Testing the evolution of correlations between supermassive black holes and their host galaxies using eight strongly lensed quasars

Xuheng Ding [1],[★] Tommaso Treu,[1][★] Simon Birrer,[2] Adriano Agnello [3] Dominique Sluse,[4] Chris Fassnacht,[5] Matthew W. Auger,[6] Kenneth C. Wong [7] Sherry H. Suyu,[8,9,10] Takahiro Morishita,[11] Cristian E. Rusu [5,12,13] and Aymeric Galan[14]

[1]*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA*
[2]*Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, Stanford, CA 94305, USA*
[3]*European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany*
[4]*STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium*
[5]*Department of Physics and Astronomy, University of California Davis, 1 Shields Avenue, Davis, CA 95616, USA*
[6]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*
[7]*Kavli IPMU (WPI), UTIAS, University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[8]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*
[9]*Physik-Department, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany*
[10]*Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), 11F of ASMAB, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan*
[11]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[12]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[13]*Subaru Telescope, National Astronomical Observatory of Japan, 650 N Aohoku Pl, Hilo, HI 96720*
[14]*Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland*

ABSTRACT

One of the main challenges in using high-redshift active galactic nuclei (AGNs) to study the correlations between the mass of a supermassive black hole ($\mathcal{M}_{BH}$) and the properties of its active host galaxy is instrumental resolution. Strong lensing magnification effectively increases instrumental resolution and thus helps to address this challenge. In this work, we study eight strongly lensed AGNs with deep *Hubble Space Telescope* imaging, using the lens modelling code LENSTRONOMY to reconstruct the image of the source. Using the reconstructed brightness of the host galaxy, we infer the host galaxy stellar mass based on stellar population models. $\mathcal{M}_{BH}$ are estimated from broad emission lines using standard methods. Our results are in good agreement with recent work based on non-lensed AGNs, demonstrating the potential of using strongly lensed AGNs to extend the study of the correlations to higher redshifts. At the moment, the sample size of lensed AGNs is small and thus they provide mostly a consistency check on systematic errors related to resolution for non-lensed AGNs. However, the number of known lensed AGNs is expected to increase dramatically in the next few years, through dedicated searches in ground- and space-based wide-field surveys, and they may become a key diagnostic of black holes and galaxy co-evolution.

Key words: gravitational lensing: strong – galaxies: active – galaxies: evolution.

1 INTRODUCTION

The tight correlations between the masses ($\mathcal{M}_{BH}$) of supermassive black holes (BHs) and the properties of their host galaxies, including stellar mass M_*, stellar velocity dispersion σ_* and luminosity L_{host}, known as scaling relations, are usually considered as a result of their co-evolution (e.g. Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2001; Marconi & Hunt 2003; Häring & Rix 2004; Gültekin et al. 2009; Beifiori et al. 2012; Graham et al. 2011). Cosmological simulations of structure formation are able to reproduce the local tight relation based on the physical mechanism by invoking active galactic nucleus (AGN) feedback as the physical

connection (Springel et al. 2005; Di Matteo et al. 2008; Hopkins et al. 2008; DeGraf et al. 2015) or having them share a common gas supply (Cen 2015; Menci et al. 2016). However, it has also been suggested that the correlations arise statistically, without any physical coupling, as a result of stochastic mergers (Peng 2007; Hirschmann et al. 2010; Jahnke & Macciò 2011).

A powerful way to understand the origin of the correlations is to study them as a function of redshift, determining how and when they emerge and evolve over cosmic time (e.g. Treu, Malkan & Blandford 2004; Salviander et al. 2006; Woo et al. 2006; Jahnke et al. 2009; Schramm & Silverman 2013; Park et al. 2015; Sun et al. 2015). Recently, based on a sample of 32 X-ray-selected type-1 AGNs in deep survey fields, Ding et al. (2020b, hereafter D20) measured the scaling relations in the redshift range $1.2 < z < 1.7$ using multicolour Wide Field Camera 3 (WFC3) *Hubble Space Telescope*

★ E-mail: dingxuheng@mail.bnu.edu.cn (XD); tt@astro.ucla.edu (TT)

Table 1. Summary of observational details (also given by Ding et al. 2017a) for the lensed AGNs. The references are: (1) Wisotzki et al. (2002); (2) Sluse et al. (2012); (3) Sluse et al. (2003); (4) Sluse et al. (2007); (5) Morgan et al. (2004); (6) Oguri et al. (2005); (7) Eulaers et al. (2013); (8) Wisotzki et al. (1993); (9) Smette et al. (1995); (10) Inada et al. (2005); (11) Eigenbrod, Courbin & Meylan (2007); (12) Hagen, Engels & Reimers (1999); (13) Chantry, Sluse & Magain (2010); (14) Wisotzki et al. (2004); (15) Ofek et al. (2006).

Object ID	z_s	Camera	Filter	Exposure time (s)	Programme ID	PI	Pixel scale (drizzled)	References
HE0435–1223	1.69	WFC3-IR	F160W	9340	12889	S. H. Suyu	0″.08	(1), (2)
RXJ1131–1231	0.65	ACS	F814W	1980	9744	C.S. Kochanek	0″.05	(3), (4)
WFI2033–4723	1.66	WFC3-IR	F160W	26257	12889	S. H. Suyu	0″.08	(5), (2)
SDSS1206+4332	1.79	WFC3-IR	F160W	8457	14254	T. Treu	0″.08	(6), (7)
HE1104–1805	2.32	WFC3-IR	F160W	14698	12889	S. H. Suyu	0″.08	(8), (9)
SDSS0246–0825	1.68	WFC3-UVIS	F814W	8481	14254	T. Treu	0″.03	(10), (11)
HS2209+1914[a]	1.07	WFC3-UVIS	F814W	9696 + 4542	14254	T. Treu	0″.03	(12), (13)
HE0047–1756	1.66	WFC3-UVIS	F814W	9712	14254	T. Treu	0″.03	(14), (15)

[a] HS2209 was visited by the *HST* twice (*vis05* and *vis06*) at different orientations. The exposure time is thus given separately.

(*HST*) imaging. Combining the new sample with published samples in both local and intermediate (i.e. $0.35 < z \lesssim 1.2$) redshift ranges, D20 strengthen the support for an evolution scenario in which the growth of BH evolution pre-dates that of the bulge component of the host galaxy. In a follow-up paper, Ding et al. (2020a) compared the D20 measurements to the predictions by the numerical simulations including the hydrodynamic simulation MassiveBlackII (Khandai et al. 2015) and a semi-analytical model (Menci et al. 2014). The observed tightness of the scaling relations at high redshift is consistent with the hypothesis that AGN feedback drives the scaling correlations, and disfavours the hypothesis of the correlations being purely stochastic in nature.

As the samples of AGNs grow to increase statistical power, it is very important to make sure that systematics do not dominate the error budget. One of the main potential sources of systematics is the finite resolution of *HST* images. Even with modern techniques, AGN hosts remain barely resolved at *HST* resolution, and therefore it would be very useful to verify the results at higher resolution. This is the goal of this work.

It has been long recognized that lensed AGNs, by virtue of magnification, can provide unique insights into the scaling relations of the distant Universe (Peng et al. 2006), provided that the lens modelling can be accurately derived. With the aim of verifying the fidelity of the modern lens modelling technique, Ding et al. (2017a) carried out extensive and realistic simulations tests based on the deep *HST* observations for a sample of eight lens systems in the project, H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW;[1] Suyu et al. 2017). They confirm that the reconstruction of the lensed host galaxy properties can be recovered with better precision and accuracy than the typical $\mathcal{M}_{BH}$ uncertainty. Then, Ding et al. (2017b) applied the advanced techniques to two strongly lensed systems analysed by the H0LiCOW collaboration (Suyu et al. 2013; Wong et al. 2017) to study their $\mathcal{M}_{BH}$–L_{host} relations. They obtained consistent scaling relations compared with the samples from the literature.

In this work, we expand the measurements of the $\mathcal{M}_{BH}$–M_* relation using the full sample of eight lensed AGNs introduced by Ding et al. (2017a). In order to take advantage of the excellent quality of the data, we develop an independent approach to achieve a one-step inference of the host galaxy photometry from the eight lensed AGNs. We adopt a set of extended modelling choices to estimate the uncertainty level of the host property and to ensure the accuracy of our measurements. We compare the inference of our new measurements

to the ones that have been modelled by the H0LiCOW collaboration to make a cross-check. To obtain an accurate inference of stellar mass, we utilize the multiband imaging data taken with the *HST* to obtain the colour information for three of our eight targets. We assume a typical stellar population for the other five AGNs of the sample, consistent with D20. Furthermore, we adopt a class of self-consistent recipes to recalibrate the $\mathcal{M}_{BH}$ of our sample, in a manner consistent with D20. Given the similar redshift range, the high data quality and consistent techniques, this sample of lensed AGNs provides an excellent validation of the D20 results.

This paper is structured as follows. In Section 2, we introduce the sample including imaging data and BH masses. In Section 3, we describe our approach designed to infer the lens models and reconstruct the host galaxy. We use the inferred photometry to derive the stellar mass and the scaling relations and we combine with the D20 sample to study the evolution in Section 4. We draw conclusions in Section 5 and we discuss the results further in Appendix B. Throughout this paper, we adopt a standard concordance cosmology with parameters adopted as $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.30$ and $\Omega_\Lambda = 0.70$, to compute the luminosity distance and to estimate the host's absolute brightness. Magnitudes are presented in the AB system. A Chabrier initial mass function (IMF) is employed for the sample, to be consistent with D20.

2 SAMPLE SELECTION AND BLACK HOLE MASS ESTIMATES

We adopt eight lens systems from the H0LiCOW collaboration including HE0435–1223, RXJ1131–1231, WFI2033–4723, HE1104–1805, SDSS1206+4332, SDSS0246–0825, HE0047–1756 and HS2209+1914. We refer to Suyu et al. (2017) and Ding et al. (2017a) for the descriptions of these data. For conciseness, in the rest of this paper, we abbreviate each lens name to four digits (e.g. HE0435–1223 to HE0435).

Based on the observational data of these eight systems, Ding et al. (2017a) performed an extensive and realistic simulation exercise using the *HST* image data and confirmed that the source reconstruction using the lens modelling technique is trustworthy. We summarize the information for the eight systems, including their redshift, data properties and references in Table 1. Besides the imaging data shown in this table, we also analysed the multiband *HST* imaging data to derive the host colour information for three of the eight systems. As we show in Section 4.1, we use the colour information to fit for the best stellar population to improve the accuracy of the estimate of M_*.

The sample of lensed AGNs is too limited in size to constrain evolution by itself. Therefore, we use it primarily to verify the results of D20. The D20 sample includes 32 AGN measurements in the redshift range of $1.2 < z < 1.9$. They also collected 59 intermediate redshift (i.e. $0.35 < z \lesssim 1.2$) AGN measurements (Bennert et al. 2011b; Cisternas et al. 2011; Schramm & Silverman 2013) and 55 local (i.e. $z \lesssim 0.007$) measurements (Häring & Rix 2004; Bennert et al. 2011a). It is worth noting that this is, so far, the largest *HST* imaging AGN sample with redshift range up to $z \sim 1.9$.

To ensure the consistency of $\mathcal{M}_{BH}$ estimates based on different broad lines, we adopt the following set of self-consistent recipes following D20:

$$\log \left(\frac{\mathcal{M}_{BH}}{M_\odot} \right) = a + b \log \left(\frac{L_{\lambda_{line}}}{10^{44} \, \text{erg s}^{-1}} \right)$$
$$+ 2 \log \left[\frac{\text{FWHM(line)}}{1000 \, \text{km s}^{-1}} \right], \tag{1}$$

where λ_{line} is the reference wavelength of the local continuum luminosities for different emission lines. The following values are adopted: $a\{\text{C IV, Mg II, H}\beta\} = \{6.322, 6.623, 6.910\}$, $b\{\text{C IV, Mg II, H}\beta\} = \{0.53, 0.47, 0.50\}$, $\lambda_{line}\{\text{C IV, Mg II, H}\beta\} = \{1350, 3000, 5100\}$ (Å). The broad line properties of our samples are adopted from the literature, with a few corrections/exceptions. The line properties of SDSS1206 have been inferred by Shen et al. (2011). However, this system was investigated as a non-lensed AGN and the lensing magnification on the intrinsic continuum luminosity was not considered. We follow Birrer et al. (2019) and apply the same magnification correction on the $\log (L_\lambda)$ in this work. For the rest of the lens sample except HS2209, their broad line properties are adopted from the literature (Sluse et al. 2012; Peng et al. 2006) taking into account lensing magnification. For HS2209, the spectrum was observed at the Keck-II Telescope in September 2015. We derived the line properties from the spectrum using the same approach as Sluse et al. (2012), based on Mg II. The $\mathcal{M}_{BH}$ measurements, together with the properties of the broad line, are listed in Table 2. The uncertainty of the $\mathcal{M}_{BH}$ is estimated to be 0.4 dex. We note that the magnification correction for these systems are not fully self-consistent, in the sense that the magnification is slightly different for the host galaxy light and the BH mass. This difference could in principle introduce some systematics on the correlations. However, these errors are smaller than the calibration error on $\mathcal{M}_{BH}(0.4 \text{ dex})$, given that the virial relations depend on roughly the square root of the $\log (L_\lambda)$.

3 SURFACE PHOTOMETRY INFERENCE

In this section, we describe how we derived surface photometry of the lensed galaxies by taking lensing effects into account. Lens models of four of the eight systems, namely HE0435 (Wong et al. 2017), RXJ1131 (Suyu et al. 2013), WFI2033 (Rusu et al. 2019) and SDSS1206 (Birrer et al. 2019), have been published by the H0LiCOW project. The goal of those models was strong lens time-delay cosmography (Refsdal 1966; Treu & Marshall 2016). The reconstruction of the host galaxy's light was a byproduct – via pixel-lated distribution (Suyu et al. 2006) or shapelets (Refregier 2003). Ding et al. (2017b) used those reconstructions for two of the systems (HE0435 and RXJ1131) and then fitted a simply parametrized surface brightness profile to the reconstruction to measure the host properties as the non-lensed AGN case.

In this work, in order to reproduce more closely and uniformly what is done in non-lensed AGNs, we develop a strategy to obtain a one-step measurement of the host galaxy's light described by

Table 2. Summary of $\mathcal{M}_{BH}$ estimates and broad line properties, based on equation (1). The references are: (1) Sluse et al. (2012); (2) Shen et al. (2011); (3) Peng et al. (2006).

Object ID	Line(s)	FWHM (km s^{-1})	$\log(L_\lambda)$ (erg s^{-1})	$\log \mathcal{M}_{BH}$ (M$_\odot$)	Ref.
HE0435	Mg II	4930	45.14	8.54	(1)
RXJ1131	Mg II/Hβ	5630/4545	44.29/44.02	8.26/8.23	(1)
WFI2033	Mg II	3960	45.19	8.38	(1)
SDSS1206	Mg II	5632	45.01	8.60	(2)
HE1104	C IV	6004	46.18	9.03	(3)
SDSS0246	Mg II	3700	45.19	8.32	(1)
HS2209	Mg II	3245	45.71	8.45	This paper
HE0047	Mg II	4145	45.59	8.60	(1)

a Sérsic surface brightness profile. For four of the eight systems already studied by the H0LiCOW project, we make a comparison to characterize systematic uncertainties related to the modelling techniques.

3.1 Data preparation and modelling set-up

We follow the standard procedure as described by D20 to prepare the fitting ingredients including the lensing imaging data, noise level map and the point spread function (PSF) information. The imaging data are first drizzled to a higher resolution with a Gaussian kernel; the adopted resolutions are listed in Table 1. We then adopt the PHOTUTILS (Bradley et al. 2016) Python package to model the global background light in two dimensions, based on the SEXTRACTOR algorithm. We remove the background light and cut the clear image data into a postage stamp at a suitable size. We draw the image mask for each system to define the region in which the pixels will be used to calculate the likelihood; see the top-middle panel in each subplot in Fig. A1.

We carry out a forward modelling process to simultaneously constrain the lens model, subtract the central AGN light, and infer the photometry of the host galaxy. For any extended objects, including the lensing galaxy and source galaxy, we assume their surface brightness can be described by the two-dimensional elliptical Sérsic profile. We start with a single Sérsic profile and consider using two Sérsic profiles if any significant residual indicates multiple components. For a single Sérsic profile, we set the prior of the Sérsic index value n in the range [0.5–5.0] to avoid unphysical results.[2] The bright nucleus is unresolved and modelled by a scaled PSF in the image plane. We also impose that the AGN and its host galaxy have the same centre. Following standard practice that has been shown to produce good models (e.g. Treu 2010, and references therein), we adopt elliptical power-law density profiles to define the surface mass density of the deflector, with an external shear.

We employ the imaging modelling tool LENSTRONOMY[3] (Birrer & Amara 2018) to perform the fitting task, using the 'particle swarm optimizer' mode. Building on D20, we adopt a set of modelling choices and fit each system multiple times. Then, we perform a statistical analysis of the measurements, and apply a weighting algorithm to derive the final inference and estimate the uncertainty level. The modelling choices that we consider include the following.

---

[2]It has been shown in Ding et al. (2017a) that choosing this prior for the Sérsic index n yields unbiased host magnitude inferences.
[3]https://github.com/sibirrer/lenstronomy

Table 3. Summary of inference of the lensed AGN host galaxy properties. Columns 2–6 show photometry derived using the imaging data listed in Table 1. Column 7 gives the adopted age of the stellar population with solar metallicity. Column 8 gives theinferred stellar mass. The uncertainty of M_* is estimated to be 0.2 dex.

Object ID (1)	Intrinsic magnitude (source plane) (2)	Magnitude (image plane) (3)	Host flux ratio (to total) (%, total = host + AGN) (4)	$R_{\rm eff}$ (arcsec) (5)	Sérsic n (6)	Stellar population age (Gyr) (7)	$\log(M_*)$ ($M_\odot$) (8)
HE0435	$21.49^{+0.40}_{-0.29}$	$18.58^{+0.30}_{-0.23}$	36.0 ± 11.1	0.28 ± 0.02	2.71 ± 0.20	1.50	10.91
RXJ1131$_{\rm bulge}$	$21.80^{+0.23}_{-0.19}$	$18.70^{+0.07}_{-0.06}$	7.1 ± 1.4	0.13 ± 0.02	fix to 4	3.00	10.39
RXJ1131$_{\rm disc}$	$19.33^{+0.17}_{-0.15}$	$17.14^{+0.08}_{-0.07}$	69.2 ± 10.1	0.90 ± 0.06	fix to 1	1.50	11.08
WFI2033	$21.78^{+0.28}_{-0.23}$	$19.07^{+0.35}_{-0.26}$	19.6 ± 4.5	0.28 ± 0.02	0.52 ± 0.01	0.625	10.51
SDSS1206	$21.31^{+0.23}_{-0.19}$	$18.30^{+0.05}_{-0.05}$	33.3 ± 6.4	0.11 ± 0.02	4.57 ± 0.53	0.625	10.77
HE1104	$21.25^{+0.16}_{-0.14}$	$19.16^{+0.02}_{-0.05}$	14.0 ± 2.0	0.27 ± 0.02	1.05 ± 0.04	0.625	11.05
SDSS0246	$23.44^{+0.28}_{-0.22}$	$20.85^{+0.08}_{-0.07}$	4.0 ± 0.9	0.44 ± 0.08	4.96 ± 0.08	0.625	10.75
HS2209	$20.72^{+0.26}_{-0.21}$	$19.20^{+0.04}_{-0.04}$	12.5 ± 2.7	1.96 ± 1.28	3.15 ± 1.40	1.00	11.04
HE0047	$22.92^{+0.48}_{-0.33}$	$20.37^{+0.20}_{-0.17}$	2.3 ± 0.8	0.32 ± 0.15	4.18 ± 0.75	0.625	10.91

(i) Following common practice, we select all the bright, isolated stars across the image frame of targets to define the PSF. Each selected star is considered as an initial PSF estimate for the fit.

(ii) The central pixels of the AGNs are very bright and can be affected by large systematic errors during the interpolation of the subsampled PSF. To avoid overfitting the noise, we adopt two different modelling options: noise boost (i.e. manually boosting the noise estimate in the central area to effectively infinite); PSF iteration (i.e. performing the iterative PSF estimation as introduced by Chen et al. 2016 and Birrer et al. 2019).

(iii) To calculate the ray tracing under a higher resolution grid relative to the pixel sizes in the image plane, we choose to oversample the model by a factor 2×2 or 3×3 pixel^{-1}.

(iv) Using the lensing imaging alone, it is difficult to constrain the power-law slope, especially with our simplified model of the host's light distribution. To mitigate the overfitting of the slope value, we repeat the fit for three values [1.9, 2.0, 2.1] that are meant to bracket the range observed in lens galaxies (e.g. Koopmans et al. 2006; Auger et al. 2010).

In general, for one lens system with a number of N initial PSF estimates, we perform in total $N \times 12$ fits; that is, 2 by (ii) $\times$ 2 by (iii) $\times$ 3 by (iv). After all fits are completed, we rank their performance based on their best-fitting χ^2 value. As there is no evidence of better performance for the options between noise boost and PSF iteration, we selected the top four fittings from each of them and then combined their best-fitting results based on the weighting algorithm introduced by D20. The degrees of freedom are the same for each of the models, ensuring that the weighting scheme is internally self-consistent. The weights are defined by

$$w_i = \exp\left[-\alpha \frac{(\chi_i^2 - \chi_{\rm best}^2)}{2\chi_{\rm best}^2}\right], \tag{2}$$

where α is an inflation parameter[4] so that, when $i = 4$,

$$\alpha \frac{\chi_{i=4}^2 - \chi_{\rm best}^2}{2\chi_{\rm best}^2} = 2. \tag{3}$$

Then, the results for noise boost and PSF iteration are combined equally to derive the value of host properties and the root-mean-

---

[4]Defining α as the inflation parameter, it is assumed that it is larger than 1, so as to err on the side of caution and include more choices that would be allowed strictly by statistical noise considerations.

square error (i.e. σ level) based on the weights of the 4+4 (i.e. noise boost and PSF iteration) options by

$$\bar{x} = \frac{\sum_{i=1}^N x_i * w_i}{\sum w_i} \tag{4}$$

and

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 * w_i}{\sum w_i}}. \tag{5}$$

Our approach uses the relative goodness of fit and ensures that at least eight sets of best fits are used to estimate the range of systematic uncertainties. Note that the slope of the mass profile of the deflector is fixed in each fit, and that the statistical uncertainty is much smaller than the systematic uncertainty.

The inferred photometric properties of the host galaxy for all eight systems are listed in Table 3, columns 2−6. Detailed information on the fitting process for each system is given in Appendix A.

4 RESULTS

In this section, we describe the approaches and assumptions used to estimate the stellar populations in our sample. Then, we adopt stellar population templates to derive the stellar mass. For this step, we infer the colour information for the multiband imaging data taken with the *HST*. We study the $\mathcal{M}_{\rm BH}$–M_* relations and compare our measurements with those taken from the literature.

4.1 Stellar population and mass

Besides the imaging data analysed in the last section, some of the systems have also been imaged by the *HST* through other bands, providing colour information. Given that we have used the highest signal-to-noise ratio data for our primary models, the analysis of the other bands has lower fidelity and we use it only to infer the colours so as to assist in the estimation of the stellar mass (i.e. Table 3, column 2).

HE0435. In addition to WFC3/F160W, HE0435 has been observed through bands ACS/F814W and ACS/F555W (GO-9744; PI: C. S. Kochanek). Our aim is to derive the brightness of the host galaxy through all the three bands to investigate the colour and stellar population in the image plane. We use the modelling approach introduced in Appendix A to perform this inference for the F814W and F555W bands. To save computer time, we only use the PSF iteration approach. We also fix the lens mass slope value to 1.9 as this is closer to the inference by Wong et al. (2017; i.e. $\gamma \sim 1.93$).

(a) HE0435 F555W

(b) HE0435 F814W

(c) Host galaxy stellar template inference
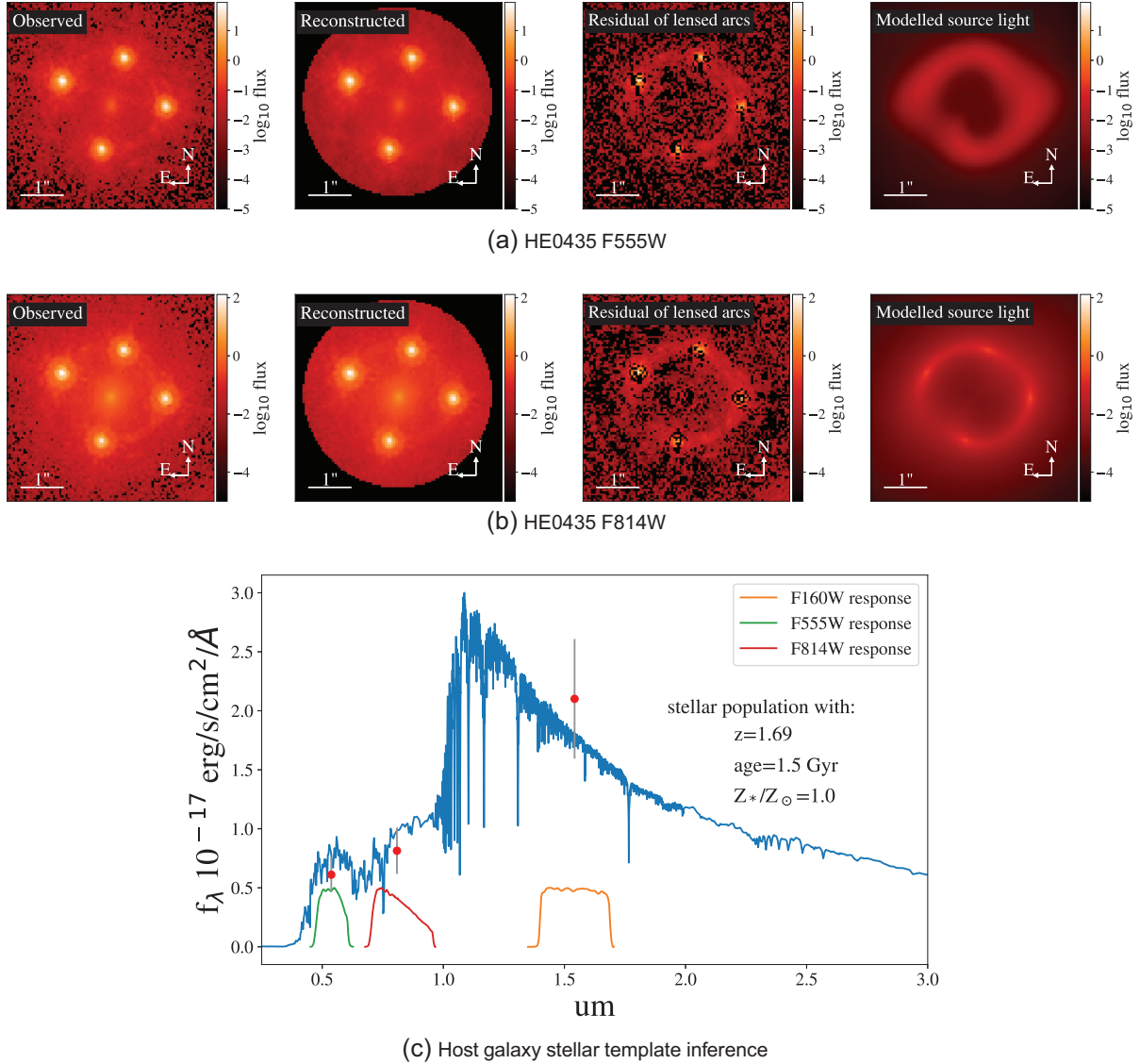
Figure 1. Illustrations of the inference of HE0435 using the multiband data. Top two panels: best-fitting models of the lensed arcs shown as Fig. A1. Bottom panel: SED inference. The red points with error bars represent the inferred image plane host flux in the three bands.

The inference of the fittings is shown in Figs 1(a) and (b). Having obtained the magnitude of the lensed host in the three bands, we use the GSF package[5] (Morishita et al. 2019) to perform the spectral energy distribution (SED) fitting. A range of ages (up to 3.0 Gyr) is used in this fit with a constant star formation rate and a flexible form for star formation histories. Note that there is a well-known degeneracy between age and metallicity; however, this degeneracy has little effect on the M_* inference (Bell & de Jong 2001). In this work, we fix the metallicity to infer the age. In the end, a stellar population with an age of 1.5 Gyr and solar metallicity provides the best fit to the colours, as shown in Fig. 1(c), although this choice is by no means unique. This stellar population is used to estimate the host stellar mass.
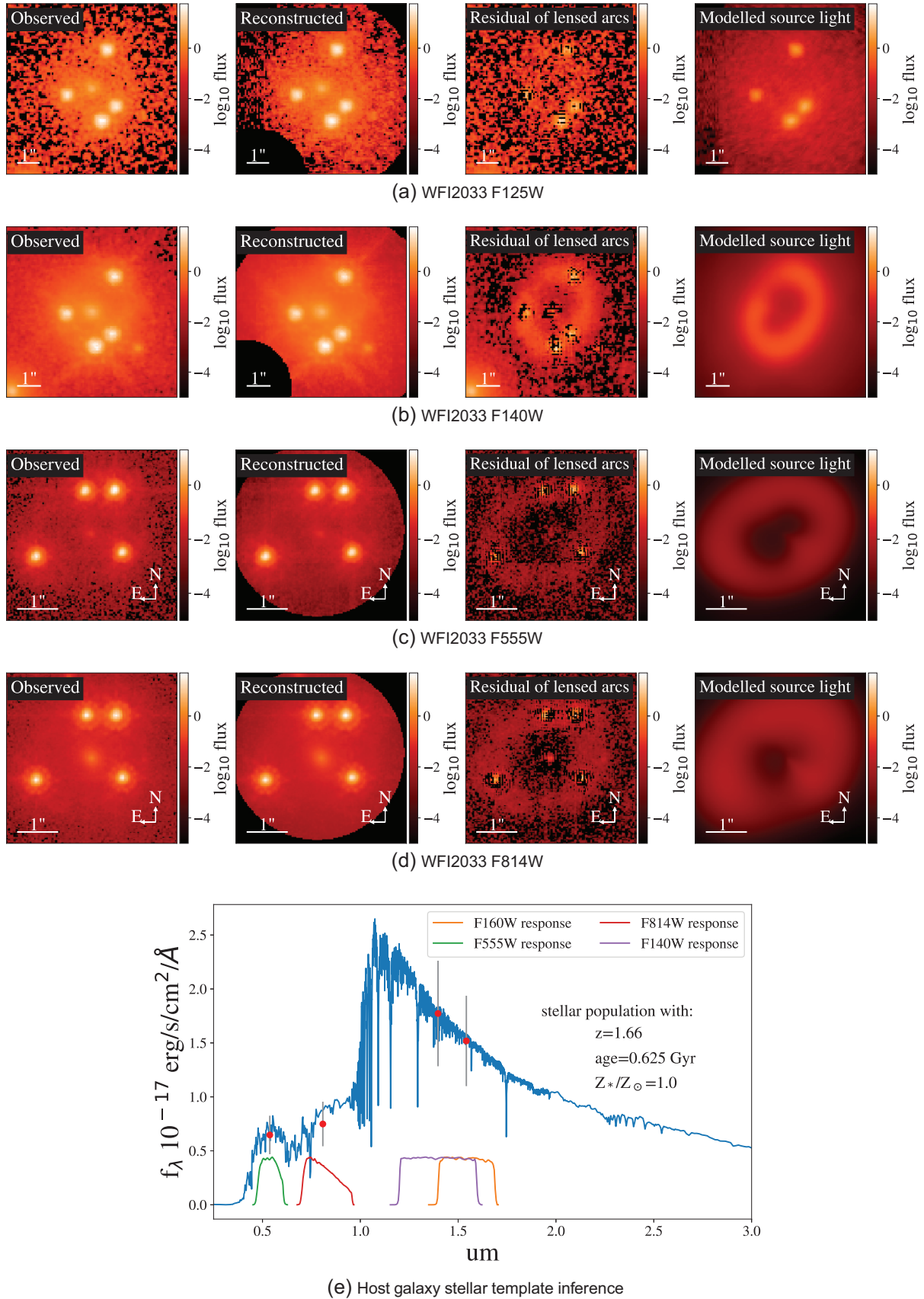
RXJ1131. The host galaxy of RXJ1131 is lensed to a very extended arc in the image plane. The SED of the arc can be directly inferred in the image plane, because lensing is achromatic. Based on

the *HST* imaging data through the three filters F814W, F555W and F160W, we adopt the SED estimated by Ding et al. (2017b) with stellar populations of 3 Gyr and 1.5 Gyr (solar metallicity) for its bulge and disc, respectively. We refer the interested reader to that paper for more details.

WFI2033. In addition to WFC3/F160W, WFI2033 has been observed by the *HST* in four bands: WFC3/F125W (GO-12874; PI: D. Floyd), WFC3/F140W (GO-13732; PI: A. Nierenberg) and ACS/F555W+F814W (GO-9744; PI: C. S. Kochanek). Similar to HE0435, the lens modelling and the host photometry of WFI2033 have been inferred through these four bands. The results are presented in Figs 2(a)–(d). Unfortunately, the F125W data are too shallow to robustly detect the host. Thus, we do not use the F125W band for the SED fitting. Using four-band photometry, we find a stellar population with an age of 0.625 Gyr provides a good match to the colours, as shown in Fig. 2(e).

HE1104 and the other systems. HE1104 has also been observed by the *HST* through the F555W and F814W bands. However, given the limited exposure time in these two bands, the lensed arcs are too

(a) WFI2033 F125W

(b) WFI2033 F140W

(c) WFI2033 F555W

(d) WFI2033 F814W

(e) Host galaxy stellar template inference

Figure 2. Same as Fig. 1 but for WFI2033. The F125W data are too shallow to detect the host and are thus omitted in the fit.
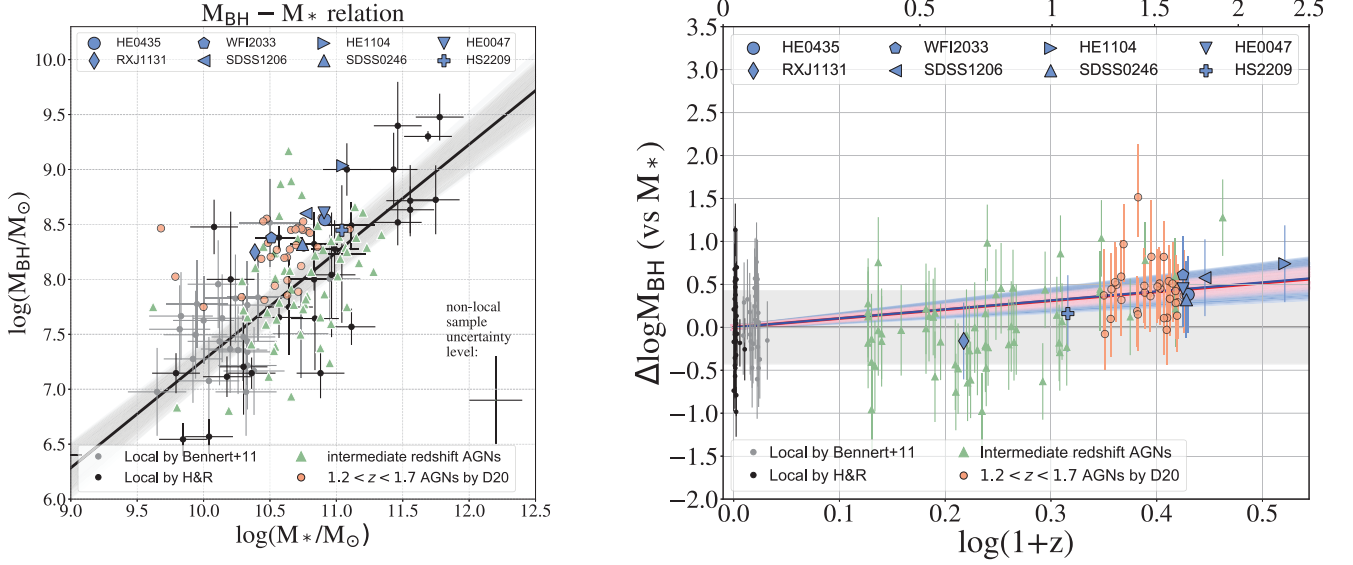
Figure 3. Left: BH masses versus stellar mass correlation ($\mathcal{M}_{BH}$–M_*). The black line and the grey-shaded region indicate the best fit and 1σ confidence interval of the form given by equation (6). Right: offset of $\log(\mathcal{M}_{BH})$ at fixed M_* as a function of redshift. Our new measurements are overlaid on samples of non-lensed AGNs taken from the literature and measured in a self-consistent way to facilitate a direct comparison. The red line and shaded region are the best fit and 1σ results of fitting equation (7) to the 32 high-redshift AGNs measured by D20. The blue line and shaded region are the results based on the eight lensed systems presented in this work. The blue and red areas are in agreement within the errors.

faint to be detected; thus, we are not able to infer the colour of the host for HE1104. No multiband information is available for the other systems. Thus, for these cases, we follow D20 and assume a typical stellar population age, i.e. 1 Gyr and 0.625 Gyr for systems as $z <$ 1.44 and $z >$ 1.44, respectively. Of course, this choice is not unique. However, we stress that as we are mostly interested in a comparison with D20, this strategy is meant to minimize differences between the two approaches.

A summary of the adopted stellar population ages is given in Table 3, column 7. Applying these templates to the filter magnitudes obtained in last section, we derive the stellar mass of our system (Table 3, column 8).

Considering the simulations by Ding et al. (2017a) and the fact that we are able to obtain very consistent host magnitudes for HE0435, RXJ1131 and WFI2033 using the independent approaches, the fidelity of the inferred magnitude is expected to be well characterized by the quoted uncertainties. Our systems do not have sufficient multi-band information to constrain their metallicity and star formation histories. Therefore, we adopt a simple stellar population template to calculate M_*, but reflect the lack of information in the associated uncertainty. Following Bell & de Jong (2001), we estimate a typical uncertainty of 0.2 dex for all the M_* measurements in this paper.

4.2 The $\mathcal{M}_{BH}$–M_* relation

In the left-hand panel of Fig. 3, we plot $\mathcal{M}_{BH}$ versus M_* for our sample, together with the comparison sample introduced in Section 2. We expect that the uncertainty of $\mathcal{M}_{BH}$ (i.e. 0.4 dex) dominates the error budget for the entire sample. Following D20, we adopt, as a baseline, a local $\mathcal{M}_{BH}$–M_* of the form

$$\log\left(\frac{\mathcal{M}_{BH}}{10^7\,\mathrm{M}_\odot}\right) = \alpha + \beta \log\left(\frac{M_*}{10^{10}\,\mathrm{M}_\odot}\right), \tag{6}$$

with $\alpha = 0.27$ and $\beta = 0.98$ based on the local sample of 55 objects measured in a consistent manner. We find that our lensed systems

are above the local relation, consistent with the inference from the 32 non-lensed AGNs published by D20 in a similar redshift range. To quantify the evolution as a function of redshift, we parametrize it as

$$\Delta \log \mathcal{M}_{BH} = \gamma \log(1+z), \tag{7}$$

where $\Delta \log \mathcal{M}_{BH}$ is the offset of $\mathcal{M}_{BH}$ with respect to the local baseline at fixed M_*. To make a direct comparison, we reproduce the plot shown in fig. 8 of D20. Then, we add our new lensed AGN measurements and show the result in the right-hand panel of Fig. 3. We find that the offset from the local relationship is similar for the two samples. Based on equation (7), we fit the evolution of the eight lensed systems and obtain $\gamma = 1.05 \pm 0.44$, which agrees with the results of D20 ($\gamma = 1.03 \pm 0.25$ using 32 AGNs) within the 1σ level. The consistency between the two measurements provides an important verification of the accuracy of the results and strengthens the conclusions drawn by D20 that the observed value of $\mathcal{M}_{BH}$ at a fixed M_* tends to be larger at higher redshift than the ones in the local Universe. It is possible that the uncertainty of M_* could be higher than our assumed 0.2 dex, given that the star formation histories could vary widely at high redshift ($z \sim 2$). However, increasing the uncertainty on the stellar mass does not change our conclusions significantly, as the error budget is dominated by the uncertainty in the BH mass. For example, increasing the stellar mass uncertainty to 0.3 (0.4) dex, we find $\gamma = 1.05 \pm 0.48$ (1.03 ± 0.53).

Note that this apparent evolution is obtained directly using the observed sample, before considering selection effects (Treu et al. 2007; Bennert et al. 2011a; Schulze & Wisotzki 2011, 2014; Park et al. 2015). For instance, adopting the framework introduced in Schulze et al. (2015), D20 estimated that accounting for selection biases would yield a more modest evolution, $\gamma = 0.50 \pm 0.25$. The lensed AGN systems considered in this work were selected for time-delay cosmography, based on the availability of a time delay and the known detectability of the host galaxy. This is a complex selection function, different from that used by D20. Although it is encouraging that the

two samples present the same apparent evolution, inferring the true underlying evolution requires modelling the selection function, as studied by D20. A full modelling of the complex selection function is not warranted, given the small size of the lensed quasar sample. Thus, at this stage, the lensed quasars should be considered as a check on possible systematic measurement error in the D20 analysis rather than a standalone measurement.

We consider all random and systematic effects in Appendix B. To summarize, the uncertainty level of M_* in this work is assumed to be 0.2 dex, as mentioned in the previous section. The uncertainty in the BH mass estimates is at a level of 0.4 dex, which dominates the overall error budget.

5 CONCLUSIONS

Using eight strongly lensed AGN systems, we present new measurements of the correlations of the mass of supermassive BHs with the stellar masses of their host galaxy. We adopt state-of-the-art lens modelling techniques to estimate the magnitude of the host galaxy, in terms of a standard Sérsic profile. We estimated the $\mathcal{M}_{BH}$ of our sample using a set of calibrated single-epoch estimators to assure self-consistency and consistency with the comparison samples taken from the literature.

We directly compare our sample with the recent measurements by D20, who used the same approach to derive M_* and calibrated $\mathcal{M}_{BH}$ using consistent recipes. The $\mathcal{M}_{BH}$–M_* correlation and its evolution with cosmic time are in excellent agreement with the results obtained by D20, as shown in Fig. 3. Currently, the sample of lensed AGNs (i.e. eight systems) is limited in size; as a result, given the statistical errors, it is not worth modelling the selection effect. Nevertheless, the good agreement with the D20 sample strengthens their conclusions, which can be summarized as follows. First, the growth of the supermassive BH pre-dates that of its host galaxy during their co-evolution, even when considering total stellar mass, as is often done at high redshift. However, the actual morphologies of these hosts are likely to be more complex, including a disc and bulge component, as in the case of RXJ1131. In contrast, the galaxies in the local sample are typically bulge-dominated and the bulge mass is adopted as their M_*. Thus, the reported evolution is weaker than would be inferred by comparing $\mathcal{M}_{BH}$ with the bulge M_* at all redshifts (Bennert et al. 2011a). Taken together, these results are consistent with a scenario in which the stellar mass is transferred from disc to bulge at a faster rate than the growth of $\mathcal{M}_{BH}$ as $z \sim 2$.

Our work based on highly magnified AGNs showcases the power of strong lensing to effectively increase the resolution of a telescope. It shows that uncertainties related to lens modelling are subdominant with respect to other sources of uncertainty such as BH mass. Furthermore, our work provides a powerful verification of the fidelity of the host galaxy reconstruction in non-lensed AGNs.

In conclusion, lensed AGNs have great potential to extend the study of the $\mathcal{M}_{BH}$–M_* correlation to higher redshifts than those considered here. So far, this type of work has been limited by sample size. However, given the pace of discovery of lensed quasars in imaging and spectroscopic surveys (e.g. Oguri & Marshall 2010; Agnello et al. 2015, 2018; More et al. 2016; Ostrovski et al. 2017; Schechter et al. 2017; Treu et al. 2018; Lemon et al. 2020), the samples of lensed AGNs with hosts that can be recovered with high fidelity are likely to continue to grow in wide field imaging and spectroscopic surveys. The forthcoming launch of the *James Webb Space Telescope* and the first light of adaptive optics-assisted extremely large telescopes may provide high-quality imaging data of AGNs up to the highest redshift at which they have been discovered.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

Agnello A. et al., 2015, MNRAS, 454, 1260
Agnello A. et al., 2018, MNRAS, 479, 4345
Astropy Collaboration et al., 2013, A&A, 558, A33
Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511
Beifiori A., Courteau S., Corsini E. M., Zhu Y., 2012, MNRAS, 419, 2497
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212
Bennert V. N., Auger M. W., Treu T., Woo J.-H., Malkan M. A., 2011a, ApJ, 726, 59
Bennert V. N., Auger M. W., Treu T., Woo J.-H., Malkan M. A., 2011b, ApJ, 742, 107
Birrer S., Amara A., 2018, Physics of the Dark Universe, 22, 189
Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102
Birrer S. et al., 2019, MNRAS, 484, 4726
Bradley L. et al. , 2016, astropy/photutils: v0.3, Zenodo, https://doi.org/10.5281/zenodo.164986
Cen R., 2015, ApJ, 805, L9
Chantry V., Sluse D., Magain P., 2010, A&A, 522, A95
Chen G. C. F. et al., 2016, MNRAS, 462, 3457
Cisternas M. et al., 2011, ApJ, 741, L11
DeGraf C., Di Matteo T., Treu T., Feng Y., Woo J-H., Park D., 2015, MNRAS, 454, 913
Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33
Ding X. et al., 2017a, MNRAS, 465, 4634
Ding X. et al., 2017b, MNRAS, 472, 90
Ding X., Treu T., Silverman J. D., Bhowmick A. K., Menci N., Di Matteo T., 2020a, ApJ, 896, 159
Ding X. et al., 2020b, ApJ, 888, 37 (D20)

Eigenbrod A., Courbin F., Meylan G., 2007, A&A, 465, 51
Eulaers E. et al., 2013, A&A, 553, A121
Falco E. E. et al., 1999, ApJ, 523, 617
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Gebhardt K. et al., 2001, ApJ, 555, L75
Graham A. W., Onken C. A., Athanassoula E., Combes F., 2011, MNRAS, 412, 2211
Gültekin K. et al., 2009, ApJ, 698, 198
Hagen H.-J., Engels D., Reimers D., 1999, A&AS, 134, 483
Halkola A., Hildebrandt H., Schrabback T., Lombardi M., Bradač M., Erben T., Schneider P., Wuttke D., 2008, A&A, 481, 65
Häring N., Rix H.-W., 2004, ApJ, 604, L89
Hirschmann M., Khochfar S., Burkert A., Naab T., Genel S., Somerville R. S., 2010, MNRAS, 407, 1016
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Inada N. et al., 2005, AJ, 130, 1967
Jahnke K., Macciò A. V., 2011, ApJ, 734, 92
Jahnke K. et al., 2009, ApJ, 706, L215
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M-S., 2015, MNRAS, 450, 1349
Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599
Lemon C. et al., 2020, MNRAS, 494, 3491
Magorrian J. et al., 1998, AJ, 115, 2285
Marconi A., Hunt L. K., 2003, ApJ, 589, L21
Menci N., Gatti M., Fiore F., Lamastra A., 2014, A&A, 569, A37
Menci N., Fiore F., Bongiorno A., Lamastra A., 2016, A&A, 594, A99
More A. et al., 2016, MNRAS, 456, 1595
Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, AJ, 127, 2617
Morishita T. et al., 2019, ApJ, 877, 141
Ofek E. O., Maoz D., Rix H.-W., Kochanek C. S., Falco E. E., 2006, ApJ, 641, 70
Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579
Oguri M. et al., 2005, ApJ, 622, 106
Østman L., Goobar A., Mörtsell E., 2008, A&A, 485, 403
Ostrovski F. et al., 2017, MNRAS, 465, 4325
Park D., Woo J-H., Bennert V. N., Treu T., Auger M. W., Malkan M. A., 2015, ApJ, 799, 164
Peng C. Y., 2007, ApJ, 671, 1098
Peng C. Y., Impey C. D., Rix H.-W., Kochanek C. S., Keeton C. R., Falco E. E., Lehár J., McLeod B. A., 2006, ApJ, 649, 616
Refregier A., 2003, MNRAS, 338, 35
Refsdal S., 1966, MNRAS, 132, 101
Rusu C. E. et al., 2019, MNRAS, 498, 1440
Salviander S., Shields G. A., Gebhardt K., Bonning E. W., 2006, New Astron. Review, 50, 803
Schechter P. L., Morgan N. D., Chehade B., Metcalfe N., Shanks T., McDonald M., 2017, AJ, 153, 219
Schramm M., Silverman J. D., 2013, ApJ, 767, 13
Schulze A., Wisotzki L., 2011, A&A, 535, A87
Schulze A., Wisotzki L., 2014, MNRAS, 438, 3422
Schulze A. et al., 2015, MNRAS, 447, 2085
Shen Y. et al., 2011, ApJS, 194, 45
Sluse D. et al., 2003, A&A, 406, L43
Sluse D., Claeskens J.-F., Hutsemékers D., Surdej J., 2007, A&A, 468, 885
Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, A&A, 544, A62
Smette A., Robertson J. G., Shaver P. A., Reimers D., Wisotzki L., Koehler T., 1995, A&AS, 113, 199
Springel V. et al., 2005, Nature, 435, 629
Sun M. et al., 2015, ApJ, 802, 14
Suyu S. H., Halkola A., 2010, A&A, 524, A94
Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983
Suyu S. H. et al., 2012, ApJ, 750, 10
Suyu S. H. et al., 2013, ApJ, 766, 70
Suyu S. H. et al., 2017, MNRAS, 468, 2590
Treu T., 2010, ARA&A, 48, 87
Treu T., Marshall P. J., 2016, A&AR, 24, 11
Treu T., Malkan M. A., Blandford R. D., 2004, ApJ, 615, L97
Treu T., Woo J.-H., Malkan M. A., Blandford R. D., 2007, ApJ, 667, 117
Treu T. et al., 2018, MNRAS, 481, 1041
Wisotzki L., Koehler T., Kayser R., Reimers D., 1993, A&A, 278, L15
Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, A&A, 395, 17
Wisotzki L., Schechter P. L., Chen H-W., Richstone D., Jahnke K., Sánchez S. F., Reimers D., 2004, A&A, 419, L31
Wong K. C. et al., 2017, MNRAS, 465, 4895
Woo J., Treu T., Malkan M. A., Blandford R. D., 2006, ApJ, 645, 900

APPENDIX A: PHOTOMETRY INFERENCE

We describe the details of the fitting for each system and present the inference of the photometry of the host galaxy.

A1 HE0435

We follow the approach described in Section 3.1. We select five isolated stars in this field as initial PSFs to input to the fitting, for a total of 60 fits. Based on the top eight choices, we perform the weighting algorithm and measure the host flux, host-to-total flux ratio, effective radius and Sérsic index. The inference results are shown in Table 3, columns 2−6.
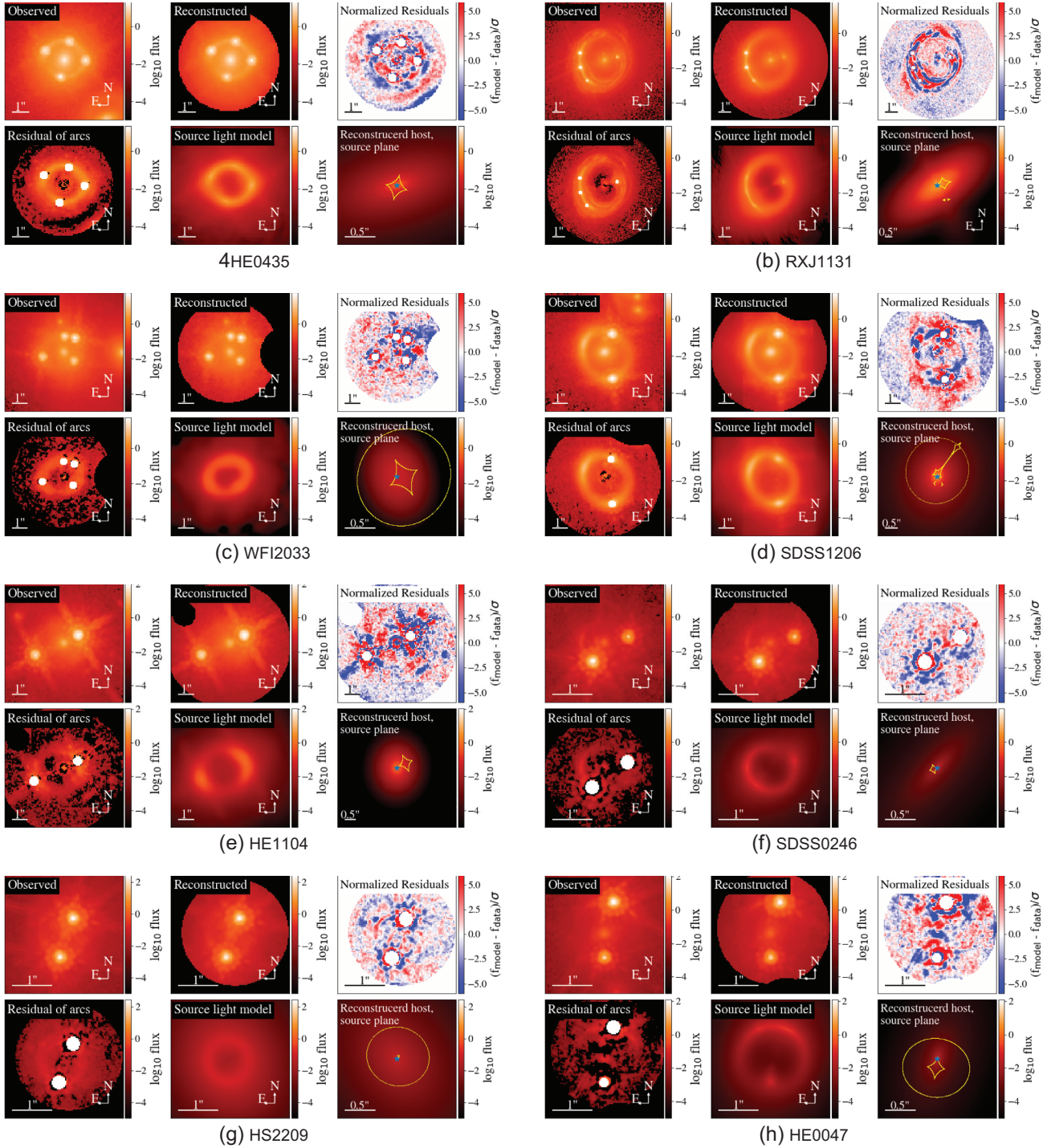
The inference by the best-fitting lens model for HE0435 is shown in Fig. A1(a). Not surprisingly, we note that the level of residuals in the normalized plot appears to be larger than that presented by Wong et al. (2017). This is primarily because the surface brightness of the host galaxy in our model is defined as a Sérsic profile, which is relatively simple and smooth compared with the pixellated reconstruction technique adopted by GLEE (Suyu et al. 2006, 2012; Halkola et al. 2008; Suyu & Halkola 2010) or the shapelet technique adopted by LENSTRONOMY (Birrer, Amara & Refregier 2015). The smooth features of the Sérsic profile cannot capture the clumps in the host galaxy, such as the star-forming regions. Nevertheless, our approach is sufficient to derive a self-consistent one-step inference of the global host light in terms of the Sérsic flux (i.e. the quantity commonly measured in the literature for non-lensed samples).

As a cross-check, we compare our host magnitude to the inference of HE0435 by Ding et al. (2017b). They inferred the Sérsic magnitude by fitting to the pixelized host galaxy as reconstructed by Wong et al. (2017). The host magnitude measured by Ding et al. (2017b) is: *mag* = 21.75 ± 0.13. The results are in excellent agreement with the measurement reported here (i.e. *mag* = 21.50 ± 0.35).

A2 RXJ1131

A lens model of RXJ1131 based on ACS/F814W data has been presented by Suyu et al. (2013). The host galaxy of this system is lensed to an extended arc. A clear bulge and disc component can be identified. Thus, we describe the host galaxy with two Sérsic profiles with index values *n* fixed to 1 and 4, respectively, to mimic the light distribution of the disc and bulge. In addition, following Suyu et al. (2013), we consider the perturbations by the small object (0″.5 in the north) and describe it as a singular isothermal sphere (SIS) and Sérsic profile for its mass and light, respectively.

The photometry of the RXJ1131 host galaxy is fitted with a set of four initial PSFs. The results are summarized in Table 3, with the best-fitting result shown in Fig. A1(b).

Figure A1. Illustrations of the inference with the best-fitting lens model for each system using the AGN centre noise level boosted approach. For each figure, the panels from left to right are as follows. Top row: (1) observed data; (2) best-fitting model; (3) residuals divided by the uncertainty level. Some of the residuals look clumpy; however, they do not affect the host property measurements (for more details, see Appendix B). Bottom row: (4) data minus the model PSF and deflector (i.e. pure lensed arc image); (5) the model of the lensed arc; (6) reconstructed host galaxy in the source plane with the caustic line drawn as a yellow line. Across all the systems, the lensed arc feature can be clearly seen in the fourth panel, indicating strong evidence of a detection. In panels (3) and (4), we use white regions to indicate the areas where the noise level is boosted.

We also compare our measurement with the previous reconstructions by Ding et al. (2017b). Based on the reconstructions by Suyu et al. (2013), the inference of the host magnitudes by Ding et al. (2017b) are $mag_\mathrm{bulge} = 21.81 \pm 0.28$ and $mag_\mathrm{disc} = 20.07 \pm 0.06$. The results are in excellent agreement with our inferred bulge

magnitude (21.80 ± 0.21) shown in Table 3. However, our inferred disc magnitude (19.33 ± 0.16) is brighter than that reported by Ding et al. (2017b). This difference is not surprising because our reconstruction of the host galaxy is based on a much more extended region to collect the disc light, compared with Suyu et al.

(2013) who performed the lens modelling using a smaller lens mask (see fig. 4 therein). Note that, at variance with the procedure described here, Ding et al. (2017b) used a Sérsic model to fit the pixellated source reconstructed by Suyu et al. (2013) in the source plane. We checked that finite grid effects did not introduce any substantial difference. As a sanity check, we find that our inferred effective radius is very consistent: $0\overset{''}{.}90 \pm 0\overset{''}{.}06$ (this paper) and $0\overset{''}{.}84 \pm 0\overset{''}{.}09$ (Ding et al. 2017b). The difference between the Suyu et al. (2013) reconstruction and the one presented here makes sense in terms of the different goals of the two studies. While our primary aim is to reconstruct the host galaxy photometry, for Suyu et al. (2013) it was only a byproduct on the way to time-delay cosmography.

A3 WFI2033

WFI2033 is the last quadruply lensed system in our sample whose lens model has been previously investigated (Rusu et al. 2019). There is a satellite galaxy in the north of the lens. However, the satellite galaxy has a much smaller mass than the main deflector. In addition, there is a galaxy west of the main target, which also has a small effect on the total macro-magnification (< 10 per cent). Thus, we ignore their influence on the magnification but only fit the light of the satellite galaxy using a Sérsic model. We select a total of eight initial PSF stars to model this system.

The final inference results are presented in Table 3 and Fig. A1(c). We compare our inference with the previously reconstructed host galaxies. Modelling the reconstructed host by Rusu et al. (2019) (i.e. their fig. 4 bottom-right plane) as a Sérsic profile, we infer *mag* = 21.98 ± 0.15, which is very consistent with our inference (i.e. *mag* = 21.78 ± 0.25).

A4 SDSS1206

SDSS1206 is a unique system – the AGN is doubly imaged by the deflector while most of the host falls inside the inner caustic and ends up being quadruply imaged. Following Birrer et al. (2019), we consider the galaxy triplet group at the north-west and use an SIS model to denote their overall mass perturbation. Moreover, as noted by Birrer et al. (2019), a subclump is located in the north, which is hardly visible (see fig. 1 in Birrer et al. 2019). We model this subclump as an SIS mass model and a circular Sérsic light model with joint centroids. It is worth noting that we are using the same imaging modelling tool as Birrer et al. (2019).

Because of the limited number of stars in the field of view, there are only two stars available as initial PSFs. To expand the volume of modelling options, we also take the stack of two bright stars, as derived by Birrer et al. (2019), as a third initial PSF. We find visible residuals at the fitted lensed arc region using a single Sérsic model as the host. However, a double Sérsic model does not significantly improve the goodness of fit. Thus, we adopt the single Sérsic model in our final inference. Our inferred results are presented in Table 3 and Fig. A1(d).

A5 HE1104

HE1104 is a typical doubly imaged quasar. We have selected, in total, five initial PSFs to perform the fit. There is an object in the north-east. However, as we do not know its redshift and considering it is further away from the lens, we do not model it explicitly but just mask it out in the fitting.

The inference is presented in Table 3 and Fig. A1(e). The lensed arcs can be clearly seen in the bottom-left panel, indicating that the host galaxy is well detected.

It is worth noting that HE1104 has been modelled by Peng et al. (2006) based on the *HST*/NICMOS *H*-band (F160W) imaging data. Their inferred host light is 20.14 ± 0.30 mag in the Vega system, which is also consistent with our inference (20.00 ± 0.15 mag in Vega).

A6 SDSS0246

Having been imaged with WFC3-UVIS/F814W, the resolution of the data for this system (together with the remaining two systems) is much higher than for those imaged in the infrared, with a drizzled pixel scale of $0\overset{''}{.}03$. However, the arcs are much fainter compared with those of other systems imaged in the infrared band. As a result, fewer pixels with signals are available for the fit than for other systems. Nevertheless, the host inference is successfully reconstructed, as shown in Fig. A1(f) and Table 3 (three initial PSF estimates were adopted).

A7 HS2209

HS2209 was imaged by the *HST* during two visits (*vis*05 and *vis*06) at different orientations. We modelled the two visits separately and recovered mutually consistent host galaxy magnitudes. However, we found that the data from *vis06* can be modelled with smaller residuals; thus, the inference based on *vis06* was adopted as our best estimate (using seven initial PSF stars), as listed in Table 3. The inference for this system is summarized in Fig. A1(g).

A8 HE0047

HE0047 is the most challenging system in our sample with the lowest signal-to-noise ratio of the lensed arcs. The results, based on three initial PSFs, are summarized in Fig. A1(f) and Table 3. The host magnitude of the HE0047 system has a relatively large uncertainty as reflected in the error bars.

APPENDIX B: SYSTEMATIC ERRORS

We considered a set of modelling choices to perform the fitting and the final inference is based on a weighting of top-ranked choices. In particular, we treat the two different modelling approaches (i.e. noise boost and PSF iteration) equally, to derive the averaged results. Of course, the results are dependent on the weighting scheme (e.g. the dispersion of the results by the top-ranked choices). However, the dependence is smaller than other sources of uncertainty. Thus, the use of different weighting schemes would only change the results marginally (<0.1 dex).

We used a range of mass slope values (i.e. 1.9, 2.0, 2.1) to perform the lens modelling. Then, we used our weighting algorithm introduced in Section 3.1 to estimate the systematic uncertainty of our inference and assumed it covers the truth. We apply this method to the entire sample to ensure self-consistency within our sample, even though four systems (HE0435, RXJ1131, WFI2033 and SDSS1206) have been analysed by the H0LiCOW Collaboration, and have high precision slope measurements available. As a sanity check, we calculate the weighted slope value and make a direct comparison with the H0LiCOW inference. The results are the following (here, versus H0LiCOW, the error bars are at the 1σ level): HE0435 (2.032 ± 0.07 versus 1.93 ± 0.02), RXJ1131 (2.02 ± 0.06

versus 1.95 ± 0.045), WFI2033 (1.94 ± 0.07 versus 1.95 ± 0.02) and SDSS1206 (1.98 ± 0.06 versus 1.95 ± 0.05). The consistency of the results supports the robustness of the systematic uncertainty estimated in this work.

Following standard practice in galaxy evolution studies, we use a Sérsic model to describe the surface brightness of the host galaxy. The Sérsic profile is relatively simple and smooth, and cannot capture the clumps in the host galaxy. Thus, the smoothness of the Sérsic profile leads to the relatively large residuals shown in Fig. A1. However, as mention in Appendix A1, our goal is to derive a self-consistent one-step inference of the host properties to compare with the measurements of the non-lensed AGN samples, which are measured using the same methodology. Note that the methodology has to be consistent between lensed and unlensed AGNs, as the use of a different host model may introduce systematic errors. In fact, it is common to have significant residuals when fitting, for example, a Sérsic profile to a galaxy (e.g. D20). These residuals, of course, affect the quality of the fit and could increase the systematics. However, these systematics are much smaller than our target precision of 0.4 dex. The difference in residuals between this work and the H0LiCOW analysis is once again a reflection of the different purposes of the two studies. Whereas fitting the host surface brightness to the noise level is important to determine the gravitational potential with sufficient precision to infer the Hubble constant, it is not necessary when the goal is to infer the luminosity of the host.

In addition, we adopt simple stellar populations to derive the stellar mass. For five of the eight systems in our sample, we did not have colour information, and we used instead a fixed age depending on redshift. The lack of colour information is reflected in an increase of the uncertainty in the inferred M_*. It is important to stress once again that the main goal of this work is to provide an independent test of the D20 measurement. Because we are using the same stellar population models, any uncertainty in the models or other stellar population assumptions will cause an absolute change, but those will cancel out when looking at relative consistency between this work and D20.

We do not expect foreground extinction to be significant, because the deflectors are all massive elliptical galaxies. As a sanity check, Falco et al. (1999) and Østman, Goobar & Mörtsell (2008) report estimates for HE0435, WFI2033 and HE1104 through filter F160W. For HE1104, the estimated extinction is negative, while for HE0435 and WFI2033, the authors do not report an extinction value because standard extinction laws did not fit the data. If we focus on the 13 ellipticals from Falco et al. (1999), the total median extinction is −0.03, which justifies our choice not to apply any correction. We interpret the negative values reported in the literature as resulting from the small effect of dust being overshadowed by chromatic microlensing or variability.

In this work, some assumptions have been made to measure the evolution of M_*–$\mathcal{M}_{BH}$. For example, a Chabrier IMF was assumed to measure the stellar mass for all samples. To compare our high-redshift measurements with the local ones, we adopted the local sample from Bennert et al. (2011a) and Häring & Rix (2004), rather than other samples available in the literature. We adopt our own recipes to calibrate $\mathcal{M}_{BH}$. Of course, different options would shift the absolute value of the inferred M_* and $\mathcal{M}_{BH}$. However, because the entire sample is self-consistent, a different assumption would only shift the global M_*–$\mathcal{M}_{BH}$ together, leaving the offset value and the evolution conclusion the same. More details can be found in section 6 of D20.

This paper has been typeset from a TeX/LaTeX file prepared by the author.