

University of Groningen

## Tracing the rise of supermassive black holes A panchromatic search for faint, unobscured quasars at $z < 6$ with COSMOS-Web and other surveys

Andika, Irham T.; Jahnke, Knud; Onoue, Masafusa; Silverman, John D.; Fitriana, Itsna K.; Bongiorno, Angela; Brinch, Malte; Casey, Caitlin M.; Faisst, Andreas; Gillman, Steven

*Published in:*  
Astronomy and Astrophysics

*DOI:*  
[10.1051/0004-6361/202349025](https://doi.org/10.1051/0004-6361/202349025)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2024

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Andika, I. T., Jahnke, K., Onoue, M., Silverman, J. D., Fitriana, I. K., Bongiorno, A., Brinch, M., Casey, C. M., Faisst, A., Gillman, S., Gozaliasl, G., Hayward, C. C., Hirschmann, M., Kocevski, D., Koekemoer, A. M., Kokorev, V., Lambrides, E., Lee, M. M., Rich, R. M., ... Vijayan, A. P. (2024). Tracing the rise of supermassive black holes A panchromatic search for faint, unobscured quasars at  $z < 6$  with COSMOS-Web and other surveys. *Astronomy and Astrophysics*, 685, Article A25. <https://doi.org/10.1051/0004-6361/202349025>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.













### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Tracing the rise of supermassive black holes

## A panchromatic search for faint, unobscured quasars at $z \gtrsim 6$ with COSMOS-Web and other surveys<sup>★</sup>

Irham T. Andika<sup>1,2</sup> , Knud Jahnke<sup>3</sup> , Masafusa Onoue<sup>4,5</sup>, John D. Silverman<sup>4,6</sup>, Itsna K. Fitriana<sup>7</sup> , Angela Bongiorno<sup>8</sup>, Malte Brinch<sup>9,10</sup> , Caitlin M. Casey<sup>11,9</sup>, Andreas Faisst<sup>12</sup> , Steven Gillman<sup>9,10</sup> , Ghassem Gozaliasl<sup>13,14</sup> , Christopher C. Hayward<sup>15</sup> , Michaela Hirschmann<sup>16,17</sup>, Dale Kocevski<sup>18</sup>, Anton M. Koekemoer<sup>19</sup> , Vasily Kokorev<sup>20</sup>, Erini Lambrides<sup>21</sup>, Minju M. Lee<sup>9,10</sup> , Robert Michael Rich<sup>22</sup>, Benny Trakhtenbrot<sup>23</sup> , Claudia Megan Urry<sup>24</sup>, Stephen M. Wilkins<sup>25,26</sup> , and Aswin P. Vijayan<sup>9,10</sup>

(Affiliations can be found after the references)

Received 19 December 2023 / Accepted 25 January 2024

### ABSTRACT

We report the identification of 64 new candidates of compact galaxies, potentially hosting faint quasars with bolometric luminosities of  $L_{\text{bol}} = 10^{43} - 10^{46}$  erg  $\text{s}^{-1}$ , residing in the reionization epoch within the redshift range of  $6 \lesssim z \lesssim 8$ . These candidates were selected by harnessing the rich multiband datasets provided by the emerging JWST-driven extragalactic surveys, focusing on COSMOS-Web, as well as JADES, UNCOVER, CEERS, and PRIMER. Our search strategy includes two stages: applying stringent photometric cuts to catalog-level data and detailed spectral energy distribution fitting. These techniques effectively isolate the quasar candidates while mitigating contamination from low-redshift interlopers, such as brown dwarfs and nearby galaxies. The selected candidates indicate physical traits compatible with low-luminosity active galactic nuclei, likely hosting  $\approx 10^5 - 10^7 M_{\odot}$  supermassive black holes (SMBHs) living in galaxies with stellar masses of  $\approx 10^8 - 10^{10} M_{\odot}$ . The SMBHs selected in this study, on average, exhibit an elevated mass compared to their hosts, with the mass ratio distribution slightly higher than those of galaxies in the local Universe. As with other high- $z$  studies, this is at least in part due to the selection method for these quasars. An extensive Monte Carlo analysis provides compelling evidence that heavy black hole seeds from the direct collapse scenario appear to be the preferred pathway to mature this specific subset of SMBHs by  $z \approx 7$ . Notably, most of the selected candidates might have emerged from seeds with masses of  $\sim 10^5 M_{\odot}$ , assuming a thin disk accretion with an average Eddington ratio of  $f_{\text{Edd}} = 0.6 \pm 0.3$  and a radiative efficiency of  $\epsilon = 0.2 \pm 0.1$ . This work underscores the significance of further spectroscopic observations, as the quasar candidates presented here offer exceptional opportunities to delve into the nature of the earliest galaxies and SMBHs that formed during cosmic infancy.

**Key words.** methods: data analysis – methods: observational – galaxies: active – galaxies: high-redshift – quasars: general – quasars: supermassive black holes

### 1. Introduction

Powered by gas and dust accretion onto supermassive black holes (SMBHs), quasars are among the brightest entities in the Universe with the corresponding active galactic nucleus (AGN) bolometric luminosities reaching  $L_{\text{bol}} \gtrsim 10^{46}$  erg  $\text{s}^{-1}$ . Thanks to various wide-field sky surveys, to date, more than 200 quasars hosting  $\gtrsim 10^9 M_{\odot}$  black holes have been discovered at  $z \gtrsim 6$ , with a select number of them already shining brightly when the cosmos was just less than 800 Myr old (see, e.g., Fan et al. 2023, for a recent review). Assuming that such SMBHs originate from less massive seeds (i.e.,  $\approx 10^2 - 10^6 M_{\odot}$ ), assembling those enormous amounts of mass is challenging, requiring highly efficient matter accretions with additions of black hole mergers (Woods et al. 2019; Pacucci & Loeb 2020). Hence, these high- $z$  quasars, with their extreme characteristics compared to inactive galaxies, are ideal targets for examining the assembly of the earliest galaxies and SMBHs during cosmic infancy (Pacucci & Loeb 2022).

Several studies have proposed explanations for constructing the black hole seeds, although the comprehensive solution to this

problem is still open-ended. These theories include the idea that the first generation of low-mass black holes are presumably produced at the same time when the first-generation stars (hereafter Population III stars) are populating the Universe at  $z \sim 20 - 30$ , or around 200 Myr since the Big Bang (Volonteri et al. 2021). In line with that, black hole seeds are often separated into two classes, depending on their initial mass: (i) heavy seeds with a mass range of  $10^4 - 10^6 M_{\odot}$  and (ii) light seeds with masses of  $10 - 100 M_{\odot}$  (see, e.g., Inayoshi et al. 2020, and references therein).

One challenge of growing light seeds to form  $10^9 M_{\odot}$  SMBHs by  $z \approx 7$  is there is simply not enough time unless episodes of super- or even hyper-Eddington accretion can be sustained (e.g., Middleton et al. 2013; Madau et al. 2014; Dubois et al. 2014; Valiante et al. 2016; Pezzulli et al. 2016; Pacucci et al. 2017; Natarajan 2021). While the super-Eddington accretion rate is just slightly above the Eddington-limit rate but still around the same order of magnitude, hyper-Eddington events can have values that are hundreds of times higher owing to photon trapping mechanisms reducing the radiation pressure effect on the infalling matter (Begelman & Volonteri 2017). However, since most of the quasars discovered today are observed as having instantaneous accretion rates below or around the Eddington limit

<sup>★</sup> FITS files and full Table B.1 are available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/685/A25>

(Trakhtenbrot et al. 2017; Fragione & Pacucci 2023), the theory on heavy seeds is thus being explored further to ease the time-limited SMBH growth issue and possibly jump-start the formation of high- $z$  quasars (e.g., Yoo & Miralda-Escudé 2004; Volonteri 2010; Mayer & Bonoli 2019). As the first possibility, heavy seeds could form by collapsing primeval gas residing in the atomic-cooling halo, potentially producing short-lived supermassive stars (or quasi-stars) as by-products with a mass range of  $10^5$ – $10^6 M_{\odot}$  (Bromm & Loeb 2003; Lodato & Natarajan 2006; Hosokawa et al. 2013; Smith & Bromm 2019). The second possibility of heavy seed formation is that runaway collisions and mergers of either black holes or Population III stars within a gas-dense environment – namely, a dense star cluster – could produce seeds with masses of  $10^3$ – $10^4 M_{\odot}$  (Alexander & Natarajan 2014; Lupi et al. 2016; Boekholt et al. 2018; Latif et al. 2021; Massonnet et al. 2023; Trinca et al. 2023). Heavy seeds might reduce the discrepancy between the theoretical model of SMBH growth and the observed quasar properties. However, such objects have yet to be detected (Nabizadeh et al. 2024; Natarajan et al. 2024).

Discovering more quasars in the reionization era is one obvious pathway for understanding early SMBH formation. In particular, finding less massive black holes ( $\approx 10^6$ – $10^8 M_{\odot}$ ) at higher redshifts might give more information on whether heavy seeds are the dominant channel to explain the majority of the  $z \gtrsim 7$  quasar population. Only the most luminous quasars, and hence, the largest, rarest SMBHs, could be discovered before the launch of the *James Webb* Space Telescope (JWST; Mortlock et al. 2011; Bañados et al. 2019; Matsuoka et al. 2019; Venemans et al. 2020; Wang et al. 2021; Yang et al. 2021; Izumi et al. 2021; Andika et al. 2022). Today, JWST is allowing for high- $z$  lower-luminosity AGNs ( $L_{\text{bol}} \approx 10^{43}$ – $10^{45}$  erg s $^{-1}$ ) to be hunted where the stellar light might dominate the total emission or where the central emission from the accretion process is obscured (e.g., Labbe et al. 2023; Maiolino et al. 2023, 2024; Larson et al. 2023a; Fujimoto et al. 2023a; Goulding et al. 2023; Furtak et al. 2023; Kokorev et al. 2023, 2024; Greene et al. 2024; Williams et al. 2023a; Pérez-González et al. 2024). About 30 lower mass SMBHs have been reported so far, and these objects might be the missing connection between the earliest bright quasars and black hole seeds.

Given the necessity of understanding how the first SMBHs and galaxies evolve, we present 64 new compact sources, potentially harboring quasars with  $L_{\text{bol}} \lesssim 10^{46}$  erg s $^{-1}$  and  $\lesssim 10^8 M_{\odot}$  SMBHs at  $6 \lesssim z \lesssim 8$ , selected utilizing various ground- and space-based imaging data. Specifically, we exploit publicly available archival datasets covering the COSMOS, GOODS-S/N, Abell 2744, EGS HST legacy, and PRIMER extragalactic fields. If spectroscopically confirmed, our candidates will double the number of quasars in the mass, luminosity, and redshift ranges mentioned earlier. Furthermore, combining our samples with other published quasars in the literature will allow us to perform statistical analysis on this intriguing population and check their black hole and host galaxy characteristics.

The outline of this paper is as follows. We start with the details on data acquisition and main database construction in Sect. 2. Then, the method for identifying quasar candidates via photometric and spectral energy distribution (SED) modeling will be presented in Sect. 3. After that, we deliver the results and discuss the properties of the new candidates in Sect. 4. Finally, we end this paper with a summary and conclusions in Sect. 5. For simplification and ease of reference within this paper, we subsequently define “quasar” as an interchangeable term for quasi-stellar object (QSO) and active galactic nucleus (AGN). On sev-

eral occasions, low-luminosity AGNs with  $L_{\text{bol}} \lesssim 10^{46}$  erg s $^{-1}$ , whose emission could be overwhelmed by the host galaxy’s light but the AGN contribution is still detectable are also considered as quasars. The magnitudes written in this paper are reported using the AB system. We further adopt the flat  $\Lambda$ CDM cosmological framework, where we assume  $\Omega_{\Lambda} = 0.7$ ,  $\Omega_{\text{m}} = 0.3$ , and  $H_0 = 70$  km s $^{-1}$  Mpc $^{-1}$ . Consequently, at  $z = 7$ , the Universe’s age is 0.748 Gyr, and the angular scale of  $\theta = 1''$  corresponds to a linear scale of 5.3 kpc.

## 2. Multi-survey datasets

This section outlines the multiband photometric datasets used for the high- $z$  quasar selection in several major JWST extragalactic fields: COSMOS, GOODS-S/N, Abell 2744, EGS HST legacy, and PRIMER. Some details on each of these surveys, data processing, and catalog construction will also be discussed here. The unified database is then utilized to perform preselection and SED modeling to find promising candidates.

### 2.1. The COSMOS-Web survey

The first dataset is based on the COSMOS-Web program (GO #1727, PIs Kartaltepe & Casey), a deep imaging program covering 0.54 deg $^2$  with 255 h total integration time. COSMOS-Web uses four JWST/NIRCam bands ( $F115W$ ,  $F150W$ ,  $F277W$ , and  $F444W$ ) and one MIRI filter ( $F770W$ ) in parallel. More details on the survey description and observing strategy are presented by Casey et al. (2023). Our work utilizes the first two epochs of COSMOS-Web data obtained in January and April 2023. The current available NIRCam mosaics cover approximately 0.28 deg $^2$ ; on the other hand, MIRI data contains 0.07 deg $^2$  of the COSMOS-Web field.

Data reduction for the NIRCam images is carried out utilizing the standard JWST Calibration Pipeline (Bushouse et al. 2022). In addition to that, custom processing steps are implemented to improve the image quality. This includes 1/f noise and low-level background subtraction (e.g., Bagley et al. 2024) and astrometric correction bootstrapped from the *Hubble* Space Telescope (HST) imaging in the  $F814W$  filter (Koekemoer et al. 2007) and the COSMOS2020 catalogs (Weaver et al. 2022), anchored to the *Gaia*-EDR3 data (Gaia Collaboration 2023). The resulting multi-band image mosaics with 0'':03/pixel have an astrometric normalized median absolute deviation below 12 mas. Accordingly, MIRI data are reduced using a similar process to produce 0'':06/pixel mosaics. While we only give a short overview here, two forthcoming papers will discuss details of the reduction process (Franco et al., in prep.; Harish et al., in prep.).

We complement the JWST data with multiwavelength information from various surveys performed on the Cosmic Evolution Survey (COSMOS) field. This includes photometric datasets from HST/ $F814W$  (Scoville et al. 2007; Koekemoer et al. 2007), *Spitzer*/IRAC (Euclid Collaboration 2022), Subaru/HSC PDR3 (Aihara et al. 2022), and UltraVISTA DR5 (McCracken et al. 2012). A detailed summary of how these data are compiled and reprocessed is provided by Weaver et al. (2022). Furthermore, we add submillimeter measurements from the A3COSMOS catalog (Liu et al. 2019) when available.

The COSMOS-Web photometric catalog is produced using the SourceXtractor++ code (SE++; Bertin et al. 2020, 2022). To create a detection image for reference, we first stack all four NIRCam bands via a chi-square ( $\chi^2$ ) combination (Szalay et al. 1999). Flux measurements are then performed on each band using model-based photometry, including the ancillary data from



HST and other ground-based observations. We note that model-based photometry enables flux extraction on images with diverse point spread functions (PSFs) without degrading their quality (see also Weaver et al. 2023). Specifically, this approach allows us to include constraints from ground-based data without sacrificing space-based data’s resolution and, consequently, photometric accuracy. In total, 342 435 sources are obtained from this catalog.

It should be noted that the flux errors of faint or undetected targets are often underestimated due to the flexibility given to the SE++ catalog construction. To handle this issue, we set a noise floor in each band equivalent to the shot noise calculated using circular apertures placed randomly with sizes of 0’3 and 1’ for space-based and ground-based data, respectively. Furthermore, to compute the source detection’s significance, parameterized with the signal-to-noise ratio (S/N), we also consider the flux-to-error ratio extracted using an aperture of 1’5 diameter. This measurement is more robust than the model-based photometry S/N, and the aperture size is large enough to capture the whole source light, given the different PSF sizes between image filters. The details on the photometric catalog creation will be described in a separate work (Shuntov et al., in prep.).

## 2.2. The JADES project

Multiband data of the Great Observatories Origins Deep Survey South (GOODS-S) sky field is taken from the first public release of the JWST Advanced Deep Extragalactic Survey (JADES<sup>1</sup>; Eisenstein et al. 2023) observations. This dataset covers the “deep” portion of the images with exposure time per filter of 3.9–16.7 h obtained in September/October 2022, resulting in a sky area of 25 arcmin<sup>2</sup> with a nominal 5 $\sigma$  depth of around 29.9 mag. The JWST/NIRCam filters utilized by the JADES project include *F090W*, *F115W*, *F150W*, *F200W*, *F277W*, *F335W*, *F356W*, *F410M*, *F444W* – that is, spanning the wavelengths of 0.8–5.0  $\mu$ m. Photometry for 47 181 unique targets is provided in the catalog, where the source extractions and measurements are explained in detail by Hainline et al. (2024).

The JADES catalog also makes use of the JWST Extragalactic Medium-band Survey (JEMS; Williams et al. 2023b) data, adding *F182M*, *F210M*, *F430M*, *F460M*, and *F480W* filters. Moreover, observations from the First Reionization Epoch Spectroscopic COmplete survey (FRESCO; Oesch et al. 2023) are also included, complementing the JADES catalog with *F182M*, *F210M*, and *F444W* filters when available. As for the bluer wavebands, JADES utilized the existing deep HST/ACS and WFC3 mosaics from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) and the Hubble Legacy Field dataset (Whitaker et al. 2019) containing *F435W*, *F606W*, *F775W*, *F814W*, and *F850LP* images. Finally, it is worth mentioning that all fluxes we use for the SED fitting later are based on the ones measured within a circular aperture with a radius of 0’15, corrected for flux losses, in the “CIRC\_CONV” table of the JADES catalog.

## 2.3. The UNCOVER program

The search on the lensing cluster Abel 2744 region will be conducted using the data provided by the Ultradeep NIRSpec and NIRCam Observations before the Epoch of Reionization (UNCOVER<sup>2</sup>; Bezanson et al. 2022) Cycle 1 JWST Treasury

<sup>1</sup> <https://jades-survey.github.io>

<sup>2</sup> <https://jwst-uncover.github.io>

program. The second version of the photometric catalog released by this program is constructed based on the 49 arcmin<sup>2</sup> image mosaics of seven NIRCam filters – that is, *F115W*, *F150W*, *F200W*, *F277W*, *F356W*, *F410M*, and *F444W* – together with numerous HST/ACS and WFC3 ancillary data.

For photometry purposes, all UNCOVER mosaics are PSF-matched to the *F444W* band, and the detection images for source extractions are created by combining the *F277W*, *F356W*, and *F444W* mosaics exploiting the noise-equalized technique. Specifically, the so-called “SUPER” catalog that we will use here, where the photometry is calculated from optimally selected color apertures in the range of 0’32 – 1’4 diameter on 0.04’’/pixel mosaics, reaches a nominal 5 $\sigma$  magnitude limit of around 30 mag. We note that the fluxes in that catalog are corrected to total values using the Kron Radius measured in the detection image, with an additional correction of approximately 5–10% applied to account for missing light beyond a 1’’ radius, guided by the *F444W* curve of growth (Weaver et al. 2024). By default, fluxes for 61 648 unique sources in the UNCOVER catalog are reported in the unit of 10 nJy or correspond to the AB magnitude zero point of 28.9. This catalog is further enriched with submillimeter measurements from the Deep UNCOVER-ALMA Legacy High-Z (DUALZ) Survey, featuring ALMA band 6 observations with a 30-GHz wide frequency band down to a sensitivity of 32.7  $\mu$ Jy beam<sup>-1</sup> (Fujimoto et al. 2023b).

## 2.4. Additional archival data

In addition to the previously mentioned datasets, we also make use of the JWST data targeting some public extragalactic fields, which were processed with grizli (Brammer et al. 2022) and msaexp (Brammer 2023) by the Cosmic Dawn Center (DAWN), stored in the DAWN JWST Archive (DJA<sup>3</sup>; Valentino et al. 2023). Specifically, we first mined the Cosmic Evolution Early Release Science Survey (CEERS<sup>4</sup>; Finkelstein et al. 2023) data provided by DJA to expand our candidates list. We refer the reader to Bagley et al. (2023) for a complete description of the official CEERS data products. In short, the dataset consists of NIRCam imaging in *F115W*, *F150W*, *F200W*, *F277W*, *F356W*, *F410M*, and *F444W* bands targeting the Extended Groth Strip (EGS) HST legacy field with the current area coverage of 91 arcmin<sup>2</sup> and a 5 $\sigma$  depth of 28.3–28.8 mag.

Along with that, we also exploited the DJA’s version of the Public Release IMaging for Extragalactic Research (PRIMER<sup>5</sup>; Dunlop et al. 2021) dataset. The PRIMER survey was performed utilizing the NIRCam and MIRI imaging on two contiguous equatorial regions, namely, the Ultra-Deep Survey (UDS; Lawrence et al. 2007) and COSMOS (Scoville et al. 2007) fields. While the used filters are similar to CEERS, PRIMER further enriches the covered wavelengths by adding *F090W*, *F770W*, and *F1800W* bands. In total, the areas covered by the PRIMER-UDS and PRIMER-COSMOS reach about 212 arcmin<sup>2</sup> and 164 arcmin<sup>2</sup>, respectively, with a 5 $\sigma$  limiting magnitude of 27.4–27.9 mag. As further information, complementary to the CEERS and PRIMER’s JWST data, DJA also provides photometric measurements based on the existing HST archival images (i.e., Grogin et al. 2011; Koekemoer et al. 2011; Kokorev et al. 2022).

At the time of writing, the photometric catalog of the GOODS-N field, along with some GOODS-S regions from the

<sup>3</sup> <https://dawn-cph.github.io/dja>

<sup>4</sup> <https://ceers.github.io>

<sup>5</sup> <https://primer-jwst.github.io>

**Table 1.** Overview of the employed selection that we used to detect the high- $z$  quasars.

Step	Selection	COSMOS-Web	JADES/ GOODS-S	GOODS-N	UNCOVER	CEERS	PRIMER- COSMOS	PRIMER- UDS
1	All sources	342 435	70 899	37 890	61 648	76 300	118 794	143 552
2	SED modeling	247	383	61	105	237	172	185
3	Visual inspection	30	58	16	32	54	69	91
4	$M_{\text{BH}} > 10^5 M_{\odot}$ and $f_{\text{AGN}} \geq 0.2$	18	11	6	3	6	13	7
•	Sky coverage (arcmin <sup>2</sup> )	1008	57	55	49	91	164	212
•	Faintest magnitude	26.1	29.4	27.4	28.1	27.7	28.0	27.8

**Notes.** At the end of our search, we found 350 compact sources, including 64 showing attributes consistent with low-luminosity AGNs. We also report the sky area covered by each dataset and the faintest  $F444W$  magnitude of the candidates.

“non-deep” portion of the JADES programs, has yet to be released by the official JADES collaboration (Eisenstein et al. 2023). Fortunately, a subset of their NIRCcam mosaics is publicly available and processed by DJA, covering the area of about 55 arcmin<sup>2</sup> and 57 arcmin<sup>2</sup> for the northern and southern datasets, respectively. These images include additional data from the observing programs of JADES Medium, 1210/1286 Parallel, and northwest and southeast pointings. Similar to the dataset introduced in Sect. 2.2, DJA complemented the JADES GOODS-S/N data with the NIR imaging from the JEMS (Williams et al. 2023b) and FRESCO (Oesch et al. 2023) projects, as well as the optical photometry from the Hubble Legacy Fields program (Whitaker et al. 2019). With all data in hand, we ultimately consider the aperture-based photometry, corrected for flux losses, calculated with a diameter of 0.36 for CEERS, PRIMER-UDS, PRIMER-COSMOS, GOODS-N, GOODS-S catalogs produced by DJA, each containing 76 300, 143 552, 118 794, 37 890, and 52 427 objects, respectively. It is important to note that the GOODS-S dataset constructed here and the one obtained in Sect. 2.2 are then merged to remove duplicated sources by crossmatching these two catalogs using a 1'' radius. This combined catalog is hereafter called “JADES” to differentiate them from the GOODS-N data.

At this point, we then compile the catalogs from the COSMOS-Web, JADES GOODS-S/N, UNCOVER, CEERS, and PRIMER projects. All fluxes are converted to nJy unit, corresponding to AB zero point of 31.4 mag. To further ensure that bright flux values do not excessively influence the SED fitting process later and to accommodate potential uncertainties in photometric calibration, we designate a lower limit of 5% as the error floor for the photometric measurements (e.g., Hainline et al. 2024). The effect of Galactic extinction is then corrected using the dust map of Schlegel et al. (1998) and reddening correction of Fitzpatrick (1999), applied using the software from Green (2018).

This resulting parent catalog comprises 851 518 unique sources (see Table 1 for the breakdown), which includes a mix of galaxies and quasars at all redshifts, stars, substellar objects, and artifacts. The following sections will describe various steps to extract the actual quasar content and resulting AGN properties, with the steps already listed in Table 1. First, the SEDs of all objects are modeled with composite SED templates representative of galaxies with and without AGN, as well as stars and substellar objects, including dust reddening. This SED modeling will robustly remove all non-galaxies from the catalog, low-redshift galaxies, and AGN with a photometric  $z < 5.5$ . The resulting much smaller sample of high- $z$  candidates is then visually inspected to remove objects with SEDs impacted by cosmic ray

hits, hot pixels, stray light residuals, etc. This approach will provide a high-probability set of high- $z$  candidate galaxies and AGN we already discussed. Then, in the final step, detailed independent SED fitting is used to extract relative galaxy and AGN flux contributions in these high-probability candidates. We demonstrate the robustness and limits of these estimates and then use the resulting AGN flux to infer AGN properties for the sample.

### 3. Quasar search via SED fitting

#### 3.1. Photometric redshift estimation and initial selection

We implement the first SED modeling step – to find the quasar candidates and separate them from other contaminants, such as low- $z$  galaxies, brown dwarfs, detector artifacts, etc. (e.g., Andika et al. 2020, 2023a) – using eazy-py<sup>6</sup>, a Python-based photometric redshift estimator (Brammer et al. 2008). By iterating through a user-defined grid of spectral templates and redshifts, eazy-py tries to find the best model that matches the observed photometry.

Here, the templates for quasar SEDs are derived empirically from the observational data of XMM-COSMOS AGNs and galaxies, provided and discussed in detail by Ananna et al. (2017). Although the original template list includes a wide variety of galaxy types, we exclusively use the spectra of bright quasars showing broad emission lines and a blue rest-frame ultraviolet (UV) continuum for our purposes (e.g., Andika et al. 2023b). As done by Duncan et al. (2021), we further append the effect of dust extinction using attenuation levels ( $A_V$ ) ranging from 0 to 2 with a step of 0.2, following the model from Calzetti et al. (2000).

Also, the built-in templates for inactive galaxies provided by eazy-py are constructed based on the Flexible Stellar Population Synthesis code (FSPS; Conroy et al. 2009, 2010; Conroy & Gunn 2010) and one high-equivalent-width galaxy from Erb et al. (2010). These SEDs contain a mixture of stellar, nebular, and dust-reprocessed emission components. It is important to note that, since young, high- $z$  galaxies could show very blue UV continuum slopes due to their high star formation rate (SFR), lower metallicity, and less dusty nature, we put to use additional bluer templates from Larson et al. (2023b) complementing the available SED models. That is, we use the “reduced Ly $\alpha$ ” sets in Larson et al. (2023b), optimized to fit galaxies at  $4 \leq z \leq 7$ .

The set of main sequence stellar SEDs is taken directly from the PHOENIX stellar library, encompassing a wide range of spectral types, luminosities, and effective temperatures

<sup>6</sup> <https://github.com/gbrammer/eazy-py>

(Husser et al. 2013). In line with that, brown dwarf spectra are obtained from the Sonora models, covering diverse properties of self-luminous extrasolar planets along with type L, T, and Y brown dwarfs (Marley et al. 2021). After compiling the required SED models, we define a redshift grid of  $0.05 \leq z \leq 20$  with  $\Delta z = 0.05$  step and distribute the quasar and galaxy templates accordingly. Furthermore, we create one additional grid for the galaxy template, forcing the redshifts to be  $z \leq 5.5$  to ensure that our candidates are distinct from the low- $z$  sources. After that, depending on the  $z$  values, intergalactic medium (IGM) attenuation is applied following the analytical equation proposed by Inoue et al. (2014). In contrast, we set the redshifts to be close to zero for the star and brown dwarf models.

Each quasar candidate will be modeled with four classes of templates: quasar, galaxy, star, and low- $z$  source. The likelihood of the source being a quasar is subsequently determined by comparing its  $\chi^2$  divided by the number of bands employed in the SED fitting (hereafter  $\chi_n^2$ ). To be exact, we define the goodness-of-fit for quasar, galaxy, star, and low- $z$  source model as  $\chi_{n,q}^2$ ,  $\chi_{n,g}^2$ ,  $\chi_{n,s}^2$ , and  $\chi_{n,lz}^2$ , respectively, and compare their values. Examples of the resulting SED fit and the image cutouts are displayed in Fig. 1. A preliminary quasar selection is then performed utilizing the criteria as follows:

1. Detections in four NIRCcam bands ( $F115W$ ,  $F150W$ ,  $F277W$ , and  $F444W$ ) with more than  $5\sigma$ . These bands cover the region redward of the expected Ly $\alpha$  emission at  $z \gtrsim 6$ .
2.  $S/N < 3$  in the optical bands blueward of the anticipated Ly $\alpha$  break. Precisely, we use both the HST/ACS  $F435W$  and  $F606W$  bands for the JADES, UNCOVER, CEERS, and PRIMER datasets, while Subaru/HSC  $g$  and  $r$  filters are utilized for the COSMOS-Web sources.
3. The best-fit model for the observed SED is not a star but either a galaxy or a quasar with the inferred  $\chi_{n,g}^2$  and  $\chi_{n,q}^2$  values being  $< 10$ . Here, we do not require the candidates to be best fitted by a pure quasar model since their host galaxy emission could dominate the observed SEDs in the lower luminosity regimes, as found in most of JWST-confirmed,  $z \gtrsim 5$  faint AGNs to date (see, for example, Harikane et al. 2023; Maiolino et al. 2023; Greene et al. 2024).
4. The source is located at high- $z$ , indicated by the estimated photometric redshift being  $z_{\text{phot}} > 6$ , both for the galaxy and quasar models.
5. The integrated redshift probability at  $z > 5.5$  should be more than 90%, that is,  $P(z > 5.5) > 0.9$ .

Of all sources identified in the combined catalogs, 1370 targets pass our initial criteria and are then visually inspected. More than half of these candidates are cosmic rays, hot pixels, stray light residues, contaminated by nearby bright sources, moving objects, or other detector artifacts. During the visual inspection stage, we also discard candidates with extended morphologies, as they could be low- $z$  dusty sources not visible in the ground-based and HST imaging. Sources with compact shapes with circularized diameter less than  $0''.5$  are preferred because their light is likely dominated by a centralized emission component around the galactic nucleus.

To identify sources that have already been spectroscopically confirmed and published in the literature, we cross-match our candidates to the DJA's JWST sources repository<sup>7</sup> and the SIMBAD Astronomical Database<sup>8</sup> (Wenger et al. 2000). Correspondingly, the current datasets that we have contain at least 36 confirmed broad-line AGNs at  $4 \lesssim z \lesssim 9$  reported in the literature, for

which 11 of them are located at  $z \gtrsim 6$  (Harikane et al. 2023; Maiolino et al. 2023; Larson et al. 2023a; Kocevski et al. 2023; Übler et al. 2023; Kokorev et al. 2023; Stone et al. 2023). Our  $z_{\text{phot}}$  estimates for these sources agree with the redshifts derived via spectroscopy considering the estimated uncertainties, indicating a good performance of our SED fitting with eazy-py. As mentioned before, most of these sources ( $\approx 80\%$ ) prefer best-fit SEDs based on the galaxy spectral templates, given the substantial brightness of their host galaxy emission, while the remaining objects opt for the pure quasar models. Also, these confirmed AGNs often show compact shapes consistent with our selection criteria. After discarding spurious sources and already published high- $z$  galaxies in the literature, our final selection yields 350 remaining quasar/galaxy candidates (see Table 1).

It is noteworthy to mention that the redshift calculated via broadband photometry can exhibit a systematic deviation from the one based on spectroscopy, which we refer to as a systematic offset bias (e.g., Carrasco Kind & Brunner 2013; Nishizawa et al. 2020). This bias is quantified as  $\Delta z = (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ . For a subset of 27 856 confirmed AGN/galaxies with available spectroscopic data from the DJA's JWST sources repository, on which we applied our SED modeling, we find that the average bias is  $\langle \Delta z \rangle = -0.05$  while its standard deviation is  $\sigma = 0.46$ . The outlier fraction, defined as the fraction of sources with  $|\Delta z| > 0.15$ , is 23%. When we focus on the subset that meets the high- $z$  source criteria outlined in the preceding paragraphs, the corresponding statistics shift to  $\langle \Delta z \rangle = 0.01$ ,  $\sigma = 0.03$ , and an outlier fraction of 6%. While there is a noticeable scatter in the accuracy of  $z_{\text{phot}}$ , these results are already sufficient to distinguish between low- and high- $z$  sources, with contamination rates of roughly 5%–25% (see Fig. 2).

### 3.2. Measurements of the galaxy properties

After robustly identifying a sample of high- $z$  galaxy and AGN candidates, which should have few interlopers or spurious members, we carry out complementary SED modeling to extract galaxy and AGN parameters from the broad-band SEDs and will also robustly estimate AGN contributions in this sample. We treat this high-confidence candidate sample as a sample of actual high- $z$  galaxies with variations in AGN contribution between 0% and 100%. We discuss the validity of this approach in the following sections.

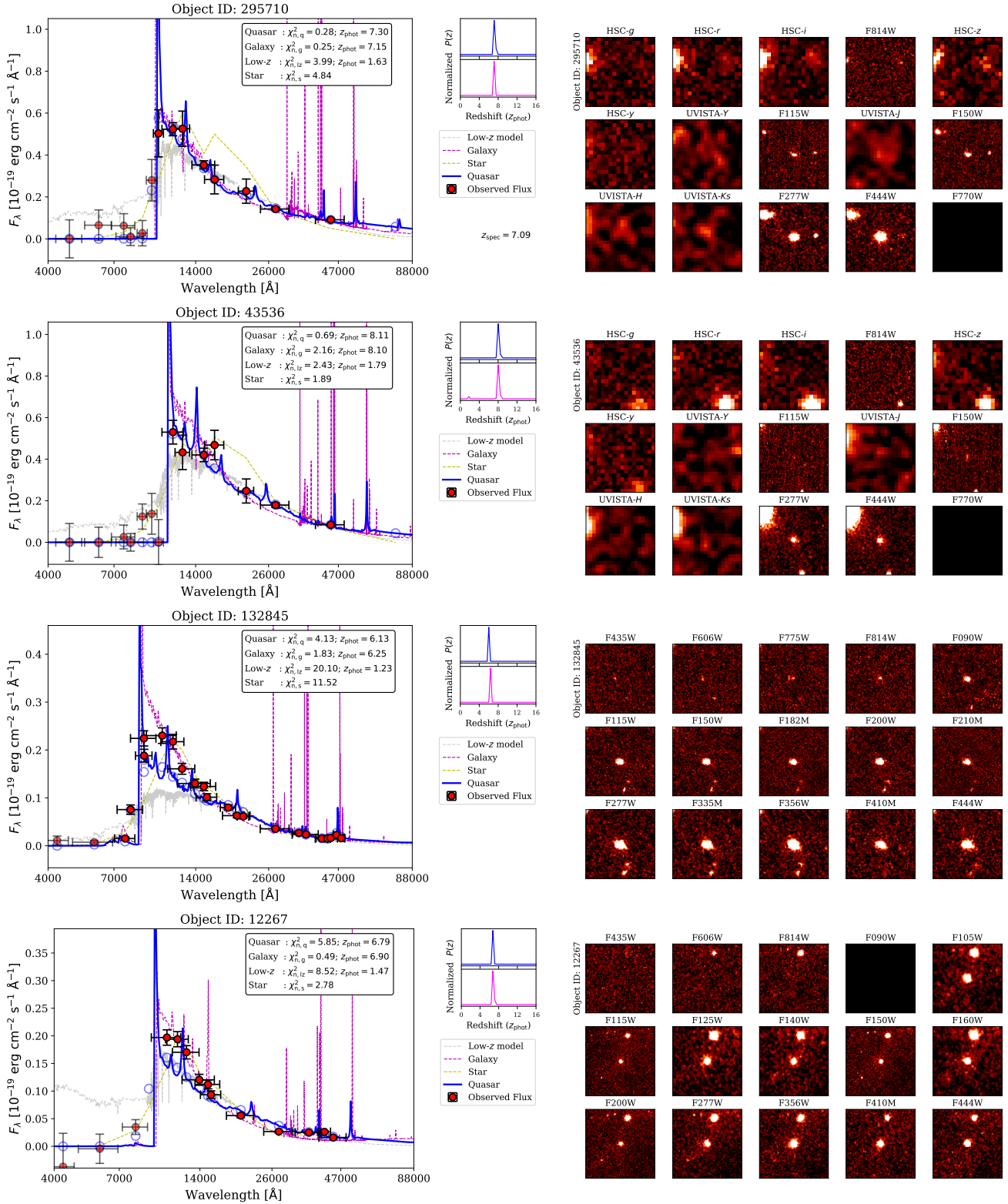
We model the SEDs using the Code Investigating GALaxy Emission (CIGALE; Boquien et al. 2019; Yang et al. 2020, 2022) package. Following the default configuration as a reference, we consider a delayed star formation history (SFH) with an e-folding time range of  $0.1 \leq \tau \leq 5$  Gyr and a recent burst, assuming Bruzual & Charlot (2003) stellar population models along with a Chabrier (2003) initial mass function (IMF) and stellar metallicity of  $Z = 0.02$ . Next, nebular emission is approximated using the Inoue (2011) model, while the dust extinction is added utilizing the combined Calzetti et al. (2000) and Leitherer et al. (2002) attenuation laws, dubbed as the modified starburst module in the CIGALE setup. We set the  $E(B - V)$  color excess for both nebular lines and stellar continuum to be between 0.05 and 2.65, which is equivalent to dust attenuation levels of  $A_V \approx 0.2$ –8.2, assuming a ratio of total-to-selective extinction of  $R_V = A_V/E(B - V) = 3.1$ . This wide range of attenuation levels is chosen since  $z \lesssim 5$  dust-enshrouded star-forming galaxies could appear as if they were sources at extremely high redshifts (Zavala et al. 2023; Meyer et al. 2024).

The ionization parameter, gas metallicity, and electron density of nebular lines are fixed to  $\log U = -2$ ,  $Z_{\text{gas}} = 0.02$ , and

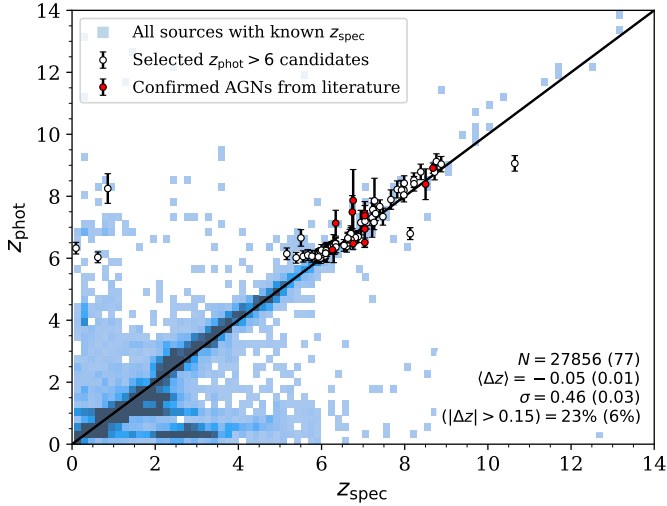
<sup>7</sup> <https://dawn-cph.github.io/dja/general/jwst-sources>

<sup>8</sup> <https://simbad.cds.unistra.fr/simbad>





**Fig. 1.** Photometric SEDs of some quasar candidates. In the left part of each panel, we model the observed photometric data points with four types of spectral templates. Fluxes with  $S/N > 3$  are marked with red circles, while those with lower values are shown with a bit transparent color. The best-fit quasar template and its associated synthetic photometry are presented with blue lines and circles. Models based on the galaxy and star/brown dwarf spectra are shown with magenta and yellow lines, with an additional fitted model of low- $z$  sources displayed in gray color. The goodness-of-fit of each model and the estimated redshift are also reported (see main text for a detailed definition). We further indicate the calculated redshift probability distribution function,  $P(z)$  for quasar and galaxy models. Finally, the right part of each panel shows the multiband images of the  $z \geq 6$  quasar candidate, each trimmed to  $6''$  size on a side.



**Fig. 2.** Comparison between  $z_{\text{phot}}$  and  $z_{\text{spec}}$ . The number count ( $N$ ), average bias ( $\langle \Delta z \rangle$ ), scatter ( $\sigma$ ), and outlier fraction ( $|\Delta z| > 0.15$ ) of all sources (blue squares) with available spectroscopic data are reported. The region with darker colors corresponds to a higher number of sources within the 2D histogram bins. We also show the metrics for a subset that satisfies our high- $z$  selection criteria (white circles with error bars). Samples of spectroscopically confirmed AGNs from the literature are depicted with red circles.

$n_e = 100 \text{ cm}^{-3}$ , respectively. We acknowledge that opting for this choice could introduce an additional uncertainty of up to 5% on the inferred AGN-to-host galaxy flux ratio, along with  $\sim 0.1$  dex in the measurements of stellar mass. However, its impact on the accuracy of photometric redshift estimations is observed to be minimal. We also note that  $U$ ,  $Z_{\text{gas}}$ , and  $n_e$  display higher sensitivity to altering the emission line strengths and lower sensitivity to modifying the continuum shape, indicating broadband photometry data alone, as we used here, would not be enough to constrain them well (Kaasinen et al. 2017; Kewley et al. 2019). Given the considerations, the introduced tradeoff is acceptable for achieving a simpler model with significantly faster computation times.

Since we are also interested in assessing how much the AGN emission contributes to the observed total fluxes, we make use of the Skirtor2016 model provided by CIGALE on top of the previous SED sets (Stalevski et al. 2012, 2016). This 3D radiative transfer AGN model includes the accretion disk emission on the UV/optical side and the torus plus polar dust emission at infrared (IR) wavelengths. In addition, the adopted AGN inclination angle could affect the resulting AGN class, namely, obscured or unobscured. Accordingly, we set this as a free parameter to cover both types. Following that, the IGM attenuation effect is appended as a function of redshift following the formula from Meiksin (2006). Finally, the SED models consisting of the galaxy and AGN components are fitted within redshift bins of  $0.05 \leq z \leq 16$  using a step size of  $\Delta z = 0.05$ . The CIGALE input file will be provided as supplementary data with this paper for reader reference and accessibility. We refer to the CIGALE documentation<sup>9</sup> for detailed information on all the spectral templates adopted here (Boquien et al. 2019; Yang et al. 2022).

Examples of the best-fit SED model made with CIGALE are portrayed in Fig. 3. Correspondingly, the current SED modeling yields posterior distributions of some physical parameters (see Fig. 4), such as the AGN fraction of the total emission ( $f_{\text{AGN}}$ ), host galaxy stellar mass ( $M_*$ ), and SFR averaged over 100 Myr,

along with  $z_{\text{phot}}$  and  $A_V$ . It should be emphasized that  $f_{\text{AGN}}$  is calculated considering only the rest-frame wavelengths from 0.1 to 0.7  $\mu\text{m}$ , which is the region constrained by our ground and space-based data while excluding ALMA submillimeter measurements. Furthermore, this wavelength range covers essential broad emission lines in the quasar SED, such as Ly $\alpha$ , H $\beta$ , and H $\alpha$ . More details on the generated CIGALE output parameters are discussed in Boquien et al. (2019). Discussion on the inferred physical characteristics of our quasar candidates will be showcased in the next section.

## 4. Results and discussion

### 4.1. List of quasar candidates and their number density

Up to this stage, we have selected 350 candidates of high- $z$  compact sources via our initial photometric cut, visual inspection, and advanced SED modeling with two independent codes. There will be unresolvable mismatches between observed SEDs and template inputs used for both modeling methods. Hence, there is space for nominal AGN components formally compensating for such template mismatch, even for fully nuclear-passive galaxies. For the subsequent analysis, we will use a threshold in formal AGN fraction to mitigate this.

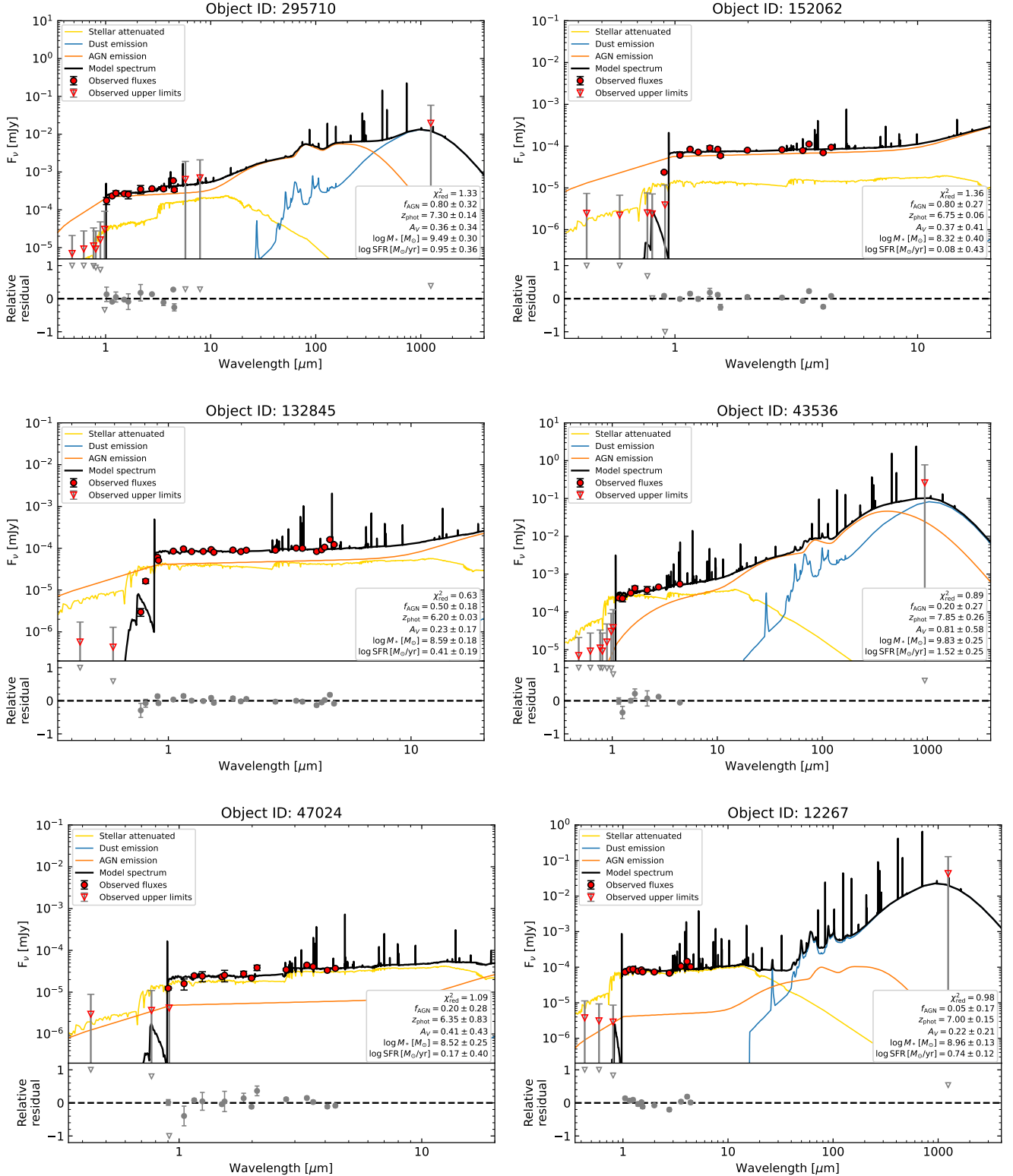
We will only consider candidates with  $f_{\text{AGN}} \geq 0.2$  to ensure the presence of actual AGN contribution to the observed emission. This threshold level is motivated by a comparative analysis of properties between active and inactive galaxies compiled from the literature, as elaborated in Appendix A. Overall, we anticipate that this cutoff will yield a completeness of approximately 80% in AGN selection, accompanied by a contamination rate as high as 30% from normal high- $z$  galaxies. We further impose a black hole mass ( $M_{\text{BH}}$ ) limit criterion, where  $M_{\text{BH}} > 10^5 M_{\odot}$  since confirming the quasar nature below this limit is challenging for numerous reasons. For instance, the bright host galaxy emission might dilute the quasar light, making the quasar signature hidden from the observers in the optical to NIR regimes (e.g., Fitriana & Murayama 2022). Furthermore, given the limitation of current observing facilities and the fact that these less massive quasars might only be capable of exhibiting H $\alpha$  with a line width of  $\lesssim 100 \text{ km s}^{-1}$ , they will be hard to differentiate from the low-velocity outflows or the narrow-line emissions of their host galaxies (Maiolino et al. 2024). Details on  $M_{\text{BH}}$  estimation will be discussed later in Sect. 4.2, but in the end, 64 sources passed these AGN fraction and mass limit criteria of the 350 parent candidates.

As a further note, out of the 11 previously confirmed AGNs at redshifts  $z \gtrsim 6$  reported by other studies (i.e., Harikane et al. 2023; Maiolino et al. 2023; Larson et al. 2023a; Kocevski et al. 2023; Übler et al. 2023; Kokorev et al. 2023; Stone et al. 2023), 9 sources met our selection criteria (see Table A.1). These known AGNs have intentionally been excluded from the final sample of the 64 quasar candidates presented here. These selected sources – that is, our final quasar candidates – are then marked as grade A while the unselected ones are labeled with grade B. All of our candidates are listed in Table B.1, which contains information on their coordinates, photometry, and derived properties. Due to the file size constraints, the full table and figures containing the SED fitting results of each source is exclusively available at the CDS.

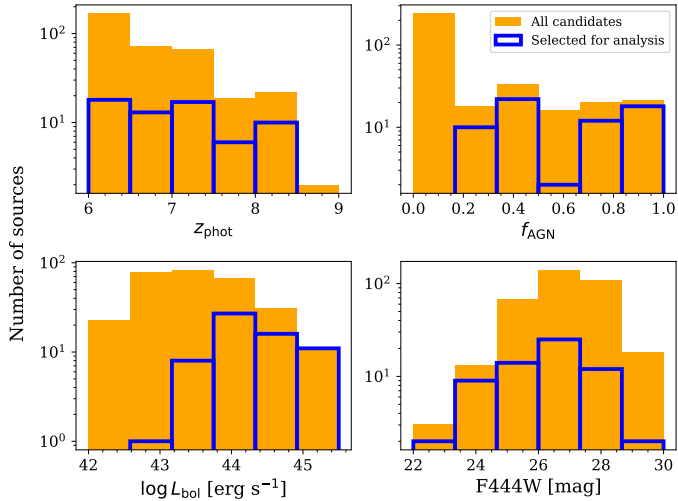
The sky coverages of each survey in the current datasets, for illustration, are approximately 0.28  $\text{deg}^2$ , 57  $\text{arcmin}^2$ , and 49  $\text{arcmin}^2$  for COSMOS-Web, JADES, and UNCOVER, respectively (see Table 1). Consequently, within the COSMOS-Web field and adopting the luminosity function of Harikane et al. (2023), we expect to find around 18 quasars at  $z = 6$ –8 having

<sup>9</sup> <https://cigale.lam.fr>





**Fig. 3.** Examples of SED fitting with CIGALE with AGN plus galaxy components. Observed and upper limit fluxes are shown in the upper part of each panel with red dots and triangles with error bars, respectively. The total model spectrum (black) in the observed-frame wavelengths, corrected for the IGM attenuation, is composed of stellar (yellow), dust, and AGN (orange) emissions. These decomposed components are shown without adding the IGM absorption model. We also report the reduced chi-square value ( $\chi^2_{\text{red}}$ ), fraction of AGN flux to the total emission ( $f_{\text{AGN}}$ ) within rest-wavelengths of 0.1–0.7  $\mu\text{m}$ , photometric redshift ( $z_{\text{phot}}$ ), dust extinction coefficient ( $A_V$ ), host galaxy stellar mass ( $M_*$ ), and star formation rate (SFR). The lower part of each panel displays the relative residual between the data and the model. Sources that are better modeled with no AGN contribution have  $f_{\text{AGN}} \leq 0.05$ , while the ones selected as quasar candidates exhibit  $f_{\text{AGN}} \geq 0.2$ .



**Fig. 4.** Distributions of the photometric redshift ( $z_{\text{phot}}$ ), fraction of AGN emission ( $f_{\text{AGN}}$ ) within the rest-wavelengths of 0.1–0.7  $\mu\text{m}$ , AGN bolometric luminosity ( $L_{\text{bol}}$ ), and  $F444W$  magnitude of the quasar candidates selected in this work. All candidates are shown with the orange histogram, while a subset with  $f_{\text{AGN}} \geq 0.2$  chosen for further analysis later in Sect. 4.3 is colored with blue.

the UV absolute magnitudes of  $M_{\text{UV}} \lesssim -21$ . With their deeper imaging, JADES and UNCOVER might recover about 12 and 23 sources brighter than  $M_{\text{UV}} \approx -19$ , respectively. Thus, the number of quasar candidates we found seems reasonable since it is within the appropriate range of the empirical predictions. We note that the luminosity function of Harikane et al. (2023) is derived based on the recent census of  $z \approx 4$ –7 low-luminosity AGNs ( $-18.5 \gtrsim M_{\text{UV}} \gtrsim -21.5$ ) detected with the JWST observations. In contrast, if we take and extrapolate the models from Matsuoka et al. (2018) or Schindler et al. (2023) into the fainter regimes, for which they were anchored initially to the bright ( $M_{\text{UV}} \lesssim -22$ ), unobscured quasar population at  $z \sim 6$ , we anticipate finding only one source in each field.

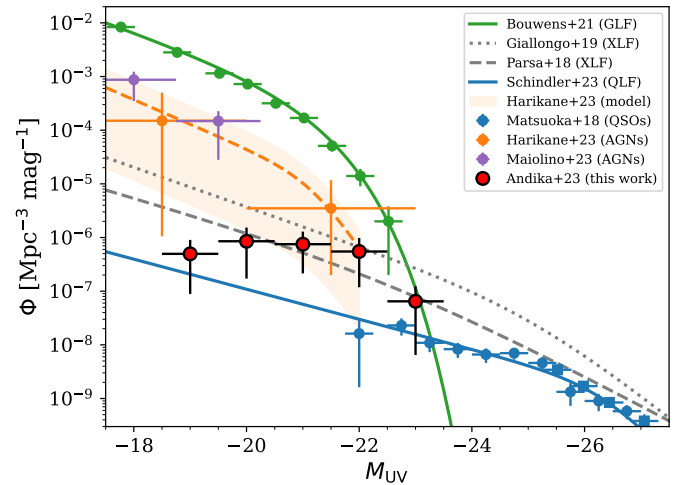
We present the number density of our quasar candidates in Table 2 and Fig. 5. Here, we consider a redshift range of  $z = 6.0$ – $8.4$ , and the total solid angle covered by our datasets is around  $0.45 \text{ deg}^2$ , which corresponds to a survey volume of approximately  $8.2 \times 10^6 \text{ Mpc}^3$ . The  $M_{\text{UV}}$  for our candidates are calculated from the flux observed at the rest-frame wavelength of  $1500 \text{ \AA}$ , derived based on our best-fit total SED model. Hence, the reported  $M_{\text{UV}}$  accounts for the total emission from the quasar plus its host galaxy component. Accordingly, to construct the UV luminosity function, we perform  $10^4$  Monte Carlo draws of our quasar candidates, incorporating their observed  $M_{\text{UV}}$  and  $z_{\text{phot}}$  along with their associated uncertainties. These random draws are necessary for instances where sources may fall outside the predefined redshift range or get counted in different magnitude bins across various iterations. Given that the quasar count depends on the chosen  $f_{\text{AGN}}$  threshold, we also vary this criterion from  $f_{\text{AGN}} \geq 0.2$  to  $0.9$  to take into account additional errors resulting from our selection method. Our error estimation also accounts for the possible presence of low- $z$  interlopers and inactive galaxies with a contamination rate of up to 30% (see, for example, Fig. 2 and Appendix A). We caution that the resulting number density estimation has not been adjusted for survey incompleteness.

In general, the number density of our quasar candidates exceeds the extrapolated values of the brighter quasar population luminosity function by a factor of  $\approx 10$  (e.g.,

**Table 2.** Number density of our  $6 \lesssim z \lesssim 8$  quasar candidates.

$M_{\text{UV}}$ [mag]	$\Phi$ [ $10^{-7} \text{ Mpc}^{-3} \text{ mag}^{-1}$ ]	$N$
$-23 \pm 0.5$	$0.65 \pm 0.96$	$1 \pm 1$
$-22 \pm 0.5$	$5.49 \pm 4.31$	$4 \pm 3$
$-21 \pm 0.5$	$7.51 \pm 5.37$	$6 \pm 4$
$-20 \pm 0.5$	$8.52 \pm 6.80$	$7 \pm 6$
$-19 \pm 0.5$	$4.97 \pm 4.07$	$4 \pm 3$

**Notes.** The columns from left to right are: (1) the UV absolute magnitude bins, (2) the average and standard deviation of the number densities, and (3) the average number of objects obtained from the Monte Carlo draws. The reported numbers are not corrected for possible survey incompleteness.



**Fig. 5.** UV luminosity functions of sources at  $z \sim 6$ . The number densities of our quasar candidates as a function of UV absolute magnitude are marked with red circles with error bars. Blue circles and squares represent the data from Matsuoka et al. (2018) and Schindler et al. (2023), where their associated best-fit quasar luminosity function (QLF) is shown with a blue line. The fitted model and observed galaxy luminosity function (GLF) from Bouwens et al. (2021) are displayed with a green line and circles, respectively. A sample of JWST-confirmed AGNs from Harikane et al. (2023) is designated with orange circles, and their double power-law model, along with its uncertainty, is portrayed with an orange dashed line and shaded region. We also show the AGN luminosity function at  $z = 4$ – $6$  reported by Maiolino et al. (2023) with purple circles for comparison. The number density of our quasar candidates is higher than the extrapolation of the bright QLF. Nevertheless, it is consistent with the X-ray selected AGN luminosity function (XLF) from Parsa et al. (2018) and Giallongo et al. (2019), which are denoted as dashed and dotted gray lines.

Matsuoka et al. 2018; Schindler et al. 2023), as shown by the blue line in Fig. 5. On the other hand, our numbers align with those reported by Harikane et al. (2023) to some extent; yet, densities at  $M_{\text{UV}} \gtrsim -19$  are largely uncertain given the source faintness and potential incompleteness in our quasar search method. Interestingly, our samples are consistent with the faint, X-ray-selected AGN luminosity function presented by Parsa et al. (2018) and Giallongo et al. (2019). The different nature of the bright and faint quasar populations might cause a large discrepancy between the luminosity functions mentioned earlier. At the same time, many of these faint sources are just being detected with JWST, and it is likely that much remains to be revealed. Below, we will discuss the constraints on the black hole and host galaxy characteristics of our quasar candidates.

#### 4.2. Black hole and host galaxy masses

The distribution of the central black hole mass to the host galaxy's stellar mass ratio – that is,  $M_{\text{BH}}/M_*$  – is a tracer of the supermassive black hole (SMBH) formation history (Volonteri 2012). We want to again treat our high-probability quasar candidates as actual quasars and, under that assumption, infer black hole and stellar masses for them. To estimate  $M_{\text{BH}}$ , we first adopt the canonical normalized accretion rate parameterized by the Eddington ratio,  $f_{\text{Edd}} \equiv L_{\text{bol}}/L_{\text{Edd}}$ , where  $L_{\text{bol}}$  and  $L_{\text{Edd}}$  are the bolometric and Eddington luminosities, respectively (e.g., Wu & Shen 2022). In this case,  $L_{\text{bol}}$  is calculated by multiplying a bolometric correction factor of 5.15 (Richards et al. 2006) with the monochromatic luminosity at the rest-frame wavelength of 3000 Å – that is,  $L_{3000} = \lambda L_{\lambda}(3000 \text{ \AA})$  – derived based on our best-fit AGN SED model obtained in Sect. 3.2. Then, we derive the lower limit  $M_{\text{BH}}$  of our quasar candidates, assuming  $f_{\text{Edd}} = 1$  and considering that Eddington luminosity can be approximated using:

$$L_{\text{Edd}} = 1.3 \times 10^{38} (M_{\text{BH}}/M_{\odot}) \text{ erg s}^{-1}. \quad (1)$$

As the values derived here represent lower limits, the true  $M_{\text{BH}}$  could potentially be significantly higher. A comparison between our SED-based  $M_{\text{BH}}$  and those determined through broad emission line spectroscopy reveals an actual  $M_{\text{BH}}$  that is  $\approx 1.6$  dex higher, as demonstrated in Appendix A. The observed offset is anticipated, given the significant influence of  $f_{\text{Edd}}$  on our  $M_{\text{BH}}$  estimates. Adjusting the assumed  $f_{\text{Edd}}$  to a much lower value, such as 0.1, results in a 1 dex increase in our data points, bringing them closer to spectroscopic  $M_{\text{BH}}$  values.

The inferred  $M_{\text{BH}}/M_*$  distribution of our quasar candidates inferred from Eq. (1) and Sect. 3.2 is displayed in Fig. 6. This distribution assumes that our quasar candidates may exhibit  $f_{\text{Edd}}$  values ranging from 0.1 to 1 and includes that uncertainty. While our quasar candidates display a  $M_*-M_{\text{BH}}$  distribution slightly higher than that of galaxies at  $z \sim 0$  (e.g., Kormendy & Ho 2013; Reines & Volonteri 2015), with properties consistent with observed samples of other high- $z$  low-luminosity AGNs (e.g., Harikane et al. 2023; Kocevski et al. 2023), we emphasize that the derived  $M_{\text{BH}}$  values represent lower limits.

Luminous quasars hosting massive black holes tend to reside within galaxies with larger stellar masses, the  $M_{\text{BH}}$  to  $M_*$  ratios show a large diversity (see, e.g., Inayoshi et al. 2020; Fan et al. 2023). For instance,  $z \gtrsim 5$  bright quasars examined by Yue et al. (2023) display  $M_{\text{BH}}/M_*$  reaching as high as 10%, which is significantly more prominent compared to the sources in the nearby Universe (e.g., Kormendy & Ho 2013). On the other hand, less luminous objects, such as samples of  $4 \lesssim z \lesssim 7$  AGNs from Harikane et al. (2023) are characterized by relatively lower  $M_{\text{BH}}/M_* \sim 1\%$ . To add further support of this diversity, Larson et al. (2023a) reported a presence of a broad-line AGN at  $z = 8.679$  exhibiting an  $M_{\text{BH}}/M_* \approx 0.3\%$ , while, conversely, Furtak et al. (2023) presented an AGN at  $z = 7.045$  having  $M_{\text{BH}}/M_* \gtrsim 3\%$ . Here, we need to note that for all samples, there are strong selection effects at play (e.g., Li et al. 2022), biasing against the ability to see low-luminosity AGN in bright galaxies. The exact impact will depend on the selection method but might imply limits by SED preselection, color-color cuts, emission-line strengths, or – as for our approach – a minimal required AGN fraction of the total flux. What all methods have in common is that they will preferentially find massive SMBHs. With that in mind, the comparison mentioned above implies that the growth of SMBHs at the upper envelope of these actually selected bright quasars may have preceded the star formation

in their host galaxies (Kokorev et al. 2023; Maiolino et al. 2024; Pacucci et al. 2023).

Whether the  $M_{\text{BH}}-M_*$  relation evolves with redshift is still a subject of debate. For example, Caplar et al. (2018) proposed an increasing SMBH to host mass ratio at higher redshifts, that is,  $M_{\text{BH}}/M_* \propto (1+z)^{1.5}$ , which was inferred using an analytical approach to obtain the  $M_{\text{BH}}-M_*$  relation that fits the observed quasar luminosity function and SFR density (see also Pacucci & Loeb 2024). On the other hand, considering various observable SMBH and host galaxy properties, including mass functions and quasar distributions, Zhang et al. (2023) demonstrated that there is no significant evolution of  $M_{\text{BH}}-M_*$  up to  $z \sim 10$  (see also, for example, Suh et al. 2020; Ding et al. 2020; Li et al. 2021). In addition, in flux-limited surveys, quasars harboring overmassive black holes – e.g.,  $M_{\text{BH}}/M_* > 0.01$  – could dominate the picked-up samples due to selection effects (Lauer et al. 2007). As seen in Fig. 6, luminous quasars investigated by Yue et al. (2023), Übler et al. (2023), and Stone et al. (2023) lie way above the local  $M_{\text{BH}}-M_*$  relation, indicating a potential bias mentioned earlier. This bias might occur because larger SMBH masses could produce higher quasar luminosities, which are more accessible to locate in flux-limited observations.

#### 4.3. Possible pathways for SMBH growth

The significant diversity observed in the most distant SMBHs and their host galaxies might suggest a range of distinct growth histories and progenitors, which we will discuss further here. While the exact seeding mechanisms remain elusive, it is generally accepted that early SMBHs might originate from at least two types of progenitors: (i) light seeds arising from the remnants of Population III stars having masses of  $\approx 10-100 M_{\odot}$  and (ii) heavy seeds with a mass range of  $10^4-10^6 M_{\odot}$  produced by the collapse of primordial gas or dense star clusters (Inayoshi et al. 2020).

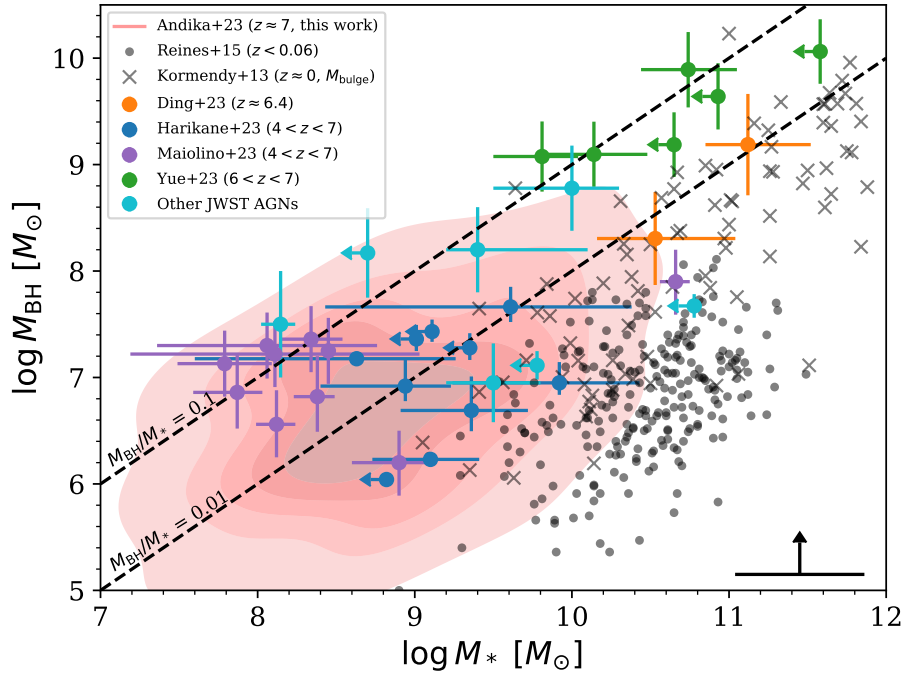
Here, we want to trace back the growth of our quasar candidates following the method presented by Pacucci & Loeb (2022). As the first step, we describe the connection between the initial seed mass  $M_{\text{seed}}$  and the accumulated black hole mass  $M_{\text{BH}}$  at a specific cosmic time  $t$  using the relation:

$$M_{\text{BH}}(t) = M_{\text{seed}} \exp\left(f_{\text{Edd}} D \frac{1-\epsilon}{\epsilon} \frac{\Delta t}{t_{\text{Edd}}}\right). \quad (2)$$

In this case,  $t_{\text{Edd}}$  is a mixture of constants where its typical value is  $\approx 450$  Myr (Pacucci & Loeb 2022),  $f_{\text{Edd}}$  is the average Eddington ratio across the accretion time interval of  $\Delta t$ , and  $\epsilon$  is the mean radiative efficiency over the  $\Delta t$ . The time interval is expressed as  $\Delta t = t - t_{\text{seed}}$ , where  $t_{\text{seed}}$  corresponds to the black hole seeding epoch. Assuming that a black hole seed is assembled at  $z = 25$ , this would equal a cosmic time of  $t_{\text{seed}} \approx 130$  Myr. For an accretion mode following the thin disk model,  $\epsilon$  ranges from 0.34 down to 0.057, depending on whether the central black hole is maximally rotating or nonrotating (Fabian & Lasenby 2019; Pacucci & Loeb 2020; Ananna et al. 2020). Then, the fraction of the quasar lifetime for which the accretion occurs is parametrized with the duty cycle  $D$ . Unfortunately,  $f_{\text{Edd}}$  and  $D$  are degenerate, meaning that one can obtain an identical value of  $M_{\text{BH}}$  by combining different values of both growth parameters. Due to this reason, we assume that the sources are actively accreting throughout their entire lifetime so that  $D$  can be set to unity for simplicity.

We subsequently simulate the SMBH growth using a simple Monte Carlo strategy with 2000 realizations, exploiting Eq. (2) as the target function. Starting from smaller seeds, we aim to match the masses of our quasar candidates that are, on average,





**Fig. 6.** Relation between the black hole mass ( $M_{\text{BH}}$ ) and its host galaxy stellar mass ( $M_*$ ). The red contour represents our quasar candidates at  $z \geq 6$ , where our measurements can only provide lower limits for  $M_{\text{BH}}$ , considering Eddington ratio values ranging from 0.1 to 1. The typical statistical errors for  $M_*$  are indicated in the lower right corner of the panel. High- $z$  quasar samples with available JWST spectroscopic data from Harikane et al. (2023), Yue et al. (2023), Ding et al. (2023), and Maiolino et al. (2023, excluding dual AGNs) are presented with blue, green, orange, and purple circles with error bars. Additional AGN samples from Larson et al. (2023a), Übler et al. (2023), Stone et al. (2023), Kocevski et al. (2023), Kokorev et al. (2023), and Goulding et al. (2023) are indicated with cyan circles. The gray dots and crosses are nearby galaxies and AGNs from Kormendy & Ho (2013) and Reines & Volonteri (2015). The black dashed lines mark the limits where  $M_{\text{BH}}/M_*$  equals 0.1 and 0.01. Our candidates show a slightly higher  $M_{\text{BH}}$  to  $M_*$  ratios than other galaxies at  $z \sim 0$  with consistent properties compared to high- $z$  low-luminosity quasars.

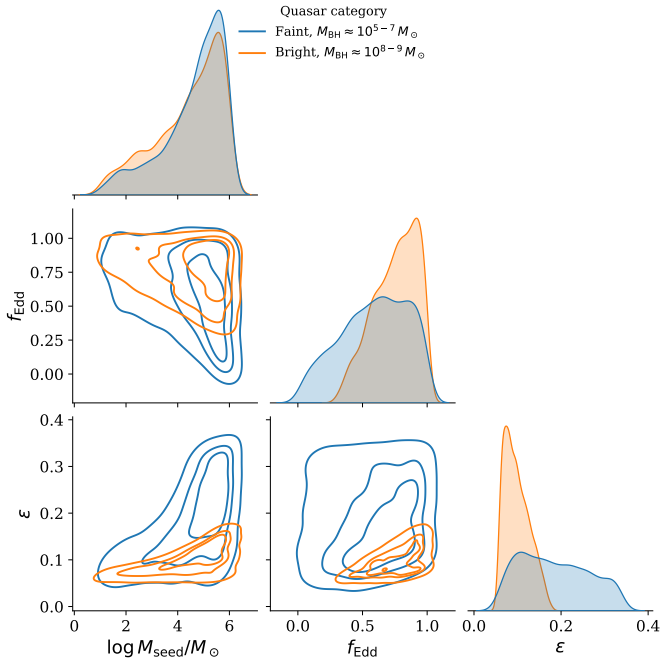
within  $M_{\text{BH}} = 10^5\text{--}10^8 M_{\odot}$ , at a median redshift of  $z = 6.7$ . Three essential parameters control our growth model, that is,  $M_{\text{seed}}$ ,  $f_{\text{Edd}}$ , and  $\epsilon$ . Correspondingly, we adopt a flat prior of  $\log M_{\text{seed}} \in [1, 6] M_{\odot}$ , covering both the light and heavy seed mass regimes, and adjust the radiative efficiency in a physical range of  $\epsilon \in [0.057, 0.34]$ . Considering that (i) many high- $z$  quasars detected so far are showing instantaneous accretion rates below or around the Eddington limit (e.g., Fan et al. 2023) and (ii) super- or hyper-Eddington accretion periods are typically short-lived ( $\Delta t \sim 0.1$  Myr), we adopt the Eddington ratio to be uniformly distributed within  $f_{\text{Edd}} \in [0, 1]$  (e.g., Fragione & Pacucci 2023). While  $M_{\text{seed}}$  is constant since the black hole is only seeded one time,  $\epsilon$  and  $f_{\text{Edd}}$ , on the other hand, may change over the period between the seed formation until it is detected at a later time. Thus, these two parameters should be viewed as averages over the quasar lifetime.

The combination of parameters that permits the formation of the central SMBHs we assume are residing in our quasar candidates is presented in Fig. 7. At the same time, the associated growth track is provided in Fig. 8. The majority of bright quasars from Fan et al. (2023) occupy the region where  $M_{\text{BH}} \gtrsim 10^8 M_{\odot}$  while our candidates, as well as some JWST-confirmed AGNs, reside in the lower mass side. (e.g., Goulding et al. 2023; Übler et al. 2023; Stone et al. 2023; Larson et al. 2023a; Kocevski et al. 2023; Harikane et al. 2023; Maiolino et al. 2023; Greene et al. 2024). Larger seeds with  $M_{\text{seed}} > 10^4 M_{\odot}$  seem to be the preferred progenitor to develop these SMBHs by  $z \approx 7$ . In particular, most of our quasar candidates might have arisen from the black hole seeds as big as  $M_{\text{seed}} \sim 10^5 M_{\odot}$ , assuming the values of  $f_{\text{Edd}} = 0.6 \pm 0.3$  and  $\epsilon = 0.2 \pm 0.1$ . If short super-Eddington episodes occur during their evolution, the required

progenitor mass could be lower, indicating that dense star cluster seeds could also be the ancestors of our quasar candidates. Distinguishing between the formation through direct collapse black hole or dense star cluster channels is complicated, given the necessity of more precise measurements of the SMBH and host galaxy masses along with the gas metallicity, denoting that extra spectroscopic data are needed (Volonteri et al. 2023).

After that, we run similar modeling as a comparison, but now targeting the bright quasars with  $M_{\text{BH}} \approx 10^8\text{--}10^9 M_{\odot}$  compiled by Fan et al. (2023). As a result, this population also gives preference for heavy seeds with the Eddington ratio pushed higher to  $f_{\text{Edd}} = 0.78 \pm 0.17$  and radiative efficiency going down to  $\epsilon = 0.09 \pm 0.03$ . We note that the parameter space occupied by this population is tighter than our less luminous quasar candidates, showing that detecting larger SMBHs at the farthest accessible distances could shrink the viable growth parameters and modes significantly. Furthermore, our simple calculation confirms that as long as the radiative efficiency is at the lower end of the range accommodated by the thin disk model and the accretion is not dominated by super-Eddington episodes, it is less likely to yield SMBHs from the light seeds.

Maturing the light seeds in a short amount of time is complicated as this process would require the growth dominated with the Eddington-limited ( $f_{\text{Edd}} = 1$ ) or even super-Eddington ( $f_{\text{Edd}} > 1$ ) accretion to match the  $z \approx 6\text{--}7$  quasar mass distribution. However, assembling such enormous masses and sustaining high accretion rates will be challenging, given the intricacy created by the enhanced stellar feedback from the host galaxies. For example, a vast number of supernova explosions will happen during the rapid mass build-up, resulting in intense heating and mixing of the gas, making the accretion inefficient and



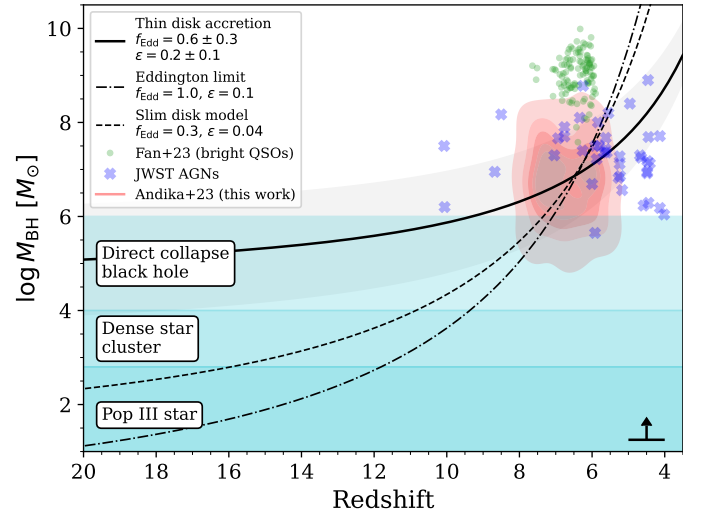
**Fig. 7.** Combination of  $M_{\text{seed}}$ ,  $f_{\text{Edd}}$ ,  $\epsilon$  values that can produce the observed  $M_{\text{BH}}$  of high- $z$  quasars. The parameter distributions of our quasar candidates and sources from Fan et al. (2023) are depicted in blue and orange colors, respectively. Assuming a thin disk model and Eddington-limited accretion, larger seed masses with  $M_{\text{seed}} > 10^4 M_{\odot}$  are the preferred channel for growing the SMBHs.

more likely to be sub-Eddington (Larson et al. 2023a). The only way to develop the light seeds into SMBHs is probably to adopt a hypothetical slim disk scenario, which lowers the radiative efficiency to  $\epsilon = 0.04$  (Abramowicz et al. 1988; Mineshige et al. 2000; Pacucci et al. 2015; Volonteri et al. 2015). With just a mild accretion of  $f_{\text{Edd}} = 0.3$ , for example, this channel could already produce  $\gtrsim 10^6 M_{\odot}$  black holes by  $z \approx 7$  as shown in Fig. 8. Whether this mass accumulation channel dominates the high- $z$  quasar population is still debatable. Therefore, further study to understand the typical accretion mode and the interplay between the growth parameters of early black holes is vital to constrain the evolution of these intriguing sources.

## 5. Summary and conclusion

We have presented 350 candidates of compact galaxies, of which 64 show a high probability of being quasars at  $z \gtrsim 6$ , selected by exploiting the rich multiband dataset of COSMOS-Web, as well as the JADES, UNCOVER, CEERS, and PRIMER projects. These surveys consist primarily of JWST/NIRCam observations. The subsequent photometric catalog creation incorporated ancillary data from HST and other ground-based surveys. Accordingly, our search strategy consists of two primary steps: photometric cut on catalog-level information and SED fitting to separate the quasars from other contaminating sources. While the initial goals of the SED fitting are to classify and estimate the photometric redshift of each candidate, we also assess their associated physical properties under the assumption that they indeed are quasars, including the SMBH and host galaxy’s stellar masses, as well as the fraction of AGN emission.

Our quasar candidates exhibit features consistent with the low-luminosity AGN population, where they potentially host less massive SMBHs with  $M_{\text{BH}} \approx 10^5\text{--}10^8 M_{\odot}$  residing in galaxies having  $M_{*} \approx 10^8\text{--}10^{10} M_{\odot}$ . Furthermore, these



**Fig. 8.** Black hole mass growth as a function of redshift. The red contour represents the expected lower limit masses of the quasar candidates identified in this study. Additionally, the typical photometric redshift uncertainty for these candidates is illustrated in the lower right corner of the panel. Green circles depict the bright quasar samples from Fan et al. (2023) while blue crosses display broad-line AGNs from the literature that have been observed with JWST spectroscopy (see main text). The cyan-shaded region shows the mass range of different progenitors. The solid black line and region show an evolutionary track along with its posterior distribution, assuming a thin disk accretion and heavy seed progenitors. On the other hand, cases where we use a thin disk accretion at the Eddington limit and a slim disk model to grow light seeds into SMBHs are shown with dash-dotted and dashed black lines.

sources display  $M_{\text{BH}}\text{--}M_{*}$  distribution that is slightly higher than those of galaxies at  $z \sim 0$  (e.g., Kormendy & Ho 2013; Reines & Volonteri 2015), or in other words, their SMBHs tend to be overmassive compared to their hosts. However, we stress that all quasars identified in these surveys are naturally biased to high  $M_{\text{BH}}/M_{*}$ -ratios. This means they are not representative of the underlying population but preferentially form the upper envelope of the distribution.

With this in mind, we then run a simple Monte Carlo simulation to explain how these SMBHs accumulate their mass by the time they are detected. Larger seeds from the direct collapse scenario, with  $M_{\text{seed}} > 10^4 M_{\odot}$ , appear to be the favored origins to develop these SMBHs by  $z \sim 7$ . Notably, most of our quasar candidates might have emerged from the black hole seeds as large as  $M_{\text{seed}} \sim 10^5 M_{\odot}$ , considering the values of  $f_{\text{Edd}} = 0.6 \pm 0.3$  and  $\epsilon = 0.2 \pm 0.1$  – that is, the Eddington limited accretion in thin disk model. If brief super-Eddington events arise during their growth, the required progenitor mass could be smaller, implying that dense star cluster seeds could also be the ancestors of our quasar candidates.

As we have offered the most promising and robust high- $z$  quasar candidates in this paper, further confirmation is vital to uncover their true nature. For example, spectroscopy with JWST would be the best opportunity to acquire the rest-frame UV/optical spectra of these quasars, allowing the detection of broad emission lines to get more precise SMBH mass measurements and gas-phase metallicity. In addition to that, we can probe the cold molecular gas, tracing the galaxy dynamics and star formation activity, with ALMA. With all of that being said, the samples presented in this work are ideal laboratories for dissecting the nature of the first galaxies and SMBHs formed during the reionization era.

**Acknowledgements.** We express our gratitude to the referee, Chiara Feruglio, for the constructive and insightful comments. We thank Sherry Suyu and Dian Triani for the helpful feedback and fruitful discussions, which have significantly enhanced the quality of this paper. This research is supported in part by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. J.S. is supported by JSPS KAKENHI (JP22H01262), the World Premier International Research Center Initiative (WPI), MEXT, Japan and the JSPS Core-to-Core Program (grant number: JPJSCA20210003). M.H. acknowledges funding from the Swiss National Science Foundation (SNF) via a PRIMA Grant PR00P2 193577 “From cosmic dawn to high noon: the role of black holes for young galaxies”. S.G. acknowledges financial support from the Villum Young Investigator grants 37440 and 13160 and the Cosmic Dawn Center (DAWN), funded by the Danish National Research Foundation (DNRF) under grant No. 140. Some of the data products presented herein were retrieved from the Dawn JWST Archive (DJA). DJA is an initiative of the Cosmic Dawn Center. B.T. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 950533) and from the Israel Science Foundation (grant 1849/19). The Flatiron Institute is supported by the Simons Foundation. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This research is based on observations made with the NASA/ESA *Hubble* Space Telescope obtained from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. This work is based in part on observations made with the NASA/ESA/CSA *James Webb* Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. Facilities. ALMA, ESO:VISTA (VIRCAM), HST (ACS, WFC3), JWST (NIRCam), *Spitzer* (IRAC), Subaru (HSC). Software. Astropy (Astropy Collaboration 2013, 2018), CIGALE (Boquien et al. 2019; Yang et al. 2020, 2022), Dask (Rocklin et al. 2015), eazy-py (Brammer et al. 2008), grizli (Brammer et al. 2022), Matplotlib (The Matplotlib Development Team 2024), msaexp (Brammer 2023), NumPy (Harris et al. 2020), Pandas (The pandas development team 2024), Seaborn (Waskom 2021), TOPCAT (Taylor 2005).

## References

- Abramowicz, M. A., Czerny, B., Lasota, J. P., & Szuszkiewicz, E. 1988, *ApJ*, **332**, 646
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, *PASJ*, **74**, 247
- Alexander, T., & Natarajan, P. 2014, *Science*, **345**, 1330
- Ananna, T. T., Salvato, M., LaMassa, S., et al. 2017, *ApJ*, **850**, 66
- Ananna, T. T., Urry, C. M., Treister, E., et al. 2020, *ApJ*, **903**, 85
- Andika, I. T., Jahnke, K., Onoue, M., et al. 2020, *ApJ*, **903**, 34
- Andika, I. T., Jahnke, K., Bañados, E., et al. 2022, *AJ*, **163**, 251
- Andika, I. T., Jahnke, K., van der Wel, A., et al. 2023a, *ApJ*, **943**, 150
- Andika, I. T., Suyu, S. H., Cañameras, R., et al. 2023b, *A&A*, **678**, A103
- Arrabal Haro, P., Dickinson, M., Finkelstein, S. L., et al. 2023, *ApJ*, **951**, L22
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2023, *ApJ*, **946**, L12
- Bagley, M. B., Finkelstein, S. L., Rojas-Ruiz, S., et al. 2024, *ApJ*, **961**, 209
- Bañados, E., Novak, M., Neeleman, M., et al. 2019, *ApJ*, **881**, L23
- Begelman, M. C., & Volonteri, M. 2017, *MNRAS*, **464**, 1102
- Bertin, E., Schefer, M., Apostolakis, N., et al. 2020, *ASP Conf. Ser.*, **527**, 461
- Bertin, E., Schefer, M., Apostolakis, N., et al. 2022, *Astrophysics Source Code Library* [record ascl:2212.018]
- Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2022, *ApJ*, submitted [arXiv:2212.04026]
- Boekholt, T. C. N., Schleicher, D. R. G., Fellhauer, M., et al. 2018, *MNRAS*, **476**, 366
- Boquien, M., Burgarella, D., Roehly, Y., et al. 2019, *A&A*, **622**, A103
- Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, *AJ*, **162**, 47
- Brammer, G. 2023, <https://doi.org/10.5281/zenodo.6672538>
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Brammer, G., Strait, V., Matharu, J., & Momcheva, I. 2022, <https://doi.org/10.5281/zenodo.6672538>
- Bromm, V., & Loeb, A. 2003, *ApJ*, **596**, 34
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2023, *A&A*, submitted [arXiv:2306.02467]
- Bushouse, H., Eisenhamer, J., Dencheva, N., et al. 2022, <https://doi.org/10.5281/zenodo.7325378>
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Caplar, N., Lilly, S. J., & Trakhtenbrot, B. 2018, *ApJ*, **867**, 148
- Carrasco Kind, M., & Brunner, R. J. 2013, *MNRAS*, **432**, 1483
- Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2023, *ApJ*, **954**, 31
- Chabrier, G. 2003, *PASP*, **115**, 763
- Conroy, C., & Gunn, J. E. 2010, *ApJ*, **712**, 833
- Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, **699**, 486
- Conroy, C., White, M., & Gunn, J. E. 2010, *ApJ*, **708**, 58
- Ding, X., Silverman, J., Treu, T., et al. 2020, *ApJ*, **888**, 37
- Ding, X., Onoue, M., Silverman, J. D., et al. 2023, *Nature*, **621**, 51
- Dubois, Y., Volonteri, M., & Silk, J. 2014, *MNRAS*, **440**, 1590
- Duncan, K. J., Kondapally, R., Brown, M. J. I., et al. 2021, *A&A*, **648**, A4
- Dunlop, J. S., Abraham, R. G., Ashby, M. L. N., et al. 2021, *JWST Proposal Cycle 1*, #1837
- Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, *ApJS*, submitted [arXiv:2306.02465]
- Erb, D. K., Pettini, M., Shapley, A. E., et al. 2010, *ApJ*, **719**, 1168
- Euclid Collaboration (Moneti, A., et al.) 2022, *A&A*, **658**, A126
- Fabian, A. C., & Lasenby, A. N. 2019, arXiv e-prints [arXiv:1911.04305]
- Fan, X., Bañados, E., & Simcoe, R. A. 2023, *ARA&A*, **61**, 373
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, **946**, L13
- Fitriana, I. K., & Murayama, T. 2022, *PASJ*, **74**, 689
- Fitzpatrick, E. L. 1999, *PASP*, **111**, 63
- Fragione, G., & Pacucci, F. 2023, *ApJ*, **958**, L24
- Fujimoto, S., Wang, B., Weaver, J., et al. 2023a, *ApJ*, submitted [arXiv:2308.11609]
- Fujimoto, S., Bezanson, R., Labbe, I., et al. 2023b, *ApJS*, submitted [arXiv:2309.07834]
- Fujimoto, S., Arrabal Haro, P., Dickinson, M., et al. 2023c, *ApJ*, **949**, L25
- Furtak, L. J., Labbé, I., Zitrin, A., et al. 2023, arXiv e-prints [arXiv:2308.05735]
- Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, **674**, A1
- Giallongo, E., Grazian, A., Fiore, F., et al. 2019, *ApJ*, **884**, 19
- Goulding, A. D., Greene, J. E., Setton, D. J., et al. 2023, *ApJ*, **955**, L24
- Green, G. M. 2018, *J. Open Source Softw.*, **3**, 695
- Greene, J. E., Labbe, I., Goulding, A. D., et al. 2024, *ApJ*, **964**, 39
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Hainline, K. N., Johnson, B. D., Robertson, B., et al. 2024, *ApJ*, **964**, 71
- Harikane, Y., Zhang, Y., Nakajima, K., et al. 2023, *ApJ*, **959**, 39
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Heintz, K. E., Watson, D., Brammer, G., et al. 2023, arXiv e-prints [arXiv:2306.00647]
- Hosokawa, T., Yorke, H. W., Inayoshi, K., Omukai, K., & Yoshida, N. 2013, *ApJ*, **778**, 178
- Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, *A&A*, **553**, A6
- Inayoshi, K., Visbal, E., & Haiman, Z. 2020, *ARA&A*, **58**, 27
- Inoue, A. K. 2011, *MNRAS*, **415**, 2920
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, **442**, 1805
- Isobe, Y., Ouchi, M., Nakajima, K., et al. 2023, *ApJ*, **956**, 139
- Izumi, T., Matsuoka, Y., Fujimoto, S., et al. 2021, *ApJ*, **914**, 36
- Kaasinen, M., Bian, F., Groves, B., Kewley, L. J., & Gupta, A. 2017, *MNRAS*, **465**, 3220
- Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, *ARA&A*, **57**, 511
- Kocevski, D. D., Onoue, M., Inayoshi, K., et al. 2023, *ApJ*, **954**, L4
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, **172**, 196
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Kokorev, V., Brammer, G., Fujimoto, S., et al. 2022, *ApJS*, **263**, 38
- Kokorev, V., Fujimoto, S., Labbe, I., et al. 2023, *ApJ*, **957**, L7
- Kokorev, V., Caputi, K. I., Greene, J. E., et al. 2024, *ApJL*, submitted [arXiv:2401.09981]
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, **51**, 511
- Labbe, I., Greene, J. E., Bezanson, R., et al. 2023, *ApJ*, submitted [arXiv:2306.07320]
- Larson, R. L., Finkelstein, S. L., Kocevski, D. D., et al. 2023a, *ApJ*, **953**, L29
- Larson, R. L., Hutchison, T. A., Bagley, M., et al. 2023b, *ApJ*, **958**, 141
- Latif, M. A., Khochfar, S., Schleicher, D., & Whalen, D. J. 2021, *MNRAS*, **508**, 1756
- Lauer, T. R., Tremaine, S., Richstone, D., & Faber, S. M. 2007, *ApJ*, **670**, 249
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, **379**, 1599
- Leitherer, C., Li, I. H., Calzetti, D., & Heckman, T. M. 2002, *ApJS*, **140**, 303
- Li, J., Silverman, J. D., Ding, X., et al. 2021, *ApJ*, **922**, 142
- Li, J., Silverman, J. D., Izumi, T., et al. 2022, *ApJ*, **931**, L11
- Liu, D., Lang, P., Magnelli, B., et al. 2019, *ApJS*, **244**, 40
- Lodato, G., & Natarajan, P. 2006, *MNRAS*, **371**, 1813
- Lupi, A., Haardt, F., Dotti, M., et al. 2016, *MNRAS*, **456**, 2993
- Madau, P., Haardt, F., & Dotti, M. 2014, *ApJ*, **784**, L38
- Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2023, *A&A*, submitted [arXiv:2308.01230]
- Maiolino, R., Scholtz, J., Witstok, J., et al. 2024, *Nature*, **627**, 59
- Marley, M. S., Saumon, D., Visscher, C., et al. 2021, *ApJ*, **920**, 85



- Massonneau, W., Volonteri, M., Dubois, Y., & Beckmann, R. S. 2023, *A&A*, **670**, A180
- Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. 2018, *ApJ*, **869**, 150
- Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2019, *ApJ*, **883**, 183
- Mayer, L., & Bonoli, S. 2019, *Rep. Prog. Phys.*, **82**, 016901
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- Meiksin, A. 2006, *MNRAS*, **365**, 807
- Meyer, R. A., Barrufet, L., Boogaard, L. A., et al. 2024, *A&A*, **681**, L3
- Middleton, M. J., Miller-Jones, J. C. A., Markoff, S., et al. 2013, *Nature*, **493**, 187
- Mineshige, S., Kawaguchi, T., Takeuchi, M., & Hayashida, K. 2000, *PASJ*, **52**, 499
- Morishita, T., Roberts-Borsani, G., Treu, T., et al. 2023, *ApJ*, **947**, L24
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Nature*, **474**, 616
- Nabizadeh, A., Zackrisson, E., Pacucci, F., et al. 2024, *A&A*, **683**, A58
- Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, *ApJS*, **269**, 33
- Natarajan, P. 2021, *MNRAS*, **501**, 1413
- Natarajan, P., Pacucci, F., Ricarte, A., et al. 2024, *ApJ*, **960**, L1
- Nishizawa, A. J., Hsieh, B.-C., Tanaka, M., & Takata, T. 2020, arXiv e-prints [arXiv:2003.01511]
- Oesch, P. A., Brammer, G., Naidu, R. P., et al. 2023, *MNRAS*, **525**, 2864
- Pacucci, F., & Loeb, A. 2020, *ApJ*, **895**, 95
- Pacucci, F., & Loeb, A. 2022, *MNRAS*, **509**, 1885
- Pacucci, F., & Loeb, A. 2024, *ApJ*, **964**, 154
- Pacucci, F., Volonteri, M., & Ferrara, A. 2015, *MNRAS*, **452**, 1922
- Pacucci, F., Natarajan, P., Volonteri, M., Cappelluti, N., & Urry, C. M. 2017, *ApJ*, **850**, L42
- Pacucci, F., Nguyen, B., Carniani, S., Maiolino, R., & Fan, X. 2023, *ApJ*, **957**, L3
- Parsa, S., Dunlop, J. S., & McLure, R. J. 2018, *MNRAS*, **474**, 2904
- Pérez-González, P. G., Barro, G., Rieke, G. H., et al. 2024, *ApJ*, submitted [arXiv:2401.08782]
- Pezzulli, E., Valiante, R., & Schneider, R. 2016, *MNRAS*, **458**, 3047
- Reines, A. E., & Volonteri, M. 2015, *ApJ*, **813**, 82
- Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, *ApJS*, **166**, 470
- Roberts-Borsani, G., Morishita, T., Treu, T., et al. 2022, *ApJ*, **938**, L13
- Rocklin, M. 2015, in *Proceedings of the 14th Python in Science Conference*, eds. K. Huff, & J. Bergstra, 130
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, *ApJ*, **962**, 24
- Saxena, A., Robertson, B. E., Bunker, A. J., et al. 2023, *A&A*, **678**, A68
- Schindler, J.-T., Bañados, E., Connor, T., et al. 2023, *ApJ*, **943**, 67
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Smith, A., & Bromm, V. 2019, *Contemp. Phys.*, **60**, 111
- Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, *MNRAS*, **420**, 2756
- Stalevski, M., Ricci, C., Ueda, Y., et al. 2016, *MNRAS*, **458**, 2288
- Stone, M. A., Lyu, J., Rieke, G. H., & Alberts, S. 2023, *ApJ*, **953**, 180
- Suh, H., Civano, F., Trakhtenbrot, B., et al. 2020, *ApJ*, **889**, 32
- Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, *AJ*, **117**, 68
- Tang, M., Stark, D. P., Chen, Z., et al. 2023, *MNRAS*, **526**, 1657
- Taylor, M. B. 2005, *ASP Conf. Ser.*, **347**, 29
- The Matplotlib Development Team 2024, <https://doi.org/10.5281/zenodo.592536>
- The pandas development team 2024, <https://doi.org/10.5281/zenodo.3509134>
- Trakhtenbrot, B., Volonteri, M., & Natarajan, P. 2017, *ApJ*, **836**, L1
- Trinca, A., Schneider, R., Maiolino, R., et al. 2023, *MNRAS*, **519**, 4753
- Übler, H., Maiolino, R., Curtis-Lake, E., et al. 2023, *A&A*, **677**, A145
- Valentino, F., Brammer, G., Gould, K. M. L., et al. 2023, *ApJ*, **947**, 20
- Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016, *MNRAS*, **457**, 3356
- Venemans, B. P., Walter, F., Neeleman, M., et al. 2020, *ApJ*, **904**, 130
- Volonteri, M. 2010, *A&A Rev.*, **18**, 279
- Volonteri, M. 2012, *Science*, **337**, 544
- Volonteri, M., Silk, J., & Dubus, G. 2015, *ApJ*, **804**, 148
- Volonteri, M., Habouzit, M., & Colpi, M. 2021, *Nat. Rev. Phys.*, **3**, 732
- Volonteri, M., Habouzit, M., & Colpi, M. 2023, *MNRAS*, **521**, 241
- Wang, F., Yang, J., Fan, X., et al. 2021, *ApJ*, **907**, L1
- Waskom, M. 2021, *J. Open Source Softw.*, **6**, 3021
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, **258**, 11
- Weaver, J. R., Zalesky, L., Kokorev, V., et al. 2023, *ApJS*, **269**, 20
- Weaver, J. R., Cutler, S. E., Pan, R., et al. 2024, *ApJS*, **270**, 7
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, **143**, 9
- Whitaker, K. E., Ashas, M., Illingworth, G., et al. 2019, *ApJS*, **244**, 16
- Williams, C. C., Alberts, S., Ji, Z., et al. 2023a, *AAS J.*, submitted [arXiv:2311.07483]
- Williams, C. C., Tacchella, S., Maseda, M. V., et al. 2023b, *ApJS*, **268**, 64
- Woods, T. E., Agarwal, B., Bromm, V., et al. 2019, *PASA*, **36**, e027
- Wu, Q., & Shen, Y. 2022, *ApJS*, **263**, 42
- Yang, G., Boquien, M., Buat, V., et al. 2020, *MNRAS*, **491**, 740
- Yang, J., Wang, F., Fan, X., et al. 2021, *ApJ*, **923**, 262
- Yang, G., Boquien, M., Brandt, W. N., et al. 2022, *ApJ*, **927**, 192
- Yoo, J., & Miralda-Escudé, J. 2004, *ApJ*, **614**, L25
- Yue, M., Eilers, A.-C., Simcoe, R. A., et al. 2023, *ApJ*, submitted [arXiv:2309.04614]
- Zavala, J. A., Buat, V., Casey, C. M., et al. 2023, *ApJ*, **943**, L9
- Zhang, H., Behroozi, P., Volonteri, M., et al. 2023, *MNRAS*, **518**, 2123

## Appendix A: Comparative analysis of properties between active and inactive galaxies

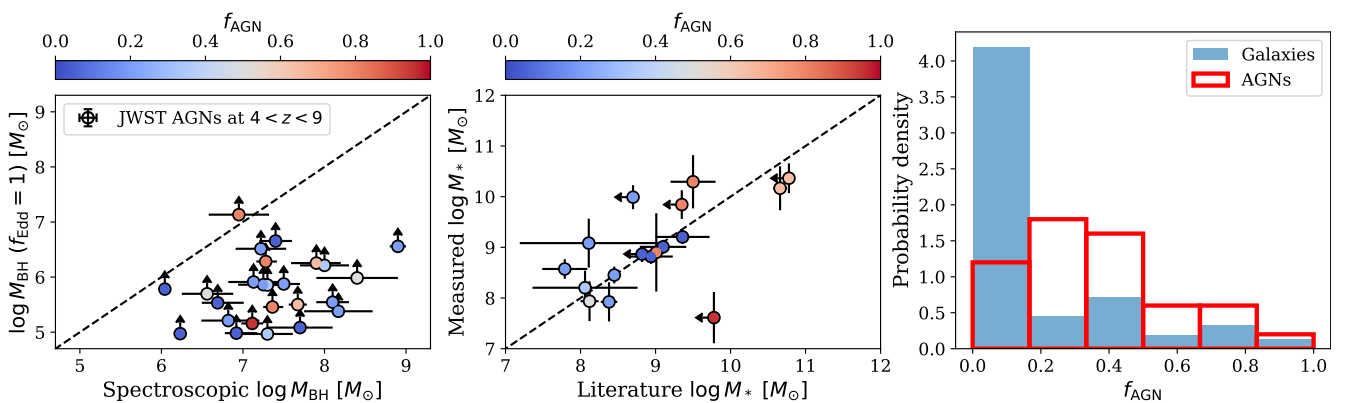
We present the properties of known sources at  $4 < z < 9$  observed with the James Webb Space Telescope (JWST) taken from the literature to investigate how well our selection criteria distinguish between active galactic nuclei (AGNs) and inactive galaxies in real data, as well as to assess the accuracy of the derived physical traits. To construct the samples for comparison, we first select sources with available spectroscopic redshifts in the DAWN JWST Archive’s (DJA; Valentino et al. 2023) version of CEERS and GOODS-S/N photometric tables as well as the second data release of UNCOVER catalog. Confirmed AGNs<sup>10</sup> reported in previous studies are listed in Table A.1, where all of them have been spectroscopically characterized with JWST. In total, there are 36 AGNs with publicly available JWST/NIRCam data (i.e., Harikane et al. 2023; Greene et al. 2024; Larson et al. 2023a; Kocevski et al. 2023; Kokorev et al. 2023), including 6 identified as dual AGN candidates (Maiolino et al. 2023). We note that these dual AGN candidates are then removed from the samples to avoid complications due to their potential peculiar properties, leaving us with the remaining 30 AGNs. To construct the inactive galaxy samples, we subsequently chose 93 sources at  $z = 4–9$  from the DJA’s JWST sources repository, characterized and classified as galaxies not containing AGNs, as defined by their originating publication (i.e., Roberts-Borsani et al. 2022; Morishita et al. 2023; Nakajima et al. 2023; Sanders et al. 2024; Tang et al. 2023; Isobe et al. 2023; Arrabal Haro et al. 2023; Fujimoto et al. 2023c; Heintz et al. 2023; Hainline et al. 2024; Bunker et al. 2023; Saxena et al. 2023).

Spectral energy distribution (SED) fitting with CIGALE is then performed to the sources compiled above to calculate their stellar mass ( $M_*$ ) and AGN fraction of the total emission ( $f_{\text{AGN}}$ ) within the rest-frame wavelengths of  $0.1–0.7 \mu\text{m}$ . The right panel of Fig. A.1 depicts the distribution of  $f_{\text{AGN}}$  for two distinct populations: AGNs versus inactive galaxies. Correspondingly, it is observed that 80% of the known AGNs (24 out of 30) have  $f_{\text{AGN}} \geq 0.2$ , whereas approximately 30% of the galaxies (28 out of 93) display  $f_{\text{AGN}} \geq 0.2$ . This might indicate that if we adopt

$f_{\text{AGN}} \geq 0.2$  as a limit for the quasar candidates selection, we would expect a contamination from the high- $z$  inactive galaxies as high as 30%. Of course, we could increase the  $f_{\text{AGN}}$  cutoff to a higher value to get a more pure quasar samples. However, given the scarcity of quasar number density in the sky, we prefer to adopt 0.2 to aim for a completeness level of up to 80%. Accordingly, of the discussed parent samples, only 11 AGNs reside at  $z \gtrsim 6$ , for which we successfully recover 9 of them using the selection method explained in Sect. 3. Two sources are missed because their  $f_{\text{AGN}}$  is less than our selection threshold of 0.2.

We then proceed to estimate the lower limit black hole masses of the 30 JWST-confirmed AGNs at  $4 < z < 9$  as explained in Section 4.2 by adopting Equation 1 and assuming Eddington ratio of  $f_{\text{Edd}} = 1$ . Accordingly, we compare these limits with the actual  $M_{\text{BH}}$  reported in the literature, determined based on the broad emission line analysis. In this case, only 25 of 30 AGNs have available spectroscopic  $M_{\text{BH}}$  (see Table A.1). As shown by the left panel of Figure A.1, our  $M_{\text{BH}}$  estimations are, on average, systematically lower by  $\approx 1.6$  dex than the spectroscopic  $M_{\text{BH}}$  reported in other studies, consistent with the notion that the  $M_{\text{BH}}$  values that we inferred are really lower limits. This offset is expected since  $f_{\text{Edd}}$  strongly affects our  $M_{\text{BH}}$  estimation. If we change the assumed  $f_{\text{Edd}}$  to be much lower, like 0.1, our data points will become 1 dex higher, closer to those spectroscopic  $M_{\text{BH}}$ .

The comparison between stellar masses ( $M_*$ ) we computed via SED modeling and values from the literature is depicted in the middle panel of Figure A.1. We note that this further excludes samples from Greene et al. (2024) since they do not provide  $M_*$  measurements, leaving us with the remaining 17 AGNs. Our measurements and other studies are reasonably consistent within the expected uncertainties as the data points are positively correlated with a scatter around the one-to-one relation of  $\approx 0.5$  dex. This scatter is expected since we calculated  $M_*$  without performing decomposition of the quasar and host galaxy lights via 2D image modeling. Instead, we performed the SED decomposition directly using CIGALE to the photometric catalog, as explained in Section 3.2.

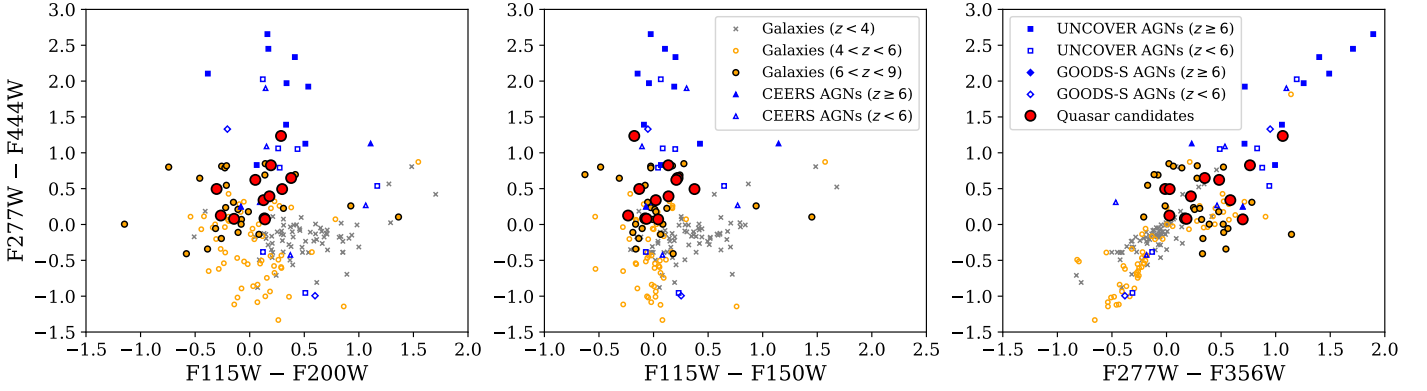


**Fig. A.1.** Distribution of black hole masses ( $M_{\text{BH}}$ ), stellar masses ( $M_*$ ), and the fraction of AGN emission ( $f_{\text{AGN}}$ ) of known sources (see text). The left panel compares the lower limit  $M_{\text{BH}}$  assuming an Eddington ratio of  $f_{\text{Edd}} = 1$  that we calculated and actual values reported in the literature. The data points are color-coded according to the inferred  $f_{\text{AGN}}$  of each source. The middle panel shows the  $M_*$  from other studies versus our own measurements. The right panel illustrates the distribution of  $f_{\text{AGN}}$  for active and inactive galaxies. To compensate for the difference in sample sizes, we normalize the bin heights of the histogram, ensuring that the integral of the distribution equals unity.

<sup>10</sup> For completeness, known bright quasars showcased previously in Figure 6 but not used for analysis in this section since they reside outside the public JWST fields are also listed in Table A.1 (i.e., Yue et al. 2023; Ding et al. 2023; Übler et al. 2023; Stone et al. 2023).

As additional information, we find that distinguishing unobscured AGNs having blue rest-frame UV continuum from galaxies is challenging due to their color similarity in specific filter pairs. To illustrate this better, we present the color distribution of our quasar candidates compared to confirmed AGNs and inactive galaxies in Figure A.2. Since a substantial overlap between the colors of unobscured AGNs and galaxies is observed, employing a more advanced technique, such as full SED fitting as done

here, is more effective than using simple color cuts for accurately identifying these blue quasars. In the future, complementing our current datasets with more mid-infrared (MIR) measurements will be instrumental in differentiating AGNs from star-forming galaxies. This distinction arises from the fact that the presence of hot dust emission in MIR bands is a unique feature not easily attributable to stellar light or cold dust within the interstellar medium.



**Fig. A.2.** JWST/NIRCam color diagram of spectroscopically confirmed sources residing in the CEERS, UNCOVER, and GOODS-S datasets. Galaxy samples from the DJA’s JWST sources repository at low, medium, and high redshifts are marked with gray crosses, orange empty circles, and orange-filled circles, respectively. Broad-line AGNs are indicated with blue colors, where filled symbols denote objects at  $6 < z < 9$ , while empty symbols show those at  $4 < z < 6$  (see the figure legend). On the other hand, samples of our quasar candidates existing in the same extragalactic fields are portrayed with red circles. Substantial overlap between the colors of unobscured AGNs – that is, those with blue rest-frame UV continuum – and galaxies make it challenging to separate them using simple color cuts, indicating that full SED fitting is a better way to recover those blue quasars.

**Table A.1.** Compilation of AGN samples that have been spectroscopically characterized with JWST from the literature.

Source	RA [J2000]	Dec [J2000]	$z_{\text{spec}}$	$\log M_{\text{BH,ref}}$ [ $M_{\odot}$ ]	$\log M_{*,\text{ref}}$ [ $M_{\odot}$ ]	$f_{\text{AGN}}$	$\log M_{\text{BH}}$ [ $M_{\odot}$ ]	$\log M_{*}$ [ $M_{\odot}$ ]	Reference
J0148+0600	177.06933	52.86397	5.98	$9.89^{+0.05}_{-0.06}$	$10.74^{+0.31}_{-0.30}$	...	...	...	Yue et al. (2023)
J159-02	159.22579	-2.54387	6.38	$9.10^{+0.01}_{-0.01}$	$10.14^{+0.34}_{-0.36}$	...	...	...	Yue et al. (2023)
J1120+0641	170.00617	6.69008	7.09	$9.08^{+0.03}_{-0.03}$	$9.81^{+0.23}_{-0.31}$	...	...	...	Yue et al. (2023)
J0100+2802	15.05425	28.04050	6.33	$10.06^{+0.01}_{-0.01}$	$< 11.58$	...	...	...	Yue et al. (2023)
J1030+0524	157.61296	5.41529	6.30	$9.19^{+0.01}_{-0.01}$	$< 10.65$	...	...	...	Yue et al. (2023)
J1148+5251	27.15683	6.00556	5.98	$9.64^{+0.01}_{-0.01}$	$< 10.93$	...	...	...	Yue et al. (2023)
J2236+0032	339.18575	0.54914	6.40	$9.19^{+0.08}_{-0.08}$	$11.12^{+0.40}_{-0.27}$	...	...	...	Ding et al. (2023)
J2255+0251	343.90850	2.85739	6.34	$8.31^{+0.04}_{-0.04}$	$10.53^{+0.51}_{-0.37}$	...	...	...	Ding et al. (2023)
CEERS01244	215.24067	53.03606	4.48	$7.18^{+0.03}_{-0.03}$	$8.63^{+0.63}_{-1.03}$	...	...	...	Harikane et al. (2023)
GLASS160133	3.58029	-30.42439	4.02	$6.04^{+0.04}_{-0.04}$	$< 8.82$	$0.05 \pm 0.20$	$> 5.78$	$8.87 \pm 0.16$	Harikane et al. (2023)
GLASS150029	3.57717	-30.42258	4.58	$6.23^{+0.03}_{-0.05}$	$9.10^{+0.31}_{-0.37}$	$0.05 \pm 0.20$	$> 4.97$	$9.01 \pm 0.14$	Harikane et al. (2023)
CEERS00746	214.80913	52.86847	5.62	$7.43^{+0.11}_{-0.10}$	$< 9.11$	...	...	...	Harikane et al. (2023)
CEERS01665	215.17821	53.05936	4.48	$6.95^{+0.20}_{-0.11}$	$9.92^{+0.51}_{-0.68}$	...	...	...	Harikane et al. (2023)
CEERS00672	214.88967	52.83297	5.67	$7.36^{+0.13}_{-0.11}$	$< 9.01$	$0.80 \pm 0.17$	$> 5.46$	$8.90 \pm 0.77$	Harikane et al. (2023)
CEERS02782	214.82346	52.83028	5.24	$7.28^{+0.14}_{-0.11}$	$< 9.35$	$0.80 \pm 0.28$	$> 6.28$	$9.84 \pm 0.28$	Harikane et al. (2023)
CEERS00397	214.83621	52.88269	6.00	$6.69^{+0.32}_{-0.19}$	$9.36^{+0.36}_{-0.45}$	$0.05 \pm 0.17$	$> 5.53$	$9.21 \pm 0.14$	Harikane et al. (2023)
CEERS00717	215.08142	52.97219	6.94	$7.66^{+0.19}_{-0.14}$	$9.61^{+0.77}_{-1.18}$	...	...	...	Harikane et al. (2023)
CEERS01236	215.14529	52.96728	4.48	$6.92^{+0.26}_{-0.14}$	$8.94^{+0.29}_{-0.54}$	$0.05 \pm 0.24$	$> 4.98$	$8.82 \pm 0.14$	Harikane et al. (2023)
MSAID2008	3.59242	-30.43283	6.74	...	...	$0.20 \pm 0.25$	$> 5.12$	$9.21 \pm 0.41$	Greene et al. (2024)
MSAID4286	3.61920	-30.42327	5.84	$8.00^{+0.30}_{-0.30}$	...	$0.35 \pm 0.08$	$> 6.21$	$11.25 \pm 0.21$	Greene et al. (2024)
MSAID10686	3.55084	-30.40660	5.05	...	...	$0.50 \pm 0.08$	$> 6.07$	$10.61 \pm 0.18$	Greene et al. (2024)
MSAID13123a	3.57983	-30.40157	7.04	$7.30^{+0.20}_{-0.20}$	...	$0.35 \pm 0.06$	$> 5.86$	$10.50 \pm 0.10$	Greene et al. (2024)
MSAID13821	3.62061	-30.39995	6.34	$8.10^{+0.20}_{-0.20}$	...	$0.20 \pm 0.11$	$> 5.54$	$10.12 \pm 0.17$	Greene et al. (2024)
MSAID15383a	3.58353	-30.39668	7.04	...	...	$0.65 \pm 0.01$	$> 5.95$	$10.43 \pm 0.11$	Greene et al. (2024)



Table A.1. Continued.

Source	RA [J2000]	Dec [J2000]	$z_{\text{spec}}$	$\log M_{\text{BH,ref}}$ [ $M_{\odot}$ ]	$\log M_{*,\text{ref}}$ [ $M_{\odot}$ ]	$f_{\text{AGN}}$	$\log M_{\text{BH}}$ [ $M_{\odot}$ ]	$\log M_*$ [ $M_{\odot}$ ]	Reference
MSAID16594a	3.59720	-30.39433	7.04	...	...	$0.50 \pm 0.06$	$> 5.61$	$10.24 \pm 0.14$	Greene et al. (2024)
MSAID23608	3.54282	-30.38065	5.80	$7.50^{+0.20}_{-0.20}$	...	$0.20 \pm 0.16$	$> 5.87$	$9.34 \pm 0.18$	Greene et al. (2024)
MSAID28876	3.56960	-30.37322	7.04	...	...	$0.50 \pm 0.11$	$> 5.54$	$10.1 \pm 0.20$	Greene et al. (2024)
MSAID32265	3.53753	-30.37017	...	...	...	...	...	...	Greene et al. (2024)
MSAID33437	3.54642	-30.36625	...	...	...	...	...	...	Greene et al. (2024)
MSAID35488	3.57898	-30.36260	6.26	$7.40^{+0.20}_{-0.20}$	...	$0.05 \pm 0.17$	$> 6.66$	$9.51 \pm 0.12$	Greene et al. (2024)
MSAID38108	3.53001	-30.35801	4.96	$8.40^{+0.50}_{-0.50}$	...	$0.50 \pm 0.07$	$> 5.98$	$10.67 \pm 0.12$	Greene et al. (2024)
MSAID39243	3.51389	-30.35602	...	...	...	...	...	...	Greene et al. (2024)
MSAID41225	3.53399	-30.35331	6.76	$7.70^{+0.40}_{-0.40}$	...	$0.05 \pm 0.18$	$> 5.08$	$9.69 \pm 0.21$	Greene et al. (2024)
MSAID45924	3.58476	-30.34363	4.46	$8.90^{+0.10}_{-0.10}$	...	$0.20 \pm 0.01$	$> 6.56$	$11.22 \pm 0.02$	Greene et al. (2024)
CEERS1019	215.03539	52.89066	8.68	$6.95^{+0.37}_{-0.37}$	$9.50^{+0.30}_{-0.30}$	$0.80 \pm 0.26$	$> 7.13$	$10.29 \pm 0.53$	Larson et al. (2023a)
GS3073	53.07888	-27.88416	5.55	$8.20^{+0.40}_{-0.40}$	$9.40^{+0.7}_{-0.20}$	...	...	...	Übler et al. (2023)
1670	214.82345	52.83028	5.24	$7.11^{+0.13}_{-0.13}$	$< 9.78$	$0.95 \pm 0.30$	$> 5.16$	$7.61 \pm 0.51$	Kocevski et al. (2023)
3210AV4	214.80914	52.86848	5.62	$7.67^{+0.11}_{-0.11}$	$< 10.78$	$0.65 \pm 0.11$	$> 5.50$	$10.36 \pm 0.29$	Kocevski et al. (2023)
J2239+0207	339.94779	2.12986	6.25	$8.78^{+0.4}_{-0.4}$	$10.00^{+0.30}_{-0.50}$	...	...	...	Stone et al. (2023)
10013704a	53.12654	-27.81809	5.92	$5.65^{+0.31}_{-0.31}$	$8.88^{+0.66}_{-0.66}$	...	...	...	Maiolino et al. (2023)
10013704b	53.12654	-27.81809	5.92	$7.50^{+0.31}_{-0.31}$	$8.88^{+0.66}_{-0.66}$	...	...	...	Maiolino et al. (2023)
8083	53.13284	-27.80186	4.65	$7.25^{+0.31}_{-0.31}$	$8.45^{+0.03}_{-0.15}$	$0.20 \pm 0.24$	$> 5.86$	$8.45 \pm 0.17$	Maiolino et al. (2023)
1093	189.17974	62.22463	5.60	$7.36^{+0.31}_{-0.31}$	$8.34^{+0.20}_{-0.20}$	...	...	...	Maiolino et al. (2023)
3608	189.11794	62.23552	5.27	$6.82^{+0.38}_{-0.33}$	$8.38^{+0.11}_{-0.15}$	$0.20 \pm 0.31$	$> 5.21$	$7.92 \pm 0.39$	Maiolino et al. (2023)
11836	189.22059	62.26368	4.41	$7.13^{+0.31}_{-0.31}$	$7.79^{+0.30}_{-0.30}$	$0.20 \pm 0.22$	$> 5.91$	$8.57 \pm 0.19$	Maiolino et al. (2023)
20621	189.12252	62.29285	4.68	$7.30^{+0.31}_{-0.31}$	$8.06^{+0.7}_{-0.7}$	$0.35 \pm 0.33$	$> 4.97$	$8.20 \pm 0.34$	Maiolino et al. (2023)
73488a	189.19740	62.17723	4.13	$6.18^{+0.30}_{-0.30}$	$9.78^{+0.20}_{-0.20}$	...	...	...	Maiolino et al. (2023)
73488b	189.19740	62.17723	4.13	$7.71^{+0.30}_{-0.30}$	$9.78^{+0.20}_{-0.20}$	...	...	...	Maiolino et al. (2023)
77652	189.29323	62.19900	5.23	$6.86^{+0.35}_{-0.34}$	$7.87^{+0.16}_{-0.28}$	...	...	...	Maiolino et al. (2023)
61888	189.16802	62.21701	5.88	$7.22^{+0.31}_{-0.31}$	$8.11^{+0.92}_{-0.92}$	$0.20 \pm 0.27$	$> 6.51$	$9.08 \pm 0.48$	Maiolino et al. (2023)
62309	189.24898	62.21835	5.17	$6.56^{+0.32}_{-0.31}$	$8.12^{+0.12}_{-0.13}$	$0.50 \pm 0.3$	$> 5.7$	$7.94 \pm 0.39$	Maiolino et al. (2023)
53757a	189.26978	62.19421	4.45	$6.29^{+0.33}_{-0.32}$	$10.18^{+0.13}_{-0.12}$	...	...	...	Maiolino et al. (2023)
53757b	189.26978	62.19421	4.45	$7.69^{+0.32}_{-0.31}$	$10.18^{+0.13}_{-0.12}$	...	...	...	Maiolino et al. (2023)
954	189.15197	62.25964	6.76	$7.90^{+0.30}_{-0.31}$	$10.66^{+0.09}_{-0.1}$	$0.65 \pm 0.2$	$> 6.25$	$10.17 \pm 0.43$	Maiolino et al. (2023)
MSAID20466	3.64041	-30.38644	8.50	$8.17^{+0.42}_{-0.42}$	$< 8.70$	$0.20 \pm 0.12$	$> 5.38$	$9.99 \pm 0.24$	Kokorev et al. (2023)

**Notes.** Column (1): name or identifier of each source. Column (2)–(3): right ascension (RA) and declination (Dec) in decimal degrees. Column (4): spectroscopic redshift. Column (5): fraction of AGN component to the total emission within the rest-wavelengths of 0.1–0.7  $\mu\text{m}$ . Column (6): lower limit black hole mass assuming an accretion at Eddington limit. Column (7): total stellar mass calculated following the method presented in this work. Column (8): original literature describing the object. Empty columns indicate that the data is unavailable in the public JWST datasets, and the corresponding source is not used to benchmark our quasar selection method.

**Appendix B: Complete list of the quasar candidates**

The complete list of our high- $z$  quasar candidates chosen following the method explained in the main text is reported here.

Table B.1 summarizes the subset of photometric properties for these sources, and the complete catalog is available at the CDS. Additionally, figures containing the SED fitting results for each quasar candidate can be provided upon reasonable request.

**Table B.1.** List of high- $z$  quasar candidates selected in this work. Here, we only show the first ten rows of the catalog as an example, while the entire catalog can be accessed at the CDS (see text).

Source	RA [J2000]	Dec [J2000]	F444W [nJy]	$z_{\text{phot}}$	$\log L_{\text{bol}}$ [erg s $^{-1}$ ]	$\log M_*$ [ $M_{\odot}$ ]	$\log \text{SFR}$ [ $M_{\odot} \text{ yr}^{-1}$ ]	$\log M_{\text{BH}}$ [ $M_{\odot}$ ]	$f_{\text{AGN}}$	Grade
CWB-663	149.75099	2.15091	351.86 $\pm$ 10.87	6.05 $\pm$ 0.12	43.66 $\pm$ 0.90	9.13 $\pm$ 0.31	0.83 $\pm$ 0.35	> 5.55	0.05 $\pm$ 0.35	B
CWB-8286	149.76383	2.19782	203.96 $\pm$ 3.92	7.81 $\pm$ 0.30	43.58 $\pm$ 1.54	9.35 $\pm$ 0.25	1.03 $\pm$ 0.22	> 5.46	0.05 $\pm$ 0.32	B
CWB-19858	149.84033	2.24807	395.54 $\pm$ 3.92	6.70 $\pm$ 0.13	43.66 $\pm$ 1.31	9.38 $\pm$ 0.16	1.08 $\pm$ 0.15	> 5.54	0.05 $\pm$ 0.19	B
CWB-24983	149.85942	2.27450	234.01 $\pm$ 3.98	6.14 $\pm$ 0.17	44.01 $\pm$ 0.67	9.12 $\pm$ 0.26	0.53 $\pm$ 0.91	> 5.89	0.20 $\pm$ 0.27	A
CWB-26445	149.80063	2.30474	183.36 $\pm$ 3.92	6.93 $\pm$ 0.27	44.81 $\pm$ 0.82	9.19 $\pm$ 0.44	0.87 $\pm$ 0.45	> 6.70	0.95 $\pm$ 0.33	A
CWB-35877	149.95743	2.11303	129.15 $\pm$ 3.92	6.88 $\pm$ 0.22	44.76 $\pm$ 0.49	8.72 $\pm$ 0.46	0.44 $\pm$ 0.47	> 6.65	0.95 $\pm$ 0.28	A
CWB-40773	149.92404	2.15948	253.99 $\pm$ 6.37	7.30 $\pm$ 0.20	44.93 $\pm$ 0.32	8.88 $\pm$ 0.39	0.62 $\pm$ 0.42	> 6.82	0.80 $\pm$ 0.29	A
CWB-41512	149.93536	2.16019	836.82 $\pm$ 5.79	6.02 $\pm$ 0.12	45.22 $\pm$ 0.29	9.75 $\pm$ 0.15	1.46 $\pm$ 0.16	> 7.11	0.20 $\pm$ 0.23	A
CWB-42214	149.93324	2.16687	1760.68 $\pm$ 5.20	6.09 $\pm$ 0.12	44.88 $\pm$ 0.79	10.21 $\pm$ 0.38	1.67 $\pm$ 0.34	> 6.77	0.50 $\pm$ 0.17	A
CWB-43536	150.02278	2.14372	545.53 $\pm$ 11.98	8.10 $\pm$ 0.19	45.29 $\pm$ 0.59	9.83 $\pm$ 0.25	1.52 $\pm$ 0.25	> 7.17	0.20 $\pm$ 0.27	A

**Notes.** Column (1): name of each candidate with specific prefixes indicating the originating dataset, that is, CWB (COSMOS-Web), JDS/GDS (JADES/GOODS-S), GDN (GOODS-N), UCV (UNCOVER), CRS (CEERS), PMC (PRIMER-COSMOS), and PMU (PRIMER-UDS). Column (2)–(3): right ascension (RA) and declination (Dec) in decimal degrees. Column (4): fluxes measured using the JWST/NIRCam images. Column (5): calculated photometric redshift of the target derived from the best-fitted SED template. Column (6): bolometric luminosity of the AGN SED component. Column (7): total stellar mass of the presumed host galaxy. Column (8): SFR averaged over 100 Myr. Column (9): lower limit of the black hole mass, assuming an accretion at the Eddington limit. Column (10): fraction of AGN component to the total spectral emission within the rest-frame wavelengths of 0.1–0.7  $\mu\text{m}$ . Column (11): grade after employing the visual inspection, black hole mass limit, and AGN fraction threshold criteria.