 clusters after this selection. Next, we select the neighboring galaxies around each of these clusters within a projected maximal distance of their member galaxies and a redshift difference  $\Delta z \leq 0.1$ . We then move these clusters and their neighboring galaxies to higher redshift. According to the magnitude limit that we applied to our galaxy sample, we discard those member galaxies that cannot make the survey magnitude limit. We can thus obtain  $N_{\text{cri}}(z)$  by counting the remaining galaxies. When moving low-redshift clusters to high redshifts, we do not take the galaxy evolution into account, as the luminosity function (the top left panel in Figure 4) shows a weak evolution at  $z \lesssim 3$ . We calculate the  $N_{\text{cri}}$  from each selected

cluster and show the minimum, median, and mean  $N_{\text{cri}}$  as a function of redshift in Figure 14. In general,  $N_{\text{cri}}$  decreases with increasing redshift. A cluster is expected to possess at least 180 neighboring galaxies at  $z \sim 2$  and 26 at  $z \sim 4$ . We quantify the probability of a protocloud candidate growing into a cluster based on the  $N_{\text{nei}}$  value relative to the  $N_{\text{cri}}$  criterion at the corresponding redshifts. We classify this probability of growth into three levels. Specifically, a protocloud candidate with  $N_{\text{nei}}(z) < N_{\text{cri,min}}(z)$  is considered to have a *low* probability of growing into a cluster, while  $N_{\text{cri,min}}(z) \leq N_{\text{nei}}(z) < N_{\text{cri,mean}}(z)$  and  $N_{\text{nei}}(z) \geq N_{\text{cri,mean}}(z)$  are considered to have *medium* and *high* probabilities, respectively. Besides, we also consider protocloud candidates to have a high probability as long as their  $M_{\text{nei}}$  is larger than  $10^{14} h^{-1} M_{\odot}$ .

The number distribution of the protocloud candidates in the S1, S2, and S3 samples, with the division of the three probabilities, is shown in Figure 15. Overall, the protocloud candidates with a large  $N_{\text{nei}}$  have a higher probability of growing into a cluster. For the S1 sample, most of the candidates have a high probability, within which quite a large number of the candidates already possess an  $M_{\text{nei}}$  larger than  $10^{14} h^{-1} M_{\odot}$ . For the S2 sample, the number of protocloud candidates with a medium probability increases compared with those in the S1 sample. At  $z \geq 4$ , only a few protocloud candidates have a high probability. The overlapping area between the different probabilities is a result of the different redshifts of the groups as judged by  $N_{\text{cri}}(z)$ .

As an illustration, we show the cases for the protocloud candidates that have high, medium, and low probabilities of evolving into clusters in the left, middle, and right panels of Figure 16. The panels in the top, central, and bottom rows correspond to the cases at different redshift ranges selected from the S1, S2, and S3 samples, respectively. The neighboring galaxies are marked with red dots. The radius adopted for counting  $N_{\text{nei}}$  is shown with the red circle. Overall, the protocloud candidates with a high probability possess sufficient and dense distributions of neighboring galaxies, whereas the systems with a low probability present sparse galaxy distributions. With the increase in redshift, the total number of neighboring galaxies,  $N_{\text{nei}}$ , decreases, as expected. In addition, the protocloud candidates in a dense environment seem to have more group member galaxies. In summary, the groups in our catalog have a large probability of growing into





**Figure 16.** The cases for the distributions of the neighboring galaxies and groups in the HSC-SSP PDR2 deep  $i$ -band images. The left, central, and right panels show the protocluster candidates with high, medium, and low probabilities of growing into clusters, respectively. The panels in the top, middle, and bottom rows indicate the cases in the S1, S2, and S3 samples, respectively. The neighboring galaxies are marked with red dots, while the members are presented with red squares. The positions of the neighboring groups with redshift differences to the central group of less than 0.1 that have at least two member galaxies are indicated with blue stars. The range of the central group ( $r_{180}$ ) is presented with the dashed magenta line. The chosen radii for counting the  $N_{nei}$  for the protocluster candidates in the top, middle, and bottom panels are, respectively, 5, 6, and 7  $h^{-1}$ Mpc comoving distances, which are shown with the red lines. The radii of the 2 and 4  $h^{-1}$ Mpc comoving distance circles are presented as the dotted and dashed blue lines, respectively. The values of  $N_{nei}$  are labeled in the top right legends.

clusters, especially for the groups at  $2 \leq z < 3$  with at least 10 member galaxies. In the group catalog, we have also provided the related  $N_{nei}$  values for those candidate groups under consideration.

## 6. Summary

We construct a catalog of galaxy groups and protocluster candidates by applying the extended halo-based method of

Yang et al. (2021) to the CLAUDS and HSC-SSP joint deep data set. The extended version of the group finder allows us to deal with a large number of galaxies with both photometric and spectroscopic redshift. In total, we obtain 2,232,134 groups at  $0 < z < 6$ , of which 41,815 groups have at least three member galaxies. We specifically explore the properties of the groups at different redshifts by showing the distributions of the galaxies in the  $i$ -band images from the HSC-SSP PDR2 deep observations. The protocluster candidates are determined based

on rich groups at  $z \geq 2$  and their surrounding galaxies and groups. Our results are summarized as follows.

1. Most groups at  $z \sim 5$  only contain galaxy pairs of similar luminosity. The large separation of these galaxy pairs relative to the halo radius indicates that galaxy or halo major mergers are not frequent at such a high redshift.
2. At  $z \sim 4$ , the majority of most of the massive groups possess three members. Typically, a pair of galaxies are close to each other and have lower luminosity compared with the third member, which implies that the distinction between BCGs and FSGs becomes significant.
3. There are 400 groups at  $z \sim 3$  with at least five member galaxies. The cases of the groups shown have one or two prominent brightest galaxies, which may indicate that these galaxies have become the dominant BCGs by devouring their closest FSGs.
4. We use the ratio between the minor and major axes of the scatter ellipse to investigate the distribution of the member galaxies or the shape of the rich groups at  $z \sim 3$ . The scatter ellipse is calculated using the R.A. and decl. coordinates of the member galaxies. The groups at this early stage are more elongated than those with similar mass at low redshifts. This is consistent with the framework of LSS formation, since galaxies distributed in clusters are more spherical than those in filaments, due to the earlier formations of filaments.
5. The groups at  $z \sim 2$  become richer. There are 914 groups with a median mass of  $1.8 \times 10^{13} h^{-1} M_{\odot}$  possessing at least 10 members. As shown in their images, some galaxies with similar redshifts to the groups frequently appear around the boundaries of the groups. These galaxies are supposed to contribute to the growth of the groups at a later time.
6. Our groups/clusters at lower redshifts are well matched with the group sample from Oguri et al. (2017), which is produced with the HSC-SSP PDR2 data set. Most of the matched groups in our sample have a halo mass larger than  $10^{13.5} h^{-1} M_{\odot}$ .
7. We use the total number of neighboring galaxies,  $N_{\text{nei}}$ , and the total halo mass of the neighboring groups,  $M_{\text{nei}}$ , to find high-redshift protocluster candidates. These candidates are determined from the high-redshift ( $z \geq 2$ ) rich groups. We judge the probability of protocluster candidates growing into clusters based on the  $N_{\text{cri}}(z)$  criterion, which is defined as the number of remaining member galaxies in lower-redshift massive groups at higher redshift according to the magnitude limit. We divide the probabilities into three levels: low, medium, and high, based on the value of  $N_{\text{nei}}$  relative to  $N_{\text{cri}}(z)$ . Most groups at  $2 \leq z < 3$  with at least 10 member galaxies can be directly regarded as protocluster candidates according to their  $M_{\text{nei}}$ .

While our catalog of groups and protocluster candidates has mostly been produced by using photometric redshifts, it is interesting to note that some samples already have a number of spectroscopic and multiwavelength observations. Besides, as the sky area used to study the science theme “Galaxy Evolution” by PFS will completely overlap the CLAUDS and HSC-SSP data set, once PFS starts operating we will continuously update our group catalogs using the massive spectroscopic data as they become available. Our catalog can be combined with the galaxy

observations from PFS to investigate the properties and evolutions of galaxies at high redshifts. Our group and protocluster candidate catalogs can be obtained online<sup>12</sup> or on Zenodo (doi:10.5281/zenodo.6516482).

We sincerely thank the anonymous referee for the useful comments. This work is supported by the National Science Foundation of China (Nos. 11833005, 11890691, 11890692, 11621303, 11890693, 11933003 and 12173025), 111 project No. B20019, and Shanghai Natural Science Foundation grant No. 19ZR1466800. We acknowledge the science research grants from the China Manned Space Project, with Nos. CMS-CSST-2021-A02 and CMS-CSST-2021-A03. Y.S.D. acknowledges the science research grants from the National Key R&D Program of China, via grant No. 2017YFA0402704, NSFC grant 11933003, and the China Manned Space Project, with No. CMS-CSST-2021-A05. W.C. is supported by the STFC AGP Grant ST/V000594/1 and the Atracción de Talento Contract No. 2020-T1/TIC-19882 granted by the Comunidad de Madrid in Spain. He further acknowledges the science research grants from the China Manned Space Project, with Nos. CMS-CSST-2021-A01 and CMS-CSST-2021-B01. J.H. acknowledges the science research grants from NSFC, grant 11973032, the National Key Basic Research and Development Program of China (No. 2018YFA0404504), and the sponsorship from Yangyang Development Fund.

These data were obtained and processed as part of the CFHT Large Area *U*-band Deep Survey (CLAUDS), which is a collaboration between astronomers from Canada, France, and China, described in Sawicki et al. (2019). CLAUDS is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the CFHT, which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. CLAUDS uses data obtained in part through the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories, Chinese Academy of Sciences, and the Special Fund for Astronomy from the Ministry of Finance of China. CLAUDS uses data products from TERAPIX and the Canadian Astronomy Data Centre (CADCC), and was carried out using resources from Compute Canada and the Canadian Advanced Network For Astrophysical Research (CANFAR).

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), the Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

<sup>12</sup> <https://gax.sjtu.edu.cn/data/PFS.html>













This paper makes use of software developed for the Vera C. Rubin Observatory. We thank the Rubin Observatory for making their code available as free software at <http://pipelines.lsst.io/>.

This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of the Center for Computational Astrophysics (CfCA), NAOJ.

This work has made use of the Gravity Supercomputer at the Department of Astronomy, Shanghai Jiao Tong University.

## ORCID iDs

Qingyang Li  <https://orcid.org/0000-0003-0771-1350>  
 Xiaohu Yang  <https://orcid.org/0000-0003-3997-4606>  
 Chengze Liu  <https://orcid.org/0000-0002-4718-3428>  
 Yipeng Jing  <https://orcid.org/0000-0002-4534-3125>  
 Y. Sophia Dai  <https://orcid.org/0000-0002-7928-416X>  
 Marcin Sawicki  <https://orcid.org/0000-0002-7712-7857>  
 Stephen Gwyn  <https://orcid.org/0000-0001-8221-8406>  
 H. J. Mo  <https://orcid.org/0000-0002-9665-5380>  
 Weiguang Cui  <https://orcid.org/0000-0002-2113-4863>  
 Jiaxin Han  <https://orcid.org/0000-0002-8010-6715>

## References

- Abdullah, M. H., Wilson, G., & Klypin, A. 2018, *ApJ*, **861**, 22  
 Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, **70**, S8  
 Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, *PASJ*, **71**, 114  
 Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, *PASJ*, **74**, 247  
 Ando, M., Shimasaku, K., Momose, R., et al. 2022, *MNRAS*, **513**, 3252  
 Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, *MNRAS*, **329**, 355  
 Banerjee, P., Szabo, T., Pierpaoli, E., et al. 2018, *NewA*, **58**, 61  
 Bates, D. J., Tojeiro, R., Newman, J. A., et al. 2019, *MNRAS*, **486**, 3059  
 Berlind, A. A., Frieman, J., Weinberg, D. H., et al. 2006, *ApJS*, **167**, 1  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393  
 Beuther, H., Leurini, S., Schilke, P., et al. 2007, *A&A*, **466**, 1065  
 Blanton, M. R., & Roweis, S. 2007, *AJ*, **133**, 734  
 Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, *ApJ*, **592**, 819  
 Bleem, L. E., Bocquet, S., Stalder, B., et al. 2020, *ApJS*, **247**, 25  
 Bluck, A. F. L., Conselice, C. J., Bouwens, R. J., et al. 2009, *MNRAS*, **394**, L51  
 Bond, N. A., Strauss, M. A., & Cen, R. 2010, *MNRAS*, **409**, 156  
 Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, *PASJ*, **70**, S5  
 Bradshaw, E. J., Almaini, O., Hartley, W. G., et al. 2013, *MNRAS*, **433**, 194  
 Cai, Z., Fan, X., Bian, F., et al. 2017, *ApJ*, **839**, 131  
 Cai, Z., Cantalupo, S., Prochaska, J. X., et al. 2019, *ApJS*, **245**, 23  
 Cai, Z., Fan, X., Peirani, S., et al. 2016, *ApJ*, **833**, 135  
 Campbell, D., van den Bosch, F. C., Hearin, A., et al. 2015, *MNRAS*, **452**, 444  
 Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, *Natur*, **470**, 233  
 Carlberg, R. G., Yee, H. K. C., Morris, S. L., et al. 1999, *RSPTA*, **357**, 167  
 Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, *MNRAS*, **441**, 2923  
 Cheema, G. K., Sawicki, M., Arcila-Osejo, L., et al. 2020, *MNRAS*, **494**, 804  
 Cheng, T., Clements, D. L., Greenslade, J., et al. 2019, *MNRAS*, **490**, 3840  
 Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, *ApJ*, **779**, 127  
 Chiang, Y.-K., Overzier, R. A., Gebhardt, K., et al. 2015, *ApJ*, **808**, 37  
 Coil, A. L., Gerke, B. F., Newman, J. A., et al. 2006, *ApJ*, **638**, 668  
 Collister, A. A., & Lahav, O. 2005, *MNRAS*, **361**, 415  
 Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, **575**, A40  
 Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, *AJ*, **110**, 2655  
 Contini, E., De Lucia, G., Hatch, N., Borgani, S., & Kang, X. 2015, *MNRAS*, **456**, 1924  
 Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E., & Kurk, J. D. 2014, *MNRAS*, **440**, 3262  
 Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, *MNRAS*, **449**, 1352  
 Crook, A. C., Huchra, J. P., Martimbeau, N., et al. 2007, *ApJ*, **655**, 790  
 Cucciati, O., Zamorani, G., Lemaux, B. C., et al. 2014, *A&A*, **570**, A16  
 Cui, W., Borgani, S., Dolag, K., Murante, G., & Tornatore, L. 2012, *MNRAS*, **423**, 2279  
 Cui, W., Qiao, J., Davé, R., et al. 2020, *MNRAS*, **497**, 5220  
 Cui, W., Dave, R., Knebe, A., et al. 2022, *MNRAS*, **514**, 977  
 Darvish, B., Scoville, N. Z., Martin, C., et al. 2020, *ApJ*, **892**, 8  
 De Lucia, G., & Blaizot, J. 2007, *MNRAS*, **375**, 2  
 Diaz-Giménez, E., & Zandivarez, A. 2015, *A&A*, **578**, A61  
 Diener, C., Lilly, S. J., Ledoux, C., et al. 2015, *ApJ*, **802**, 31  
 Driver, S. P., Robotham, A. S. G., Kelvin, L., et al. 2012, *MNRAS*, **427**, 3244  
 Duarte, M., & Mamon, G. A. 2015, *MNRAS*, **453**, 3849  
 Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, *ApJ*, **876**, 110  
 Einasto, J., Einasto, M., Tago, E., et al. 2007, *A&A*, **462**, 811  
 Eke, V. R., Baugh, C. M., Cole, S., et al. 2004, *MNRAS*, **348**, 866  
 Galametz, A., Stern, D., De Breuck, C., et al. 2012, *ApJ*, **749**, 169  
 Gerke, B. F., Newman, J. A., Davis, M., et al. 2005, *ApJ*, **625**, 6  
 Golob, A., Sawicki, M., Goulding, A. D., & Coupon, J. 2021, *MNRAS*, **503**, 4136  
 Goto, T. 2005, *MNRAS*, **359**, 1415  
 Hamana, T., Shirasaki, M., & Lin, Y.-T. 2020, *PASJ*, **72**, 78  
 Harikane, Y., Ono, Y., Ouchi, M., et al. 2022, *ApJS*, **259**, 20  
 Hasselfield, M., Hilton, M., Marriage, T. A., et al. 2013, *JCAP*, **2013**, 008  
 Hill, D. T., Driver, S. P., Cameron, E., et al. 2010, *MNRAS*, **404**, 1215  
 Hilton, M., Sifón, C., Naess, S., et al. 2021, *ApJS*, **253**, 3  
 Hu, W., Wang, J., Infante, L., et al. 2021, *NatAs*, **5**, 485  
 Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841  
 Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, *ApJ*, **675**, 1095  
 Jiang, L., Wu, J., Bian, F., et al. 2018, *NatAs*, **2**, 962  
 Jing, Y. P., Mo, H. J., & Börner, G. 1998, *ApJ*, **494**, 1  
 Katsianis, A., Yang, X., & Zheng, X. 2021, *ApJ*, **919**, 88  
 Katsianis, A., Blanc, G., Lagos, C. P., et al. 2017, *MNRAS*, **472**, 919  
 Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, *MNRAS*, **264**, 201  
 Kaviraj, S., Devriendt, J., Dubois, Y., et al. 2015, *MNRAS*, **452**, 2845  
 Koester, B. P., McKay, T. A., Annis, J., et al. 2007, *ApJ*, **660**, 239  
 Koyama, Y., Polletta, M. d. C., Tanaka, I., et al. 2021, *MNRAS*, **503**, L1  
 Kravtsov, A. V., & Borgani, S. 2012, *ARA&A*, **50**, 353  
 Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, **218**, 15  
 Lagos, C. d. P., Stevens, A. R. H., Bower, R. G., et al. 2018, *MNRAS*, **473**, 4956  
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24  
 Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, *A&A*, **559**, A14  
 Lemaux, B. C., Le Fèvre, O., Cucciati, O., et al. 2018, *A&A*, **615**, A77  
 Lilly, S. J., Fevre, O. L., Renzini, A., et al. 2007, *ApJS*, **172**, 70  
 Lim, S. H., Mo, H. J., Lu, Y., Wang, H., & Yang, X. 2017, *MNRAS*, **470**, 2982  
 Lu, Y., Yang, X., & Shen, S. 2015, *ApJ*, **804**, 55  
 Lu, Y., Yang, X., Shi, F., et al. 2016, *ApJ*, **832**, 39  
 Man, A. W. S., Toft, S., Zirm, A. W., Wuyts, S., & van der Wel, A. 2012, *ApJ*, **744**, 85  
 Mandelbaum, R., Miyatake, H., Hamana, T., et al. 2018, *PASJ*, **70**, S25  
 Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, *ApJ*, **841**, 111  
 McLure, R. J., Pearce, H. J., Dunlop, J. S., et al. 2013, *MNRAS*, **428**, 1088  
 Mehmood, R., Zhang, G., Bie, R., Dawood, H., & Ahmad, H. 2016, *Neurocomputing*, **208**, 210  
 Merchán, M., & Zandivarez, A. 2002, *MNRAS*, **335**, 216  
 Merchán, M. E., & Zandivarez, A. 2005, *ApJ*, **630**, 759  
 Miley, G. K., Overzier, R. A., Tsvetanov, Z. I., et al. 2004, *Natur*, **427**, 47  
 Miller, C. J., Nichol, R. C., Reichart, D., et al. 2005, *AJ*, **130**, 968  
 Miller, T. B., Chapman, S. C., Aravena, M., et al. 2018, *Natur*, **556**, 469  
 Miyazaki, S., Oguri, M., Hamana, T., et al. 2018, *PASJ*, **70**, S27  
 Mobasher, B., Rowan-Robinson, M., Georgakakis, A., & Eaton, N. 1996, *MNRAS*, **282**, L7  
 Moutard, T., Sawicki, M., Arnouts, S., et al. 2020, *MNRAS*, **494**, 1894  
 Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2015, *MNRAS*, **452**, 2528  
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493  
 Negrello, M., Gonzalez-Nuevo, J., De Zotti, G., et al. 2017, *MNRAS*, **470**, 2253  
 Oguri, M. 2014, *MNRAS*, **444**, 147  
 Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2017, *PASJ*, **70**, S20  
 Oguri, M., Miyazaki, S., Li, X., et al. 2021, *PASJ*, **73**, 817  
 O’Leary, J. A., Moster, B. P., Naab, T., & Somerville, R. S. 2021, *MNRAS*, **501**, 3215  
 Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, *PASJ*, **70**, S10  
 Ostriker, J. P., & Tremaine, S. D. 1975, *ApJL*, **202**, L113  
 Oteo, I., Ivison, R. J., Dunne, L., et al. 2018, *ApJ*, **856**, 72  
 Overzier, R. A. 2016, *A&ARv*, **24**, 14  
 Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, **318**, 1144



- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, [A&A](#), **596**, A100
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, [A&A](#), **641**, A6
- Polletta, M., Soucail, G., Dole, H., et al. 2021, [A&A](#), **654**, A121
- Qu, Y., Helly, J. C., Bower, R. G., et al. 2017, [MNRAS](#), **464**, 1659
- Robotham, A., Wallace, C., Phillipps, S., & De Propriis, R. 2006, [ApJ](#), **652**, 1077
- Rodriguez, F., & Merchán, M. 2020, [A&A](#), **636**, A61
- Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, [ApJS](#), **224**, 1
- Sawicki, M., Arcila-Osejo, L., Golob, A., et al. 2020, [MNRAS](#), **494**, 1366
- Sawicki, M., Arnouts, S., Huang, J., et al. 2019, [MNRAS](#), **489**, 5202
- Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, [AJ](#), **113**, 1
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, [MNRAS](#), **446**, 521
- Scodreggio, M., Guzzo, L., Garilli, B., et al. 2018, [A&A](#), **609**, A84
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, [MNRAS](#), **323**, 1
- Shi, D. D., Cai, Z., Fan, X., et al. 2021, [ApJ](#), **915**, 32
- Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, [ApJS](#), **220**, 12
- Snyder, G. F., Lotz, J. M., Rodriguez-Gomez, V., et al. 2017, [MNRAS](#), **468**, 207
- Springel, V. 2005, [MNRAS](#), **364**, 1105
- Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, [ApJ](#), **492**, 428
- Stewart, K. R., Bullock, J. S., Wechsler, R. H., Maller, A. H., & Zentner, A. R. 2008, [ApJ](#), **683**, 597
- Tago, E., Einasto, J., Saar, E., et al. 2006, [AN](#), **327**, 365
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, [PASJ](#), **66**, R1
- Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, [A&A](#), **600**, A110
- Tempel, E., Tuvikene, T., Kipper, R., & Libeskind, N. I. 2017, [A&A](#), **602**, A100
- Tinker, J. L. 2021, [ApJ](#), **923**, 154
- Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, [PASJ](#), **70**, S12
- van den Bosch, F. C., Norberg, P., Mo, H. J., & Yang, X. 2004, [MNRAS](#), **352**, 1302
- van den Bosch, F. C., Weinmann, S. M., Yang, X., et al. 2005, [MNRAS](#), **361**, 1203
- Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2007, [A&A](#), **461**, 823
- Ventou, E., Contini, T., Bouché, N., et al. 2017, [A&A](#), **608**, A9
- Ventou, E., Contini, T., Bouché, N., et al. 2019, [A&A](#), **631**, A87
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, [MNRAS](#), **444**, 1518
- Wang, H., Mo, H., Chen, S., et al. 2018, [ApJ](#), **852**, 31
- Wang, K., Mo, H. J., Li, C., Meng, J., & Chen, Y. 2020, [MNRAS](#), **499**, 89
- Wang, T., Elbaz, D., Daddi, E., et al. 2016, [ApJ](#), **828**, 56
- Wechsler, R. H., & Tinker, J. L. 2018, [ARA&A](#), **56**, 435
- Weinmann, S. M., van den Bosch, F. C., Yang, X., & Mo, H. J. 2006, [MNRAS](#), **366**, 2
- Wylezalek, D., Galametz, A., Stern, D., et al. 2013, [ApJ](#), **769**, 79
- Yang, X., Mo, H. J., Jing, Y. P., & van den Bosch, F. C. 2005a, [MNRAS](#), **358**, 217
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, [MNRAS](#), **339**, 1057
- Yang, X., Mo, H. J., & van den Bosch, F. C. 2006, [ApJL](#), **638**, L55
- Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005b, [MNRAS](#), **356**, 1293
- Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005c, [MNRAS](#), **357**, 608
- Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2005d, [MNRAS](#), **362**, 711
- Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2007, [ApJ](#), **671**, 153
- Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., & Han, J. 2012, [ApJ](#), **752**, 41
- Yang, X., Zhang, Y., Lu, T., et al. 2017, [ApJ](#), **848**, 60
- Yang, X., Xu, H., He, M., et al. 2021, [ApJ](#), **909**, 143
- Zandivarez, A., Martínez, H. J., & Merchán, M. E. 2006, [ApJ](#), **650**, 137
- Zhao, D. H., Jing, Y. P., Mo, H. J., & Börner, G. 2009, [ApJ](#), **707**, 354