DO THE MOST MASSIVE BLACK HOLES AT z = 2 GROW VIA MAJOR MERGERS?

M. Mechtley1,2, K. Jahnk1, R. A. Windhorst2, R. Andrae1, M. Cisternas3, S. H. Cohen2, T. Hewlett4, A. M. Koekemoer5, M. Schramm5, A. Schulze6, J. D. Silverman6, C. Villforth4, A. van der Wel1, and L. Wisotzki7

1 Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117, Heidelberg, Germany
2 School of Earth and Space Exploration, Arizona State University, P.O. Box 871404, Tempe, AZ 85274-1404, USA
3 Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain
4 Scottish Universities Physics Alliance (SUPA), University of St Andrews, School of Physics and Astronomy, North Haugh, St Andrews, KY16 9SS, UK
5 Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA
6 Kavli Institute for the Physics and Mathematics of the Universe, Todai Institutes for Advanced Study, the University of Tokyo, Kashiwa, 277-8583 (Kavli IPMU, WPI), Japan
7 Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany

Received 2015 October 6; revised 2016 June 29; accepted 2016 July 29; published 2016 October 19

ABSTRACT

The most frequently proposed model for the origin of quasars holds that the high accretion rates seen in luminous active galactic nuclei (AGN) are primarily triggered during major mergers between gas-rich galaxies. While plausible for decades, this model has only begun to be tested with statistical rigor in the past few years. Here, we report on a Hubble Space Telescope study to test this hypothesis for z = 2 quasars with high supermassive black hole masses (M_{BH} = 10^{8}–10^{10} M_{\odot}), which dominate cosmic black hole growth at this redshift. We compare Wide Field Camera 3 F160W (rest-frame V-band) imaging of 19 point source-subtracted quasar hosts to a matched sample of 84 inactive galaxies, testing whether the quasar hosts have greater evidence for strong gravitational interactions. Using an expert ranking procedure, we find that the quasar hosts are uniformly distributed within the merger sequence of inactive galaxies, with no preference for quasars in high-distortion hosts. Using a merger/non-merger cutoff approach, we recover distortion fractions of f_{m,qso} = 0.39 ± 0.11 for quasar hosts and f_{m,gal} = 0.30 ± 0.05 for inactive galaxies (distribution modes, 68% confidence intervals), with both measurements subjected to the same observational conditions and limitations. The slight enhancement in distorted fraction for quasar hosts over inactive galaxies is not significant, with a probability that the quasar fraction is higher P(f_{m,qso} > f_{m,gal}) = 0.78 (0.78σ), in line with results for lower mass and lower z AGN. We find no evidence that major mergers are the primary triggering mechanism for the massive quasars that dominate accretion at the peak of cosmic quasar activity.

Key words: galaxies: high-redshift – quasars: general

1. INTRODUCTION

How do active galactic nuclei (AGN) get the gas that fuels black hole growth? Proximal to the black hole, the direct mechanism for feeding is described by the unification model for AGN (e.g., Antonucci 1993): gas forms a thin accretion disk as it falls into a supermassive black hole (SMBH), with gravitational energy converted into kinetic energy that heats the gas, producing the UV/optical continuum that ultimately drives other observed properties, such as broad and narrow emission lines, and re-emission in the infrared via dust heating. The observational evidence for this general model is secure, with most work now focusing on its details. Less well understood are the physical processes that result in the transport of gas from kiloparsec-scale reservoirs in the galaxy into the central few parsecs, such that can be captured by the SMBH and accreted. The inferred active-phase lifetimes of ≤10^{7}–10^{8} year (e.g., Martini & Weinberg 2001; Yu & Tremaine 2002; Shen et al. 2007) mean that galaxies must provide a significant fraction of the final SMBH mass (roughly 1/e if accreting near the Eddington rate) of gas to the very small volume surrounding the SMBH over a comparable timescale. This requires gas transport mechanisms that efficiently strip angular momentum, allowing gas to pass close enough to the SMBH for capture and accretion.

As the most luminous, massive AGN, the gas transport mechanism most often posited for quasars is disruption due to gravitational interactions with massive galaxies, in particular via major mergers. The most popular version of this model, originally described by Sanders et al. (1988), holds that quasars are a phase in galaxy evolution that follows a starburst phase triggered during a gas-rich major merger. This is an entirely plausible scenario, given that the necessary ingredients are present: galaxy-scale torques and large gas reservoirs. Major mergers—with or without a starburst phase—almost certainly account for some fraction of quasar triggering. However, observational evidence that supports this feeding mechanism uniquely as dominant over other mechanisms (e.g., violent disk instabilities (VDIs) or direct accretion of intergalactic medium via cold gas streams) has remained elusive.

With high spatial resolution, a relatively stable point spread function (PSF), and sensitivity to low surface brightness features in distant galaxies, the Hubble Space Telescope (HST) has been the observatory of choice for many quasar host studies, especially those examining host galaxy morphology. Some early studies of quasar host galaxies with HST lent credence to the major merger model, noting the presence of merging signatures and close companions in some quasar hosts. However, these studies had explicitly biased or unknown sample selection functions, which prevent the results from being easily generalized to the parent populations. For example, Disney et al. (1995) and Bahcall et al. (1997) both contain a mix of radio-loud and radio-quiet objects, but with much higher fractions of radio-loud objects than the quasar population generally. While such studies were important in
unambiguously resolving host galaxy structures for the first time and helping to understand what kinds of galaxies can host quasars, they did not provide conclusive evidence that mergers are the unique or even a dominant trigger for quasar activity.

Later studies began to select more representative quasar samples, with better control over factors like luminosity and redshift ranges. At low redshift—$z \lesssim 0.4$, well after the peak of cosmic quasar activity—the host galaxies of the most luminous quasars were found to be mostly giant ellipticals undergoing only minor (if any) interactions (McLure et al. 1999; Dunlop et al. 2003; Floyd et al. 2004). These studies also took the crucial step of comparing the quasar hosts to inactive galaxies with similar properties (Dunlop et al. 2003), noting that such small-scale disturbances are a common feature of giant elliptical galaxies, including brightest cluster galaxies. These also included the first HST studies of rest-frame visible light in high-redshift quasar hosts (e.g., $z = 1$–3, Kukula et al. 2001; Ridgway et al. 2001), although these were generally limited to lower-luminosity quasars and did not systematically examine morphologies.

Practical considerations have limited most quasar host studies with HST to a few tens of objects each, so implicit or explicit selection biases play an important role in determining how generally a study’s conclusions may be applied. We discuss possible biases in our own sample in Section 2.3 below. Many successful studies have focused on specific areas of quasar parameter space, allowing inferences to be made about specific classes of objects. For instance, almost universal evidence for merging hosts has been found in studies using various “red” or (semi-)obscured quasar selection methods (e.g., Canalizo & Stockton 2001; Urrutia et al. 2008; Glikman et al. 2015). There is some question whether these represent an evolutionary stage or a subset of quasars with dusty, ULIRG-like hosts, which we discuss in greater detail in Section 6.4 below. Chiaberge et al. (2015) also recently examined the different merger fractions of various radio-selected AGN samples, finding significantly higher merger fractions for both faint and luminous radio-loud AGN than for radio-quiet AGN or inactive galaxies. In this study, we choose not to select explicitly based on radio-loudness, which results in an implicit focus mainly on radio-quiet AGN, as these make up $\sim 90\%$ of the luminous quasar population (e.g., Jiang et al. 2007).

Comparison to a sample of inactive galaxies is also key to demonstrating that an observed merger fraction is actually related to AGN activity. Large samples of inactive galaxies are observed at all merger stages, so it is clear that a major merger alone is not a sufficient condition for quasar activity. Thus, to conclusively demonstrate that mergers are an important channel for quasar fueling, we would need to observe an enhancement to the merger fraction in quasar hosts relative to a matched sample of inactive galaxies. Several studies of lower-luminosity AGN host morphologies with inactive control samples have been conducted in HST extragalactic survey fields (e.g., Grogin et al. 2005; Gabor et al. 2009; Cisternas et al. 2011; Schawinski et al. 2011; Kocevski et al. 2012; Böhm et al. 2013; Villforth et al. 2014). In particular, we designed our study methodology following Cisternas et al. (2011), who used visual classification to compare strong distortion signatures in moderate-luminosity X-ray selected AGN hosts to a comparison sample of inactive galaxies in the redshift range $z = 0.3$–1.0. They found no significant enhancement to the merger fraction of AGN hosts relative to inactive galaxies, demonstrating that the majority of cosmic black hole mass accretion at $z < 1.0$, i.e., in AGN with inferred SMBH masses $\lesssim 10^{8} M_{\odot}$ (Vestergaard & Osmer 2009), is not merger driven. How do we then reconcile this result with the results from the red quasar and radio galaxy studies? One possibility is that certain sub-classes of AGN may be preferentially merger driven, even though the bulk of all objects are not. In particular, a downsizing trend has been observed, such that near the peak of quasar activity at $z = 2$, higher-mass SMBHs dominate the cosmic mass accretion ($\approx 10^{9.5} M_{\odot}$, Vestergaard & Osmer 2009). It is possible that forming these most massive black holes requires major mergers, as a particularly efficient gas transport mechanism, and that the declining major merger rate of galaxies is one of the driving forces behind this downsizing trend.

In this paper, we examine the evidence for major mergers among the hosts of 19 of the highest-mass Type-1 quasars at $z = 2$—i.e., broad-line quasars with $M_{\text{BH}} = 10^{9}$–$10^{10} M_{\odot}$ that dominate cosmic accretion at this redshift (Vestergaard & Osmer 2009)—compared to a matched sample of 84 inactive galaxies at the same redshift. This is near the peak of cosmic quasar activity and is the highest redshift where the rest-frame optical emission that reliably diagnoses recent merger signatures can be observed with HST. In Section 2, we describe the selection of the 19 quasars and our observations with the Wide Field Camera 3 (WFC3) infrared channel. In Section 3, we describe the method for modeling and subtracting the central point source from the quasar images. In Section 4, we describe the selection of the comparison sample of inactive galaxies. In Section 5 we describe the procedure for producing a list of galaxies ranked by evidence for strong gravitational distortions, based on visual inspection by ten professionals. Finally, in Section 6, we discuss our statistical analysis of the ranked list and the resultant merger fractions, as well as properties of the quasar hosts such as inferred galaxy masses. Throughout the paper, we adopt a Λ CDM cosmology with $H_0 = 67$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ (Planck Collaboration et al. 2014). Unless otherwise stated, all magnitudes are on the AB system (Oke & Gunn 1983) and have been corrected for Galactic extinction using the reddening map of Schlafly & Finkbeiner (2011).

2. OBSERVATIONS AND DATA REDUCTION

2.1. Sample Definition and Existing Data

As discussed in Section 1, a well defined selection function is necessary to understand the statistical biases present in any AGN sample. Previous contiguous-field morphological studies with HST (e.g., Grogin et al. 2005; Gabor et al. 2009; Cisternas et al. 2011; Kocevski et al. 2012) are substantially volume limited and thus lacking luminous quasars at the highest black hole masses ($M_{\text{BH}} > 10^{9} M_{\odot}$). We are thus interested in extending these studies to test whether black hole accretion is merger driven in the highest-mass black holes at the highest redshift where HST studies can reliably diagnose merger signatures—$z \sim 2$, where the rest-frame optical emission shifts into the WFC3 IR F160W filter.

We selected quasars from the Sloan Digital Sky Survey (SDSS) 5th Data Release Quasar Catalog (Schneider et al. 2007), using virial black hole masses calculated from the MgII line by Shen et al. (2008). We required that the quasars be targeted using the uniform color selection of SDSS (TARGET_QSO_CAP or TARGET_QSO_SKIRT, see Richards...
et al. 2002, 2006), have a redshift in the range \( z = 1.9 - 2.1 \), and SMBH mass \( M_{\text{BH}} = 10^{9.3-10.9} \, M_\odot \). In order to ensure an unbiased sample of all optically luminous massive AGN, the quasars were selected blindly from the parent sample, with no further criteria based on spectral features, broadband colors, or detections at other wavelengths. Of this parent sample, we submitted 115 randomly drawn targets for an HST Cycle 19 SNAPshot survey (Program SNAP 12613, PI: Jahnke). Between October and 2012 September, 19 of these quasars were observed with the WFC3 infrared channel in the \( F160W \) filter (rest-frame V-band). After the HST observations were completed, we examined existing data from the SDSS spectra and wide-area surveys. These data are summarized in Table 1 and discussed below.

Shen et al. (2011) classified the spectral properties of all the SDSS Data Release 7 quasars (Schneider et al. 2010), including the 19 in our HST program. In Table 1 we reproduce those data relevant to deriving the Mg \( \text{II} \) virial BH masses—rest-frame luminosity at 3000 Å and FWHM of the broad Mg \( \text{II} \) component. We note that Shen et al. (2011) also publish a revised calibration for Mg \( \text{II} \) virial BH masses, different from the McLure & Dunlop (2004) calibration used by Shen et al. (2008) that went into our mass selection. The new calibration simply increases the scatter in BH masses—the sample is still representative of luminous AGN with the high BH masses and the median mass, \( M_{\text{BH}} = 10^{9.5} \, M_\odot \), remains unchanged. Table 1 also includes an estimate of the Eddington ratio for each quasar, calculated from the Shen et al. (2011) bolometric luminosity and Mg \( \text{II} \) BH mass.\(^8\) One object was flagged as a \( \text{CIV} \) broad absorption line (BAL) quasar. We manually examined the SDSS spectra of all 19 objects and confirmed that it is the only BAL quasar in the sample. Eighteen of the quasars were covered by the FIRST 1.4 GHz radio survey (White et al. 1997, version 14Dec17) with a detection limit sufficient to assess radio-loudness—one object is radio-loud with \( S_{1.4 \, \text{GHz}} = 900 \, \text{mJy} \) and \( R = 1670 \), and is core-dominated under the classification scheme of Jiang et al. (2007). In addition, one object is classified as hot dust-poor, defined as having a rest-frame mid-IR flux deficiency (Jiang et al. 2010; Jun & Im 2013), with a 2.3\,\mu m luminosity ratio of \( \log (L_{2.3}/L_{0.51}) = -0.71 \). One object out of 19 corresponds to an inferred fraction for each of these special object classes (BAL, radio-loud, hot dust-poor) in the range 0.035–0.15 (beta distribution, 68% confidence interval, see Section 5.2 below). This is consistent with these fractions for the \( z = 2 \) luminous quasar population as a whole (\( \approx 10\% - 15\% \) BALs Gibson et al. 2009; Allen et al. 2011, 10%–15% radio-loud Jiang et al. 2007, \( \approx 2\% \) dust-poor Jun & Im 2013). Table 1 also includes the measured \( F160W \) magnitude of each quasar from our HST imaging (see next section), and the rest-frame V-band absolute magnitude, calculated using the Vanden Berk et al. (2001) median quasar spectrum for the \( k \)-correction.

\(^8\) The Eddington ratio estimate in the Shen et al. (2011) catalog (\text{LOGEDD\_RATIO}) is calculated from the \text{CIV} black hole masses for quasars with \( z \geq 1.9 \). The Eddington ratio we report is calculated using the Mg \( \text{II} \) black hole masses.

### Table 1: Properties of Observed Quasars

| Quasar (SDSS J) | Redshift \( z \) | \( L_{3000} \text{log}(L_z) \) | FWHM Mg \( \text{II} \) (km s\(^{-1}\)) | \( M_{\text{BH}} \text{log}(M_\odot) \) | \( L/L_{\text{Edd}} \) | \( m_{\text{F160W}} \text{mag} \) | \( M_\odot \text{mag} \) | Notes |
|---------------|-----------------|-----------------------------|---------------------------------|------------------|-----------------|------------------|------------------|-------|
| 080216.69+033911.5 | 2.021 | 12.72 | 3600 | 9.3 | 0.49 | 18.55 | 37.99 | ... |
| 082510.09+031801.4 | 2.035 | 12.91 | 4900 | 9.7 | 0.28 | 17.81 | 37.34 | ... |
| 085111.74+301838.7 | 1.917 | 12.54 | 3000 | 9.0 | 0.65 | 18.85 | 37.99 | ... |
| 094373.70+110843.3 | 1.905 | 12.66 | 6200 | 9.7 | 0.11 | 19.91 | 37.99 | C IV BAL |
| 102719.13+584114.3 | 2.020 | 12.67 | 3700 | 9.3 | 0.57 | 18.67 | 37.99 | ... |
| 113820.35+565652.8 | 1.917 | 12.76 | 4300 | 9.5 | 0.30 | 18.12 | 37.99 | ... |
| 120305.42+481313.1 | 1.988 | 12.79 | 4100 | 9.4 | 0.28 | 18.71 | 37.99 | ... |
| 123011.84+401442.9 | 2.049 | 13.17 | 5400 | 9.9 | 0.28 | 18.67 | 37.99 | ... |
| 124949.65+593216.9 | 2.052 | 12.88 | 4100 | 9.5 | 0.35 | 18.11 | 37.99 | ... |
| 131501.14+533314.1 | 1.921 | 12.70 | 8400 | 10.0 | 0.10 | 18.45 | 37.99 | ... |
| 131535.42+253643.9 | 1.926 | 12.57 | 3200 | 9.1 | 0.71 | 18.50 | 37.99 | ... |
| 135851.73+540805.3 | 2.066 | 12.80 | 3700 | 9.4 | 0.54 | 18.27 | 37.99 | ... |

#### Note
Properties of \( z = 2 \) quasars. Columns 1–5 are from the catalog of Shen et al. (2011). Column 1: SDSS name, including the full sexagesimal celestial coordinates; Column 2: systematic redshift; Column 3: rest-frame 3000 Å luminosity; Column 4: FWHM of the broad component of the Mg \( \text{II} \) line emission (median fractional error is 11%); Column 5: Mg \( \text{II} \) virial BH mass (calibration of Shen et al. 2011); Column 6: Eddington ratio, derived using the Mg \( \text{II} \) BH masses and the bolometric luminosity; Column 7: \( F160W \) observed magnitude from our HST observations (photometric error is \( \approx 0.05 \) mag); Column 8: V-band absolute magnitude (\( k \)-corrected using the Vanden Berk et al. (2001) median quasar spectrum); Column 9: special notes or features.

### 2.2. HST Data

All 19 quasars were observed with the WFC3 infrared channel using the \( F160W \) filter (broad H-band). None of the quasars had any existing HST imaging at rest-frame optical wavelengths. As a SNAP program, the observations of necessity have short integration times of 1597 s per target (\( <1 \) orbit). Each observation was split into four exposures, dithered using the standard four-point half-pixel box pattern to

---

8 The Eddington ratio estimate in the Shen et al. (2011) catalog (\text{LOGEDD\_RATIO}) is calculated from the \text{CIV} black hole masses for quasars with \( z \geq 1.9 \). The Eddington ratio we report is calculated using the Mg \( \text{II} \) black hole masses.
improve the PSF sampling, and to assist in the rejection of bad pixels and cosmic rays.

Our data processing began with individual flat-fielded, flux-calibrated exposures delivered by the HST archive. The four exposures for each pointing were combined using the ASTRODRIZZLE software package (Koekemoer et al. 2002, 2013) with an output plate scale of 0′′.0060 per pixel and a pixfrac parameter of 0.8. For F160W observations, this samples the PSF with 2 pixels per FWHM and provides relatively uniform weighting of the individual pixels. We used “ERR” (minimum variance estimator) weighting for the final image combination step. We also generated variance maps that include all sources of noise, including uncertainty from the CALWF3 count rate determination, by copying the WFC3 “ERR” arrays into the standard image arrays, and re-running the drizzle process using the same parameters and weighting scheme. The variance maps are a requirement for our analysis, since the count rate from the quasar point source is significantly higher than that of the sky, and underestimated errors can lead to problems with multi-component fitting, as described in Section 3.2 below.

Despite the short exposure time, the excellent sensitivity of WFC3 IR and low on-orbit near-infrared sky allow the images to reach 1σ limiting surface brightness in the range ≈24.0–24.5 mag arcsec−2. This is sufficient to detect tidal features of major mergers between luminous galaxies, as illustrated by the WFC3 imaging of the most luminous, distorted z ≈ 2 merger in the GOODS/ERS2 field (Van Dokkum & Brammer 2010; Ferreras et al. 2012). We note that the true achieved surface brightness sensitivity is a function of distance from the quasar point source (due to shot noise from the removed point source). The sensitivity to low surface brightness features is discussed in detail in Appendix B.

2.3. Potential Selection Biases

As mentioned in Section 1, particular care must be taken in selecting quasar samples, since there is no unbiased selection method that captures the entire population. Furthermore, selecting non-random subsamples from even large survey catalogs may introduce additional biases, and observational constraints may bias sensitivity to features of interest. For our study, two kinds of bias are the most salient. First, if a universal evolutionary sequence such as that proposed by Sanders et al. (1988) exists, certain quasar selection methods may bias toward a particular phase within that evolution. Second, since we are interested in assessing evidence for mergers via morphological signatures, observational biases affecting sensitivity to those signatures are relevant, in particular surface brightness sensitivity and the rest-frame emitted wavelength. We discuss the observational biases of this program in detail in Appendix B.

Selecting optically luminous broad-line sources (Type 1 quasars) by definition selects objects where the central accretion disk is essentially unobscured along the line of sight. Highly obscured or very red (Type 2) sources missed by this selection may include objects similar to the unobscured sources but viewed from different angles, as well as fundamentally different objects that contain systematically more dust or different dust geometries. For luminous quasars such as those in this study ($L_{\text{bol}} \gtrsim 10^{47}$ erg s−1), the obscured fraction due to anisotropy of the dust torus is estimated from Type 1 quasar spectral energy distribution (SED) modeling to be ≈20%–50% (Lusso et al. 2013). Mid-infrared quasar selection methods directly estimate the total luminous Type 2 fraction—i.e., the above torus-obsured quasars, but also those obscured due to fundamentally different dust geometries—to be ≲50% (Donley et al. 2008; Assef et al. 2015). We thus expect ≲30% of luminous quasars to be hosted in systems with much larger dust covering fractions than our sample, whether these represent the putative “buried” evolution phase with the quasar completely enshrouded in dust, or simply a sub-population of luminous quasars hosted in ULIRG-like dust-rich galaxies. This also sets an upper limit on a buried phase duration of at most 30% of the average quasar lifetime. We argue in Section 6.2 below that this timescale is insufficient for dynamical effects to erase the major merger signatures we are searching for.

Besides the implicit luminosity and unobscured line-of-sight constraints, there are potential biases from explicitly selecting quasars with high-mass black holes. In particular, since the (active) black hole mass function drops sharply at high masses, there could be concern that such high-mass quasars are preferentially near the end of their accretion phases. Although the volume density of active black holes drops at high mass, the Eddington ratio distribution function does not depend strongly on black hole mass (Schulze et al. 2015), so even high-mass black holes like those in our sample seem to have accretion episodes statistically similar to lower-mass black holes. We discuss merger signatures as a function of black hole mass in Section 6.2, but note here that even if the quasar duty cycle is near unity and these quasars grow continuously near the sample average $\langle L/L_{\text{Edd}} \rangle = 0.40$, 15 of the 19 could continue growing for >108 years before they exceeded the maximum mass of our selection region, on the long side for quasar lifetime estimates (Martini & Weinberg 2001; Yu & Tremaine 2002; Shen et al. 2007). Thus, although these high-mass black holes might be experiencing their last major growth phase, there is no a priori reason to assume we are observing the quasars late in the current active phase.

3. POINT SOURCE SUBTRACTION

3.1. PSF Models

As a SNAP program, our HST data did not have dedicated observations of stars to measure the instrument and telescope PSF. The focus of all HST instruments is affected by the telescope’s thermal environment, with changes in solar illumination resulting in de-space of the secondary mirror, the so-called “spacecraft breathing” effect (Bély et al. 1993; Hershey 1998). For applications requiring precise PSF matching, the exposure focal history is best matched by extracting PSF stars from the same images. However, the WFC3 infrared channel PSF has other aberrations (e.g., coma, astigmatism) that vary with position within the WFC3 field of view. Simulated WFC3 PSF models are currently poor matches to observations (e.g., Mechtley et al. 2012; Biretta 2012), so we chose to build a library of empirical PSF models from WFC3 archival data of high signal-to-noise ratio (S/N) stars near the center of the field of view, where all the quasar targets were observed.

To find similar exposures from which to extract PSF models, we searched the HST archive for all single-orbit F160W observations using a 4-exposure dither pattern. We then identified all point sources falling within 0′′.5 of the WFC3 field of view center, and excluded known quasars or radio
sources in the NASA/IPAC Extragalactic Database or were from HST programs specifically targeting AGN. We visually inspected the remaining PSF stars, and excluded any that were contaminated by background galaxies or whose flux distribution was significantly elliptical, i.e., were likely stellar binaries. This left us with a library of eight stars\(^9\) with high S/N, but which still had an accurate count rate determination in their cores. We created drizzled images and variance maps of these stars using the same plate scale and weighting scheme described in Section 2.2 above.

While a posteriori estimation of HST focus is possible from on-orbit thermal measurements using the HST Focus Model (Cox & Niemi 2011), there is yet another (essentially) degenerate source of PSF mismatch, namely the object SED through the broadband F160W filter. Bahcall et al. (1997) noted that for diffraction-limited HST observations through a broadband filter, the color of an observed star (effectively, the SED-weighted average of the monochromatic PSFs) can significantly affect the quality of quasar point source subtractions if the quasar and star SEDs are poorly matched. To first order, redder objects will have a slightly broader PSF (e.g., as measured by FWHM), and bluer objects slightly narrower. Since we do not have detailed information about the shape of the quasar or star SEDs through the filter, and because of the degeneracy with focus, we simply leave the choice of PSF as a free parameter during the fitting procedure, which also allows matching of higher-order features (e.g., differences in diffraction spike patterns).

3.2. Bayesian Modeling Method

In the presence of detectable host galaxy flux, fitting only a single point source when attempting to remove the central quasar light tends to over-subtract the point source, especially for bright or centrally concentrated host galaxies. We, therefore, adopted a simultaneous fitting technique that models the point source and the underlying host galaxy flux distribution, approximating the latter with a Sérsic profile (following Mechtley et al. 2012). We stress that the actual morphologies of the host galaxies may differ from symmetric ellipsoids; the purpose of the Sérsic model is simply to provide a first-order approximation of the surface brightness distribution with a flexible parameterization.

We first attempted the two-component fit using the software GALFIT (Peng et al. 2002, 2010), currently the most widely used 2D surface brightness modeling software. However, this software employs several design decisions that make it less desirable for this particular problem (e.g., as noted in Section 6.2 of Peng et al. 2010).

First, GALFIT uses a standard Levenberg–Marquardt gradient descent method to perform least-squares minimization when fitting models. This method involves calculating a gradient image during each iteration to determine the parameter values to use for the subsequent iteration. Extremely compact models—e.g., a Sérsic profile with small effective radius \((r_e)\) and large index \((n)\)—have most of their gradient information contained within a single pixel, and so the Sérsic degrees of freedom are prone to fitting aberrant pixels (e.g., from PSF mismatch). This essentially creates a false minimum in parameter space, from which the gradient descent cannot escape.

A second, related problem is that GALFIT assumes that the supplied PSF is without error. Even without systematic PSF uncertainties—i.e., a PSF exactly matching the telescope focal history, spectral energy distribution of the quasar point source, etc.—the photon noise can be large enough to cause problems. While our PSF stars were selected to have higher S/N than the quasars, they exceed the quasar S/N by less than a factor of 10. This means that when performing the PSF subtraction, the PSF can contribute up to \(\approx 30\%\) of the per-pixel rms error.

Finally, the model is expected to have covariant parameters, such as the relative flux normalizations of the point source component and Sérsic component. These covariances are not quantified by GALFIT.

To address these problems, we developed our own Markov Chain Monte Carlo (MCMC) simultaneous fitting software, PSFM (Mechtley 2014).\(^{10}\) As an MCMC parameter estimator, it addresses the first and third problems intrinsically, by allowing the user to provide prior probability distributions for the model parameters, and providing as an output product the full posterior probability distribution of model parameters given the observed data. The second problem is addressed by propagating a supplied PSF variance map during the convolution process.

Our software is built upon the PYMC Python module for Bayesian stochastic modeling (Patil et al. 2010). The computational book-keeping tasks—such as implementing the MCMC sampler, setting proposal distributions (see below), evaluating prior probabilities for a given set of parameters, and saving sample traces to disk—are handled by PYMC. We provide a framework that allows the user to simultaneously model an arbitrary number of model components (at this time, point sources, Sérsic profiles, and sky background). The free parameters for each component (e.g., position, total magnitude, Sérsic index, etc.) can either be supplied as a fixed numeric value or as an arbitrary prior probability distribution. An arbitrary number of samples can be drawn from the posterior distribution.

Samples are drawn from the posterior distribution using the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970), which accepts a proposed sample probabilistically based on the ratio of its posterior probability to that of the previous sample in the chain. Each time a proposed sample is drawn from the parameter space, PSFM generates a model image of the intrinsic surface brightness distribution described by the parameters (hereafter, “raw model”). This raw model is then used to generate two further images—one convolved with the PSF (“convolved model”), and a model variance map.

The model variance map is simply the square of the raw model image convolved with the PSF variance map. The intensity, \(I_m(p)\), of a pixel \(p\) in the convolved model is given in Equation (1) below, where the summation is over all pixels \(q\).

\[ I_m(p) = I_0(q)K(p, q)\]

\(I_0(q)\) is the raw model intensity, and \(K(p, q)\) is the PSF weight for pixel \(p\) with the kernel centered at pixel \(q\). If the kernel weights have associated variances \(\sigma_q^2(p, q)\), then the variance propagates in the usual way, and the convolved variance of pixel \(p\) is given by Equation (2) below. This assumes that the off-diagonal covariance terms are zero (the noise between

\(^9\) We note that five of these eight were dedicated PSF star observations from two other quasar host programs: GO 12332 (PI: Windhorst) and GO 12974 (PI: Mechtley).

\(^{10}\) The details of the software implementation are given in Mechtley (2014). The software, documentation, examples, and source code are available at: https://github.com/mmechtley/psMC.
pixels is not correlated), which is almost true for the drizzle parameters used, and is a standard assumption in 2D surface brightness modeling

\[
I_C(p) = \sum_q I_R(q)K(p, q),
\]

\[
\sigma_C^2(p) = \sum_q I_R(q)^2\sigma_R^2(p, q).
\]

The conditional probability of each observed pixel value given the model is then calculated, using a normal distribution with mean equal to the PSF-convolved model, and a variance equal to the summed convolved model and observed variances. The likelihood function is then the joint probability of these individual pixel probabilities, which is then multiplied by the prior probabilities of the parameter values to generate the sample’s posterior probability. The sample is then accepted or rejected based on the Metropolis criterion.

3.3. Model Parameter Estimation and Convergence Checking

Each quasar in our sample was modeled by two simultaneously fit components—a point source and a Sérsic profile. Although the software enables more complex multi-component fitting, the number of required samples (and thus computation time) increases steeply with additional parameters. Since we needed to model a total of 179 images (19 quasar images and 160 comparison galaxy images, see Section 4.1), this argued against iterative hand-crafting of more complex models for hosts that show multiple nuclei or other more complicated structures. Instead, we masked out other galaxies in each image with flux peaks that were separate from the central source (effectively, any galaxy with \( \geq 1.5\) separation), excluding them from the fit. This ensures that the Sérsic profile free parameters are used to fit the light distribution of the host galaxy (or galaxies) surrounding the quasar, rather than neighboring or foreground galaxies. This 2-component model is a computational compromise, with one Sérsic profile being more appropriate than none to avoid over-subtracting the point source. So we caution against over-interpreting the model parameter estimates. For instance, positional offsets between the point sources and the Sérsic profile centers should be interpreted as evidence for asymmetric flux distributions or multiple components in the host, rather than physical separation between the black hole and the center of its host.

The prior distributions adopted for the model parameters are summarized in Table 2. At \( z \approx 2 \), the drizzled 0′060 linear pixel scale corresponds to a physical size of 0.52 kpc. The ranges on the parameter priors were selected to model the entire range of values that are both physically reasonable and detectable. In particular, the Weibull distribution is used to approximate the observed distribution of Sérsic indexes (Yoon et al. 2011; Ryan et al. 2012).\(^{11}\)

Four chains were run for each model, each 100,000 samples with the first 50,000 discarded as a burn-in period, allowing the chains an opportunity to converge in parameter space before they are retained for analysis. Each chain is fully independent, with no parallel tempering of proposal distributions used. We use the Gelman–Rubin Potential Scale Reduction Factor to assess convergence \((\hat{R}, \text{Gelman} & \text{Rubin} 1992)\). This summary statistic compares the variance within individual chains to the variance among the chains to estimate how much sharper the posterior distribution for a parameter could be made by running for more samples. To consider a parameter’s posterior estimate converged, we require \( \hat{R} < 1.05 \), i.e., the estimated potential reduction in the scale of a parameter’s univariate marginalized posterior distribution is less than 5%. Samplers whose chains had not converged (one or more parameters had \( \hat{R} \geq 1.05 \)) after 100,000 samples were continued for another 50,000 samples to increase the burn-in time, up to two additional times. Objects requiring longer convergence times are generally those where the host galaxies are only marginally detected. In all cases the point source parameters (position and magnitude) were well determined and met the above convergence criterion. In four of the 19 models, one or more Sérsic parameters still had \( \hat{R} \geq 1.05 \). Since the goal of the modeling is point source subtraction, and single Sérsic models are a contrived simplification of the true flux distributions, we consider these converged for the purpose of our experiment.

3.4. Posterior Model Analysis

The result of the MCMC fitting process is a collection of samples representing the posterior probability distribution of the free parameters in Table 2, given the observed data. The PSFMC software uses this to create model images—both before and after convolution with the PSF—and an image with all point source components subtracted. These images can either be made from the single maximum a posteriori (MAP) sample, or posterior-weighted. Posterior-weighted here means that an average image is made from the model images of all the retained MCMC samples (hence, weighted by the posterior), rather than the single image from the MAP sample. The purpose of this is primarily diagnostic—for models with well constrained parameter estimates, the posterior-weighted and MAP images are almost identical. For models with poorly constrained parameters or multiple posterior modes (e.g., SDSS J135851.73+540805.3),

\(^{11}\) The Weibull distribution is a continuous, asymmetric distribution with non-negative support. Its probability density function (PDF) is given by:

\[ P(x; \alpha, \beta) = \frac{1}{eta \Gamma \left( \frac{1}{\beta} \right) } \exp[-(x/\beta)^\alpha] . \]
these will be visible in the posterior-weighted images, but not the MAP images, since MAP represent parameter estimates from only a single sample, rather than the full distribution of parameter values. Figure 1 shows the observed quasar, the posterior-weighted model image before PSF convolution, and the image of the host galaxy after subtracting only the point source component of the model from the observed data, both with original sampling and smoothed by a $2 \times 2$ pixel Gaussian.

Several cosmetic features in these images are worth discussing. In some of the raw model images ("Posterior Model" panels), a central dot is visible. This is the location of the point source before convolution with the PSF. The point source-subtracted images for several quasars (e.g., SDSS J113820.35+565652.8 and SDSS J124949.65+593216.9) show negative cores with a ring-like residual structure at $0\farcs02$ radius (the location of the first maximum in the Airy disk). Others (e.g., SDSS J102719.13+584114.3 and SDSS J155447.85+194502.7) show strong positive core residuals (most apparent in the smoothed images). These features are related to PSF focus and/or SED mismatch rather than significant point source over- or under-subtraction, and also appear when subtracting stars from other stars. The per-pixel S/N is lowest in the center due to the point source shot noise, and these mismatch structures are faint compared to the total point source signal, so they do not significantly affect the fitting and have per-pixel S/N < 1 in the point source-subtracted images.

Figure 1. (a) Posterior-weighted model images for each quasar (see text for explanation). All images are displayed with the same arcsinh color stretch. Far left: drizzled, undistorted WFC3 F160W image with 0\farcs060 pixels. Middle left: posterior-weighted model from MCMC fitting process, before convolution with the PSF. Middle right: residual after subtracting only the posterior-weighted point source from original image. Far right: same as middle right panel, but smoothed with a $\sigma = 2$ pixel Gaussian, to suppress high-frequency visual noise. (b)-(e) Posterior-weighted models.

The Astrophysical Journal, 830:156 (29pp), 2016 October 20
Mechtley et al.
Figure 1. (Continued.)
Figure 1. (Continued.)
Figure 1. (Continued.)
Figure 1. (Continued.)
Having samples from the full posterior distribution also allows us to examine parameter covariance, as discussed in Section 3.2. The strongest of these covariances is between the total magnitude of the point source component and the total magnitude of the Sérsic component (though we note that the point source magnitude is still determined to within 0.01 mag on average, and the Sérsic magnitude to 0.1 mag). This is expected, since their combined flux must in some sense add up to the observed flux in the WFC3 image, but since some of the total flux is buried in the sky noise the exact form of the covariance function varies from object to object (arguing against removing one of the magnitude free parameters from the model). Figure 2 plots the posterior distribution for a typical object, marginalizing over each pair of parameters.

### 3.5. Host Galaxy Photometry

For non-morphological analysis (see Section 6.5 below), we measure revised $F160W$ magnitudes for all objects, rather than using the Sérsic approximations resulting from the MCMC process. This is primarily important for deriving upper limits for the non-detections, since in those cases the Sérsic approximation often fits PSF mismatch structures rather than host flux (see e.g., the posterior model image for SDSS J135851.73+540805.3). Fluxes for the 16 detected galaxies, on the other hand, are identical to their Sérsic fits, within the 1-sigma magnitude errors. We perform isophotal photometry on the point source-subtracted images, using a threshold of $S/N = 0.9$ times the rms background noise to determine the lowest isophote. In the galaxy centers, the high shot noise—including contributions from the removed quasar point source and the model PSF—results in few pixels having significant flux values (whether positive or negative, i.e., $S/N$ is lowest in the center), and are primarily due to PSF mismatch (see discussion in Section 3.4 above). We, therefore, use values from the model for pixels within a 0′′.6 diameter circle surrounding the point source location. For galaxies with total $S/N < 2.0$ (the three non-detections), we report the $2\sigma$ upper limit. The difference compared to the Sérsic fit magnitudes is $\Delta m < 0.5$ mag for the detected objects, but for non-detections the upper limits are 1.0–1.5 mag brighter than the values indicated by the fit. The observed $F160W$ magnitudes and corresponding $V$-band absolute magnitudes are reported in Table 3 in Appendix A. The host-to-nuclear luminosity ratio $L_{\text{host}}/L_{\text{Nuc}}$ of our sample spans the range from $<0.01$–0.16, with median 0.058 ($m_{\text{Nuc}}-m_{\text{Host}} = -3.2$ mag). This is roughly the range predicted for quasars with $L/L_{\text{Edd}} \gtrsim 0.1$ from SDSS spectral decompositions (which are sensitive only to host fractions $>0.1$, Vanden Berk et al. 2006).

### 4. INACTIVE GALAXY COMPARISON SAMPLE

#### 4.1. Comparison Sample Selection

As discussed in Section 1, any study that wishes to make a definitive statement about particular AGN triggering mechanisms needs to compare the AGN hosts to a matched sample of inactive galaxies. This matching, in principle, should be done for all properties except accretion rate (AGN versus non-accreting) and the property we wish to test for (morphology). Our quasars were selected by SMBH mass, but this is unfortunately not a directly measurable quantity for large samples of inactive $z = 2$ galaxies in $HST$ extragalactic survey fields. Stellar masses ($M_*$(a)) are, on the other hand, fairly well constrained in such fields from multi-wavelength SED fitting, when the available data cover the entire rest-frame ultraviolet to rest-frame near-infrared range (with the standard caveats of stellar population synthesis modeling, including an assumed universal initial mass function). The question is then how to select galaxies whose stellar masses match the distribution of stellar masses for the quasar hosts. One approach is to assume a certain $M_{\text{BH}}$–$M_*$ relation (e.g., Häring & Rix 2004; Kormendy & Ho 2013), taking into account scatter in the relation, possible redshift evolution, and biases introduced by our selection function, and then select galaxies in the resulting stellar mass range. The second approach is to match samples using some proxy for $M_*$ that is directly observable in both data sets, namely the total $F160W$ magnitude. We opted for this second option, since it requires no assumptions regarding SMBH scaling relations and still captures the full range of galaxies that might reasonably host the high-mass SMBHs. We discuss scaling relations in more detail in Section 6.5.

We obtained $F160W$ images for our comparison sample from the CANDELS $HST$ multi-cycle treasury program (Grogin et al. 2011; Koekemoer et al. 2011), selected using the redshift and photometry catalogs from the 3D-$HST$ program (v4.1, Brammer et al. 2012; Skelton et al. 2014). The five CANDELS treasury fields represent the best comparison data set since they image a moderately wide area in $F160W$ (0.22 deg$^2$), and contain a wealth of ancillary data at other wavelengths to ensure accurate redshift determinations, stellar masses, and AGN identifications. We first selected all galaxies in the redshift range $z = 1.8$–2.2. Because the CANDELS fields also have deep 3.6–8.0 μm imaging from the Spitzer Space Telescope, which samples the red side of the Balmer/4000 Å break at $z = 2$, the 3D-$HST$ photometric redshifts are sufficiently accurate ($\Delta z/(1+z) < 0.1$ at $z = 2$, Skelton et al. 2014). We next exclude probable AGN from the sample by cross-matching the $z \simeq 2$ galaxies with X-ray catalogs (Subaru/XMM-Newton Deep Survey for UDS, Ueda et al. 2008; Chandra Source Catalog for the other fields, Evans et al. 2010). The number of X-ray sources removed from the $z \simeq 2$ sample in each field were: AEGIS: 9, COSMOS: 5, GOODS-N: 4, GOODS-S: 6, UDS: 1. We examined each remaining object and excluded any spurious detections (stars and diffraction spikes). This left us with a parent sample of 1123 $z \simeq 2$ galaxies with $m_{F160W} < 23$ mag, and 150 with $m_{F160W} < 22$ mag.

We then drew ten samples from the posterior distribution of Sérsic component magnitude for each of the 16 detected quasar hosts$^{12}$, and selected the $z \simeq 2$ galaxies which most closely matched these samples in $F160W$ magnitude. The quasar hosts are luminous (median $m_{F160W} = 21.36$ mag), so there are not enough objects in CANDELS to ensure each host has ten unique magnitude-matched comparison galaxies. We, therefore, allowed galaxies to be re-used as comparisons for more than one quasar (i.e., drawn galaxies were replaced before selecting galaxies matched to the next quasar). In total, 48 comparison galaxies were used once, 18 were used twice, and 18 three or more times (with a maximum of seven times). This gave us ten redshift- and magnitude-matched inactive comparison galaxy images for each quasar, or a total of 160 images of

---

12 Three of the 19 quasars—SDSS J082510.09+031801.4, SDSS J102719.13+584114.3, and SDSS J135851.73+540805.3—had Sérsic component magnitudes consistent with the noise limit in the images, i.e., were formally undetected. We did not select comparison galaxies for these objects.
84 unique galaxies. We were originally prepared to either down-select to fewer comparison galaxies per quasar or employ weighting in analyses, but decided this was unnecessary since the weighted and unweighted stellar mass distributions were not significantly different (see Section 6.5 below). Our analyses simply consider the sample of 84 comparison galaxies without re-use, unless otherwise stated (though see Appendix A for some consistency checks that were enabled by this re-use).

4.2. Addition of Synthetic Quasar Point Sources

To allow a fair comparison between quasar hosts and inactive galaxies, we needed to ensure that merger signatures were equally detectable in both sets of images. The CANDELS F160W images are deeper than those in our SNAP program, and the quasar images retain systematic residual patterns from slight PSF mismatch, despite our best efforts at PSF matching (see Section 3.1). We, therefore, added synthetic point sources...
to the comparison galaxy images and performed the same MCMC point source subtraction procedure used for quasars.

For galaxies that were used as comparisons for multiple quasars (Section 4.1 above), the images were first transformed with a unique sequence of mirror flips and 90° rotations, so that no two images of a galaxy were exactly the same. For each comparison galaxy image, we then randomly selected one of the eight PSF stars (see Section 3.1) to act as the synthetic point source. This was then scaled to match the magnitude of the subtracted point source for the galaxy’s corresponding quasar—recalling that ten comparison galaxies were selected to match each quasar host—and inserted into the image centered at the galaxy flux centroid. We then measured the sky noise in the resulting image (the point source having contributed some noise), and added noise if necessary to match the sky noise of the quasar image. This was approximated as an uncorrelated Gaussian field rather than correlated Poisson noise like the real quasar images, but this added field is not the dominant source of noise in the final images. Of the 160 images, 116 required no added noise, with the other 44 requiring at most $\sigma = 0.0047 \, \text{e}^{-\text{s}^{-1}}$ additional noise, corresponding to 29% of the final variance after point source subtraction. Weight images were produced by summing the individual variance sources for these images: the CANDELS (sky) variance map, the CANDELS object signal (approximated as the image in electron units), the scaled point source variance map, and the variance of the added Gaussian field. We then performed point source subtraction using the same MCMC parameter estimation procedure described in Section 3.3, supplying only the remaining seven stars as PSF models, i.e., the star used for each galaxy’s synthetic point source was excluded. The resulting galaxy images then contained the same structural residuals from point source subtraction as the quasar host images.

5. DISTORTION RANKING AND MERGER FRACTION

5.1. Expert Ranking Procedure

After preparing the comparison sample, we now had images for a joint sample of 179 quasar hosts and inactive galaxies.
with similar noise properties and residuals from point source removal. Rather than classifying the galaxies into bins of merger strength or interaction stage (e.g., Cisternas et al. 2011; Kocevski et al. 2012), we instead had ten galaxy morphology experts rank the 179 galaxies by morphological evidence of distortion due to strong gravitational interactions. Since ranking is a relative measure of interaction strength, it gives a natural way to avoid personal statistical biases (e.g., what an individual considers a major versus a minor merger), and allows us to examine how our conclusions change as a function of where exactly the final distinction between merger and non-merger is drawn. While ranking the galaxies, these experts were also asked to note any instances where the galaxy was undetected (i.e., was below the noise limit or completely obscured by image artifacts from the point source subtraction).

We then combined the ranked lists to establish a single consensus sequence of galaxies from the most to the least distorted. This is more difficult than might naively be assumed, since it is equivalent to a ranked voting system and results from social choice theory apply. In particular, Arrow’s Impossibility Theorem (Arrow 1950) states that there is no way to combine

---

13 Ranking was performed by co-authors Cisternas, Cohen, Hewlett, Jahnke, Mechtley, Schulze, Silverman, Villforth, van der Wel, and Windhorst.
### Table 3
Quasar Properties in Consensus Rank Order

| Quasar (SDSS J) | \(m_{\text{host}}\) (mag) | \(m_{\text{host}}\) (mag) | \(M_V\) | \(M_{\text{host}}\) | \(L/L_{\text{Edd}}\) | Rank | SEM | ND Count |
|----------------|--------------------------|--------------------------|-------|-----------------|-----------------|-------|-----|----------|
|                | (1)                      | (2)                      | (3)   | (4)             | (5)             | (6)   | (7) | (8)      |
| 124949.65+593216.9 | 18.22                     | 21.48 ± 0.27             | −23.71 | 9.5             | 0.35            | 6     | 3.8 | 0        |
| 131535.42+253643.9 | 18.64                     | 20.76 ± 0.10             | −24.18 | 9.1             | 0.71            | 12    | 7.1 | 0        |
| 155447.85+194502.7 | 18.94                     | 21.92 ± 0.16             | −23.35 | 9.7             | 0.14            | 17    | 8.6 | 0        |
| 215954.45−002150.1 | 16.73                     | 20.58 ± 0.36             | −23.44 | 9.8             | 0.74            | 19    | 6.9 | 0        |
| 232300.06+151002.4 | 18.19                     | 21.24 ± 0.21             | −23.82 | 9.7             | 0.19            | 21    | 7.1 | 0        |
| 215006.72+120620.6 | 18.72                     | 21.13 ± 0.11             | −23.94 | 9.5             | 0.29            | 24    | 11.1| 0        |
| 145645.53+110142.6 | 18.21                     | 20.94 ± 0.14             | −24.19 | 9.2             | 0.65            | 28    | 7.9 | 0        |
| 120305.42+481313.1 | 18.24                     | 20.98 ± 0.16             | −24.09 | 9.4             | 0.27            | 33    | 8.8 | 0        |
| 143645.80+633637.9 | 17.00                     | 20.41 ± 0.29             | −24.81 | 10.1            | 0.22            | 37    | 12.7| 1        |
| 220811.62−083235.1 | 18.53                     | 21.92 ± 0.23             | −23.01 | 9.2             | 0.63            | 38    | 8.7 | 0        |
| 131501.14+533314.1 | 18.51                     | 22.26 ± 0.37             | −23.67 | 10.0            | 0.10            | 41    | 9.2 | 0        |
| 094737.70+110843.3 | 19.07                     | 21.39 ± 0.11             | −23.51 | 9.7             | 0.11            | 50    | 14.3| 0        |
| 081518.99+103711.5 | 18.97                     | 21.20 ± 0.13             | −23.92 | 9.3             | 0.49            | 66    | 8.0 | 0        |
| 082510.09+031801.4 | 17.74                     | (21.88)                  | (−23.28)| 9.7             | 0.28            | 69    | 10.6| 1        |
| 113820.35−565652.8 | 18.14                     | 21.95 ± 0.41             | −22.97 | 9.5             | 0.30            | 71    | 11.1| 0        |
| 085117.41+301838.7 | 18.88                     | 22.74 ± 0.34             | −22.18 | 9.9             | 0.65            | 72    | 9.6 | 0        |
| 102719.13+584141.4 | 18.68                     | (22.8)                   | (−23.33)| 9.3             | 0.57            | 85    | 11.6| 4        |
| 123011.84+404442.9 | 17.44                     | 19.56 ± 0.11             | −25.63 | 9.9             | 0.28            | 88    | 11.2| 2        |
| 135851.73+540805.3 | 18.25                     | (22.25)                  | (−22.97)| 9.4             | 0.54            | 95    | 5.5 | 9        |

**Note.** Quasar properties, in consensus rank order. Magnitude values in parentheses indicate 2σ upper limits. Column 1: quasar name, giving the full sexagesimal coordinates; Column 2: observed magnitude of the subtracted nuclear point source; Column 3: observed magnitude of the host galaxy; Column 4: rest-frame absolute magnitude of the host galaxy; Column 5: black hole mass; Column 6: standard error on the mean of consensus rank; Column 7: non-detection count, i.e., number of individuals who did not observe the host galaxy; Column 8: standard error on the mean of consensus rank; Column 9: non-detection count, i.e., number of individuals who flagged the host as a non-detection.

such ranked lists such that the consensus sequence has three desirable properties: (1) non-dictatorship, such that each voter has equal weight, (2) unanimity, such that if all voters agree option \(X > Y\), then the consensus sequence also has \(X > Y\), and (3) independence of irrelevant alternatives, such that the consensus preference between \(X\) and \(Y\) depends only upon the individual preferences between \(X\) and \(Y\), and not relationships between other options. Several techniques are available for relaxing one or more of these properties—we chose to relax non-dictatorship by allowing de-weighting of individual votes, clipping those that significantly disagreed with the majority for each galaxy.

The individual lists were merged by calculating the mean rank for each galaxy and its associated variance. We then clipped any individual ranks that were more than 2σ from the mean rank for each galaxy. This excluded 55 of the 1790 individual rankings, roughly evenly distributed between rankers (between zero and 11 individual ranks clipped for each ranker). A total of 237 individual rankings were additionally flagged as non-detections. For comparison, galaxies that were included multiple times (see Section 4), were averaged in the mean variance-weighted mean of the individual rankings as the final rank. If half or more of the individual rankings flagged a galaxy as non-detected, we excluded it from further analysis (as noted in the merger statistics below). Images of all the galaxies in consensus rank order are included in Appendix A.

#### 5.2. Merger Fraction Determination

As with all studies based on visual inspection, individual galaxies show more or less evidence for major mergers. We made this explicit by ranking them on a continuum from the most to least evidence. An alternative to merger fractions, then, is to consider the two samples’ distributions within this ranked continuum. These distributions are formally indistinguishable under either a one-sample Kolmogorov–Smirnov test (\(D = 0.21, p = 0.45\)) or 2-sample Anderson–Darling test (\(A^2 = −0.38, p = 0.52\)). The notion of categorizing galaxies as mergers or non-mergers is convenient for some of the discussion in Section 6, so we discuss the derivation of a merger fraction for each sample below.

To estimate the quasar host and inactive galaxy merger fractions, we visually inspected the consensus sequence and selected a particular cutoff rank, below which we could no longer find any clear merger signatures. We selected rank 30 for this cutoff (corresponding to 32% of the final sample of all quasars and inactive galaxies, minus the non-detections), but note that the particular choice of rank may differ among individual experts. In Appendix A we explore in detail how this choice affects the inferred merger fractions. We only note here that for any reasonable choice of cutoff rank, the qualitative aspects and conclusions below do not change.

With a merger/non-merger cutoff rank of 30, we identified seven quasar hosts with evidence of major mergers, 11 with no evidence of major mergers, and one indeterminate (more than half of individuals flagged them as non-detections). For the inactive galaxies, 24 had evidence of major mergers, 56 had no evidence of major mergers, and four were indeterminate. The probability distribution describing the inferred merger fraction given the finite number of experimental trials is the beta distribution. The PDF for the beta distribution is given by Equation (3):

\[
P(x; a, b) = \frac{(a + b + 1)!}{a! b!} x^a (1 - x)^b. \tag{3}
\]

Here \(a\) and \(b\) are integers denoting the number of mergers and non-mergers, respectively. The resulting probability distributions
Table 4

Inactive Galaxy Properties in Consensus Rank Order

| Field      | ID          | R.A. (degree) | Decl. (degree) | Redshift $z$ | $m_{5175}$ (mag) | $M_{*}$ (log($M_\odot$)) | $A_V$ (mag) | Rank | SEM | Use Count | ND Count |
|------------|-------------|---------------|----------------|-------------|------------------|--------------------------|--------------|------|-----|-----------|----------|
| GOODS-N    | 10993       | 150.077515    | 1.82           | 21.85       | 1.10             | 0.5                      | 13           | 13   | 1  |          | 1        |
| AEGIS      | 11930       | 215.078857    | 1.75           | 21.99       | 9.90             | 0.005                    | 14           | 6    | 1  |          | 1        |
| COSMOS     | 28565       | 215.039334    | 2.07           | 21.88       | 11.4             | 0.6                      | 15           | 19.5 | 2  |          |          |
| AEGIS      | 36852       | 215.093567    | 2.22           | 22.50       | 10.43            | 0.7                      | 16           | 22.4 | 4  |          |          |
| GOODS-S    | 30274       | 53.131084     | 2.24           | 21.31       | 11.16            | 1.2                      | 20           | 5    | 3  |          |          |
| GOODS-N    | 30283       | 189.444336    | 2.10           | 20.61       | 11.32            | 1.1                      | 23           | 5    | 5  |          |          |
| COSMOS     | 10592       | 215.070971    | 1.76           | 21.73       | 11.02            | 0.3                      | 25           | 9    | 1  |          |          |
| GOODS-N    | 10657       | 189.256929    | 2.22           | 21.58       | 11.68            | 1.9                      | 26           | 5    | 2  |          |          |
| AEGIS      | 15871       | 215.030258    | 1.21           | 19.71       | 11.07            | 1.2                      | 27           | 5    | 2  |          |          |
| GOODS-N    | 20709       | 189.464142    | 1.80           | 21.64       | 11.23            | 0.7                      | 29           | 5    | 5  |          |          |
| COSMOS     | 4536        | 150.13826     | 1.80           | 21.89       | 11.17            | 2.4                      | 30           | 6    | 3  |          |          |
| AEGIS      | 36755       | 215.08044     | 1.94           | 21.63       | 10.52            | 0.6                      | 31           | 10.8 | 1  |          |          |
| COSMOS     | 2716        | 150.177567    | 1.94           | 21.79       | 10.81            | 1.0                      | 32           | 10.2 | 1  |          |          |
| AEGIS      | 29591       | 214.799545    | 1.77           | 21.96       | 10.45            | 1.1                      | 34           | 8    | 4  |          |          |
| GOODS-S    | 4568        | 53.164497     | 2.13           | 21.37       | 10.66            | 0.3                      | 35           | 9    | 1  |          |          |
| AEGIS      | 5745        | 215.104919    | 1.76           | 22.48       | 10.63            | 0.6                      | 36           | 16.5 | 1  |          |          |
| GOODS-N    | 37243       | 150.665628    | 2.05           | 21.27       | 11.44            | 0.6                      | 43           | 8    | 3  |          |          |
| GOODS-S    | 36095       | 53.162609     | 1.76           | 21.42       | 10.79            | 1.2                      | 44           | 12.8 | 1  |          |          |
| AEGIS      | 16065       | 214.799965    | 2.19           | 22.51       | 11.14            | 2.6                      | 46           | 9    | 1  |          |          |
| GOODS-N    | 24509       | 215.017929    | 1.86           | 22.54       | 11.07            | 1.0                      | 47           | 10   | 1  |          |          |
| AEGIS      | 25526       | 215.101151    | 1.77           | 21.94       | 10.75            | 0.6                      | 48           | 12.0 | 3  | 12       |          |
| GOODS-S    | 10562       | 53.266895     | 2.19           | 21.76       | 10.80            | 0.7                      | 49           | 8    | 1  |          |          |
| UDS        | 14996       | 34.278588     | 2.07           | 21.74       | 11.20            | 1.3                      | 51           | 16.3 | 1  |          |          |
| GOODS-S    | 37243       | 53.098831     | 2.05           | 21.27       | 11.44            | 0.6                      | 43           | 8    | 3  |          |          |
| GOODS-N    | 36095       | 53.162609     | 1.76           | 21.42       | 10.79            | 1.2                      | 44           | 12.8 | 1  |          |          |
| GOODS-S    | 24509       | 215.017929    | 1.86           | 22.54       | 11.07            | 1.0                      | 47           | 10   | 1  |          |          |
| AEGIS      | 25526       | 215.101151    | 1.77           | 21.94       | 10.75            | 0.6                      | 48           | 12.0 | 3  | 12       |          |
| GOODS-N    | 10562       | 53.266895     | 2.19           | 21.76       | 10.80            | 0.7                      | 49           | 8    | 1  |          |          |
| UDS        | 14996       | 34.278588     | 2.07           | 21.74       | 11.20            | 1.3                      | 51           | 16.3 | 1  |          |          |
| GOODS-S    | 37243       | 53.098831     | 2.05           | 21.27       | 11.44            | 0.6                      | 43           | 8    | 3  |          |          |
| GOODS-N    | 36095       | 53.162609     | 1.76           | 21.42       | 10.79            | 1.2                      | 44           | 12.8 | 1  |          |          |
| AEGIS      | 25526       | 215.101151    | 1.77           | 21.94       | 10.75            | 0.6                      | 48           | 12.0 | 3  | 12       |          |
Table 4
(Continued)

| Field   | ID        | R.A. (degree) | Decl. (degree) | Redshift | $m_{F160W}$ (mag) | $M_\star$ | $A_V$ | Rank | SEM | Use Count | ND Count |
|---------|-----------|---------------|---------------|----------|------------------|----------|--------|------|-----|-----------|----------|
| COSMOS  | 19071     | 150.087845    | -2.37344     | 1.79     | 21.76            | 10.82    | 0.3    | 78   | 8.5| 1         | 0        |
| COSMOS  | 11494     | 150.073929    | +2.297985    | 2.05     | 20.84            | 11.57    | 0.9    | 79   | 5.3| 4         | 14       |
| AEGIS   | 2016      | 215.22639     | +53.017464   | 1.98     | 22.52            | 11.00    | 0.7    | 80   | 9.1| 1         | 0        |
| AEGIS   | 26649     | 215.042541    | +52.989007   | 1.87     | 21.71            | 11.13    | 1.3    | 81   | 6.4| 2         | 0        |
| COSMOS  | 16419     | 150.095642    | +2.350081    | 2.09     | 21.09            | 11.29    | 0.0    | 82   | 8.5| 5         | 22       |
| UDS     | 4701      | 34.273331     | -5.262073    | 2.11     | 21.11            | 11.18    | 0.4    | 83   | 8.3| 2         | 3        |
| COSMOS  | 20983     | 150.108505    | +2.393802    | 1.97     | 21.66            | 11.07    | 0.9    | 84   | 9.5| 1         | 0        |
| COSMOS  | 8443      | 150.079056    | +2.266573    | 1.82     | 21.11            | 10.95    | 0.7    | 86   | 6.9| 7         | 19       |
| GOODS-N | 2295      | 189.11438     | +62.142494   | 2.04     | 21.78            | 11.03    | 0.5    | 87   | 6.9| 1         | 8        |
| UDS     | 32892     | 34.389614     | -5.168086    | 1.90     | 21.24            | 10.98    | 0.3    | 89   | 6.9| 2         | 4        |
| AEGIS   | 9128      | 214.976212    | +52.872486   | 2.13     | 21.87            | 11.32    | 0.5    | 90   | 9.5| 1         | 3        |
| GOODS-N | 33780     | 189.202026    | +62.317154   | 1.86     | 21.41            | 11.29    | 1.1    | 91   | 7.8| 2         | 0        |
| UDS     | 20529     | 34.549076     | -5.208462    | 1.77     | 21.00            | 11.26    | 0.7    | 92   | 5.7| 5         | 9        |
| GOODS-N | 23187     | 189.514648    | +62.255245   | 1.77     | 21.1             | 11.08    | 0.5    | 93   | 7.6| 6         | 9        |
| UDS     | 22802     | 34.446915     | -5.200699    | 1.77     | 21.05            | 10.99    | 0.9    | 94   | 8.6| 5         | 9        |
| GOODS-N | 17735     | 189.060898    | +62.228977   | 1.84     | 21.48            | 10.94    | 0.4    | 96   | 8.0| 1         | 1        |
| UDS     | 19405     | 34.544777     | -5.211998    | 1.86     | 21.75            | 10.92    | 0.4    | 97   | 8.3| 2         | 9        |
| AEGIS   | 12417     | 215.052643    | +52.940708   | 1.82     | 22.46            | 10.94    | 0.3    | 98   | 4.5| 1         | 8        |
| AEGIS   | 35002     | 215.093201    | +53.058929   | 1.79     | 21.92            | 11.07    | 2.2    | 99   | 6.8| 3         | 2        |
| AEGIS   | 12288     | 215.131424    | +52.995796   | 1.98     | 21.86            | 10.83    | 0.4    | 100  | 9.9| 1         | 3        |
| GOODS-S | 42501     | 53.197029     | -27.712065   | 1.82     | 21.70            | 11.00    | 0.6    | 101  | 6.4| 1         | 7        |
| COSMOS  | 17255     | 150.133347    | +2.355615    | 1.76     | 21.71            | 10.82    | 0.3    | 102  | 8.8| 1         | 3        |

Note. Inactive galaxy properties, in consensus rank order. Column 1: CANDELS field name; Column 2: 3D-HST catalog ID number; Column 3: R.A.; Column 4: decl.; Column 5: Redshift (3D-HST $z_{peak}$); Column 6: observed magnitude in the F160W filter; Column 7: stellar mass (from 3D-HST SED fit); Column 8: rest-frame V-band internal dust attenuation (from 3D-HST SED fit); Column 9: consensus rank; Column 10: standard error on the mean of consensus rank; Column 11: number of quasars for which the galaxy was selected as comparison; Column 12: non-detection count, i.e., number of individuals who flagged the host as a non-detection (noting the maximum is Use Count × ten experts).

for both quasar hosts and inactive galaxies are shown in Figure 3. The modes or peaks of the distributions are simply the usual merger fractions, $a/(a+b)$. The inferred parent population merger fractions, with 68% confidence intervals, are thus $f_{m,qso} = 0.39 ± 0.11$ for quasar hosts, and $f_{m,gal} = 0.30 ± 0.05$ for inactive galaxies.

6. DISCUSSION

6.1. Comparison of Inferred Merger Fractions

Although we use the term “merger fraction,” we emphasize that the distorted fractions above should not be used as absolute merger fractions to compare to other studies of high-mass galaxies. We, as with other studies, have invariably missed a few real mergers and possibly mis-identified some non-mergers due to noise and residual patterns associated with the point source subtractions (see discussion in Appendix B). Rather, they should only be interpreted relative to each other, since both samples have the same observational limitations. The probability distribution for the quasar merger fraction in Figure 3 rules out the extreme scenario immediately. The merger fraction is not consistent with values near 1 (99.7% confidence interval: 0.13–0.72), as might be expected if every massive quasar were growing due to ongoing merger activity.

The quasar host PDF peaks at a slightly higher merger fraction than the inactive galaxy PDF, corresponding to a merger fraction enhancement of $f_{m,qso}/f_{m,gal} = 1.3$. The enhancement is not significant, however. Given two random variables X and Y, the probability that $X > Y$ is the integral of the joint PDF over the region where this inequality holds. For the two distributions in Figure 3, this probability is $P(f_{m,qso} > f_{m,gal}) = 0.78$, or 0.78σ. We can invert this to examine the statistical sensitivity of our observations—given the fixed quasar sample size, and assuming the observed PDF
for the inactive galaxies, we can calculate the minimum merger fraction that would have resulted in a significant ($>3\sigma$) signal. For 18 quasars, this would have required 13 mergers, for a merger fraction of $f_{\text{m,qso}} \geq 0.72$ or enhancement of $f_{\text{m,qso}}/f_{\text{m,gal}} \geq 2.4$. Correspondingly, if the true intrinsic distortion fractions are 0.30 and 0.39, we can ask how much larger the sample would need to be to detect this enhancement signal at a significant level. Both the quasar and comparison galaxy sample sizes would need to be increased by a factor of 7.6 to detect an enhancement signal at $2\sigma$, and a factor of 17.2 for $3\sigma$. This underscores the statistical insignificance of the observed enhancement, and the need for cautious interpretation when dealing with beta-distributed quantities inferred from small samples.

6.2. Are These High-mass Quasars Preferentially Hosted in Mergers?

Although the observed enhancement is not significant, the data are still consistent with the quasars having either a slightly enhanced merger fraction, or no enhancement. The data are not, however, consistent with the quasars having a large merger fraction enhancement. If the observed enhancement is real (i.e., if it were still present with much larger samples), it could indicate that mergers are simply one of several possible AGN triggering mechanisms, rather than the dominant or only mechanism.

One possible caveat is that of significant time lags between the appearance of merger signatures and the onset of quasar activity. That is, if gas disturbed or provided by a merging companion takes long enough to reach the SMBH (i.e., several times the dynamical timescale), or is first reprocessed via an episode of star formation, morphological signatures of a merger may no longer be observable. Observational evidence for such time lags is necessarily indirect, and the most successful studies have examined growth regimes very different from those in our study—e.g., Wild et al. (2010), who studied AGN 2.5 orders of magnitude less massive than ours at $z = 0.01–0.07$, and inferred a time lag of 250 Myr between peak starburst activity and peak AGN accretion. Theoretical models of merger-induced AGN activity predict a delay of $\approx 100$ Myr between galaxy coalescence and the peak of quasar activity (Di Matteo et al. 2005; Hopkins et al. 2008). These estimates are at any rate shorter than the timescale over which morphological merger signatures are still visible (as high as $\gtrsim 1$ Gyr for gas-rich mergers Lotz et al. 2010a, 2010b). For exceptionally long time lags, such that merger signatures are almost completely erased, the observational signatures would become essentially indistinguishable from VDIs (e.g., Bournaud et al. 2012; Trump et al. 2014). In such a case, the theoretical model might even be indistinguishable. That is, if gas transport times are several times longer than the dynamical time, is the merger responsible for the loss of angular momentum, or are secular processes like VDI simply acting upon the gas-rich merged galaxy? For the specific feeding mechanism we are testing—near-zero angular momentum gas provided directly by a major merger—the relevant timescale should be closer to the free-fall time and thus shorter than the timescale for merger signatures to be smoothed out.

We also examined the incidence of merger features as a function of black hole mass and accretion rate (Figure 4), with the caveat that our sample spans only about a factor of ten in each. No trend is observed in distortion rank as a function of black hole mass and thus no evidence that the higher-mass black holes are nearer the end of a current merger. Similarly, no trend is observed as a function of either luminosity or Eddington ratio, as might be expected if nuclear gas availability were systematically greater for merger-triggered quasars. We have planned a future study to address the question of merger triggering in high-Eddington systems specifically, as contrasted with low-Eddington AGN and inactive galaxies.

6.3. Results from AGN Environment Studies

Environmental diagnostics provide another test of merger or interaction triggering hypotheses. Small-scale environmental studies test a slightly different hypothesis compared to our study—essentially, that interactions trigger gas instabilities within a galaxy at early merger stages or in non-merging close encounters. Studies of small-scale clustering of quasars (Hennawi et al. 2006; Myers et al. 2008) found some evidence for an excess in the quasar spatial autocorrelation function at small ($\approx 10$ kpc) scales. They attributed this excess to gravitational interaction events triggering quasars in rich environments. Silverman et al. (2009) examined the AGN fraction in galaxies as a function of their local environmental density using a nearest-neighbors approach. They found that the hosts of these lower-luminosity AGN generally trace the same environments as star-forming galaxies (both processes requiring gas), with a preference toward under-dense regions for AGN hosts comparable in mass to our inactive sample ($M_* \gtrsim 10^{11} M_\odot$).

A related diagnostic for examining AGN triggering via gravitational interactions is the study of galaxies in kinematic pairs, i.e., those that are close in mass, spatial projection, and line-of-sight velocity, looking once again for an enhancement to the number of AGN in close pairs versus a field sample.
Ellison et al. (2008) studied low-redshift ($z < 0.15$) kinematic pairs from SDSS (2402 galaxies in pairs, 69,583 in the field control sample), finding the (lower-luminosity) AGN fraction in pairs to be consistent with that of the control sample. Silverman et al. (2011) performed a similar study at $z = 0.25–1.05$ using the zCOSMOS spectroscopic sample (562 galaxies in pairs, 2726 control galaxies). They detected significant differences in the AGN fractions of three subsamples: close pairs (9.7%), wide pairs (6.7%), and isolated galaxies (3.8%). They estimated that 17.8% of moderate-luminosity nuclear activity is triggered during early-stage interactions, leaving the further $\approx 80\%$ unaccounted for. Both the SDSS and zCOSMOS studies posited that later stages of major mergers may account for some of the missing triggers. Lackner et al. (2014) examined this by looking for galaxies with multiple nuclei at small separations (i.e., pairs no longer distinguishable as such in the low-resolution ground-based imaging), finding that the combination of wide pairs, close pairs, and late-stage mergers account for a total of 20% of AGN activity at $z = 0.25–1.0$. This is then consistent with the Cisternas et al. (2011) study, which found that morphologically identified major mergers are not a dominant trigger for low-to-moderate luminosity AGN. We have now further shown that such major mergers are not a dominant triggering mechanism for the high-mass quasars that dominate SMBH accretion at $z = 2$.

6.4. Results from Red and Dust-obscured Quasar Studies

As mentioned in Section 1, studies of reddened and dust-obscured quasars have generally found very high merger fractions, in contrast with most studies of the unobscured AGN or quasar populations mentioned above. Various obscured or dust-rich quasar selection methods target slightly different spectral features. The first HST studies of red quasars targeted objects with far-infrared excesses (e.g., Canalizo & Stockton 2001), finding evidence for major mergers in eight of nine hosts. Although the authors argue that chance hosting is unlikely — i.e., the quasar and starburst activity are related — such objects make up a small fraction of low-redshift quasars, and so correspondingly represent a small contribution to the total triggering budget.

Zakamska et al. (2006) selected Type 2 quasars at $z \approx 0.2–0.4$ from SDSS using the scattered emission-line luminosities as a proxy for total nuclear luminosity — i.e., objects expected to be similar to moderate-luminosity Type 1 quasars, but viewed from an angle where the circumnuclear dust torus obscures the direct AGN light, so that they are not distinct from an evolution standpoint. They found evidence of a major merger in only one of nine hosts, and evidence of tidal debris in three of nine, roughly in line with the lower merger fractions found in low-redshift Type 1 quasars (McLure et al. 1999; Dunlop 2001; Dunlop et al. 2003; Ford et al. 2004).

More recently, considerable effort has gone into the study of dust-reddened Type 1 quasars selected from a combination of radio and near-IR data (F2M quasars, Glikman et al. 2007). These highly reddened quasars make up $\approx 10\%$ of the luminous quasar population, are among the most intrinsically luminous at any given redshift (Glikman et al. 2012), and their host galaxies show a high incidence of mergers (merger fractions of $\approx 0.8–0.85$. Urrutia et al. 2008; Glikman et al. 2015).

How do we reconcile these conflicting results with our current study and other unobscured quasar host studies?

Reddening and obscuration can (at least) originate from non-evolutionary nuclear geometric effects (torus obscuration), evolutionary effects (a buried/blowout phase), and non-evolutionary host geometric effects (i.e., quasars that happen to be hosted in asymmetric, dust-rich, ULIRG-like galaxies). Any given red selection method may pick up some combination of the three. Type 2 selection like that of Zakamska et al. (2006) purports to select only for torus effects, and the similarity of those hosts to Type 1 hosts seems to support this. For red samples with high host merger fractions, additional evidence is needed to determine whether the mergers are an evolutionary feature or a host sub-population feature. The F2M quasars are the best bet for quasars triggered by early-stage galaxy mergers (see especially discussions in Glikman et al. 2012; Urrutia et al. 2012), but the universality of such triggering remains unclear. We argue in Section 6.2 above that, assuming the usual quasar lifetime estimates, we would see far more evidence for mergers in our sample if merger triggering were universal and all (or even most) quasar hosts begin as train-wreck mergers like the F2M hosts.

6.5. Bonus: Stellar Masses and the $M_{BH}–M_*$ Relation

The distribution of the inactive galaxy stellar masses taken from the 3D-HST catalogs is plotted in Figure 5. Two histograms are shown: one where each galaxy is counted only once, and one where each is weighted by the number of times it was used for comparison to a quasar host. The median stellar mass of the unweighted comparison sample is $\log(M_*/M_\odot) = 11.0$. The median stellar mass of the weighted sample (i.e., with the distribution of $F160W$ magnitudes matched to the quasar hosts) is $\log(M_*/M_\odot) = 11.1$. We note that CANDELS/3D-HST are complete to much lower masses. There is no significant trend in stellar mass as a function of distortion rank, consistent with previous studies finding no significant mass-dependence of the major merger rate over similar stellar mass ranges (e.g., Xu et al. 2012).
We can obtain crude estimates of the quasar host galaxy stellar masses from their F160W magnitudes by adopting some assumptions about their stellar populations. Since we sample the SED at a longer wavelength than the Balmer/4000 Å break, the light comes mostly from older stars that account for the bulk of stellar mass. The general procedure is to calculate a galaxy’s luminosity from some bandpass, then multiply by a stellar mass-to-light ratio ($M/L$). This assumes a particular SED, which may differ from object to object. Alternatively, we could use the empirical relation (with scatter) between F160W magnitude and stellar mass from the Skelton et al. (2014) catalog, which implicitly encodes the full range of observed SEDs or $M/L$ ratios for galaxies at a given redshift and stellar mass. The quasar hosts and inactive galaxies show a similar range of morphologies, so we take this approach.

Since this is the same population of galaxies from which the comparison sample was drawn, and since the quasar and inactive galaxy samples have the same luminosity distribution by construction, this should result in a mass distribution roughly the same as the inactive galaxies. Indeed, the quasar hosts have a stellar mass distribution of $\log(M_*/M_\odot) = 11.2 \pm 0.4$, but with an uncertainty on the mean value of 0.4 dex, due to the intrinsic scatter in the $m_{F160W}$–$M_*$ relation (in turn due to the physical range in SEDs for a fixed magnitude). There is no statistically significant difference between the distributions of stellar mass for the mergers and non-mergers, as diagnosed by either a 2-sample Kolmogorov–Smirnov test ($D = 0.33$, $p = 0.68$) or 2-sample Anderson–Darling test ($A^2 = 0.046$, $p = 0.33$). The average $M/L \simeq 0.5$ is consistent with relatively young stellar populations found in previous quasar host studies (e.g., Jahnke et al. 2004a, 2004b; Sánchez et al. 2004).

With stellar mass estimates for the quasar hosts, we can now compare them to the local $M_{BH}$–$M_*$ relation. Figure 6 shows the $z = 2$ quasar hosts alongside the local relation derived by Kormendy & Ho (2013). The error bars include contributions from the scatter in both the virial black hole mass calibration (0.3 dex) and the F160W magnitude-stellar mass relation (0.4 dex, encoding the range of SEDs for a galaxy with a given magnitude). The median black hole-to-stellar mass ratio is $\Gamma = \log(M_{BH}/M_*) = -1.7$. However, some bias in this observed ratio compared to the intrinsic relation is expected, given the sample selection function (Lauer et al. 2007; Schulze & Wisotzki 2011). In particular, when selecting at the high BH mass end of the relation, the corresponding stellar masses are in the exponentially declining regime of the stellar mass function. This leads to average stellar masses lower than the relation. We estimate the expected bias in $\Gamma$ following the framework of Schulze & Wisotzki (2011), using the galaxy stellar mass function from Ilbert et al. (2013), and the $z = 2$ black hole mass function and Eddington ratio distribution function from Schulze et al. (2015). Assuming the SDSS flux limit, and a BH mass selection limit $M_{BH} > 10^9$, this predicts a bias of $\Delta\Gamma = 0.37$ over the Kormendy & Ho (2013) value of $\Gamma = -2.19$ for black holes with $M_{BH} = 10^9.5 M_\odot$. Our quasar hosts are thus consistent with the local relation ($\Delta\Gamma = 0.1$) within uncertainties once selection bias is accounted for, as is the case with previous high-redshift studies (Schulze & Wisotzki 2014).

We note in passing that the lack of redshift evolution is consistent with a picture where the scaling relations arise through non-causal means, i.e., galaxies approach the cosmic mean $\Gamma$ via mergers rather than direct coupling via AGN feedback (Peng 2007; Jahnke & Macciò 2011). However, a lack of evolution does not itself rule out strong AGN feedback, since feedback models can be constructed that predict weak or no evolution.

### 7. CONCLUSIONS

We have performed a study of 19 quasar host galaxies and 84 inactive galaxies at $z = 2$, having ten experts blindly rank them by evidence for distortions due to major mergers. The inactive galaxies were luminosity-matched to the quasar hosts and subjected to the same MCMC modeling and point source subtraction procedure, producing comparison images with the same systematic observational biases and limitations. After synthesizing the expert rankings into a consensus distortion sequence, we have demonstrated that the quasar hosts are consistent with being uniformly distributed within the merger sequence of inactive galaxies. The inferred major merger fraction in host galaxies of massive quasars at $z = 2$ is not significantly higher than the major merger fraction for inactive galaxies, thus the bulk of black hole accretion at the peak of quasar activity is not merger-triggered. This is in line with previous findings regarding the bulk of black hole growth at $z = 0.3$–1.0 (Cisternas et al. 2011) and lower-luminosity AGN at $z = 2$ (Schawinski et al. 2011; Kocevski et al. 2012), supporting the interpretation that mergers are not the dominant fueling channel by which cosmic black hole mass is built up.
We also found no trend in specific accretion rate or black hole mass as a function of merger rank, over a modest (factor of 10) range in each.

We also show that, for a reasonable set of assumptions about the stellar populations of the quasar hosts, they have stellar masses that are consistent with the local black hole-to-stellar mass scaling relation once selection biases have been accounted for. This is consistent with Schulze & Wisotzki (2014), who find that previous high-redshift studies have also not observed a significant offset from the local scaling relations, once observational biases have been accounted for.

We thank the anonymous referee for their careful review and detailed suggestions which improved the manuscript. MM and KJ acknowledge support through the Extraterrestrialstrische Verbundforschung program of the German Space Agency, DLR, grant number 50 OR 1203. MM acknowledges support provided by NASA through grant GO-12613.011-A from the Space Telescope Science Institute, which is operated by the Association for Universities for Research in Astronomy, Inc., under NASA contract NAS 5–26555. AS acknowledges support by JSPS KAKENHI Grant Number 26800098. This research made use of Astropy, a community-developed core Python package for Astronomy (The Astropy Collaboration et al. 2013). MCMC corner plots make use of corner.py (Foreman-Mackey 2016).

Facility: HST (WFC3 IR).

APPENDIX A
CONSENSUS DISTORTION SEQUENCE AND CHOICE OF THE MERGER/NON-MERGER CUTOFF

As discussed in Section 5.2, the choice of where to draw the merger/non-merger distinction is somewhat arbitrary and may differ even among galaxy morphology experts. We chose rank 30 as a reasonable fiducial cutoff rank for our discussion, but also examined how the inferred merger fraction distributions were affected by selecting a different cutoff rank. Figure 7 shows the result of this experiment for cutoff ranks between 15 and 45. The qualitative interpretation—that $f_{m,qso}$ is slightly higher than $f_{m,gal}$, but not significantly—is essentially independent of cutoff rank for any reasonable choice. That is, for no cutoff rank in the range 15–45, is the enhancement of $f_{m,qso}$ over $f_{m,gal}$ significant (>3$\sigma$). In Figure 10, we reproduce the point source-subtracted images of all quasar hosts and inactive galaxies in consensus sequence order, so that readers may make their own assessment of the appropriate cutoff rank. The images are presented with the asinh stretch and color map that was provided to the experts when ordering the sequence. For objects that were ranked multiple times (see Section 4.1), the image with the smallest rank variance (i.e., the image that dominated the consensus rank determination) is shown.

Another concern is the precision with which the consensus sequence is determined, i.e., how certain we are of the final ordering of objects in terms of distortion, and whether the uncertainty of this determination could affect the results? This is determined by the precision with which we can estimate the mean rank for each galaxy, approximated by the standard error on the sample mean after excluding non-detections and those removed by sigma clipping (see Section 5.1). The standard error on the mean is in turn determined by the rank variance among experts (sample variance) and the number of classifications for that galaxy. We calculate the standard error on the mean rank for each galaxy, and generate 10,000 simulated sequences, with the merger cutoff performed at rank 30. Figure 8 shows the result of these simulations. Since there are many more inactive galaxies, their peak merger fraction is primarily determined by the choice of cutoff rank. Roughly 48% of the simulations result in exactly the same quasar merger fraction (the number of quasars above and below the cutoff is unchanged), while for 46% the number of quasar mergers differs only by one. Thus, while more expert classifiers could reduce the uncertainties on the mean ranks, such an improvement in precision is unlikely to change the qualitative result. Finally, since we had 36 inactive galaxies that were used as comparisons more than once (see Section 4.1), we can examine how brighter point sources affect our sensitivity to merger signatures. Figure 9 shows the change in consensus rank as a function of the nuclear-to-host flux ratio, $m_{Nuc} - m_{Host}$, relative to the image with the smallest magnitude difference (which has the greatest sensitivity to faint features). There is significant scatter with increasing point source magnitude, as well as some bias toward lower ranks for the brightest point sources (i.e., experts identified them as having fewer merger signatures). We have not attempted to remove this bias from the consensus sequence.
sequence, because the same observational limitations apply to both the quasar and the inactive galaxy samples, since by construction the inactive sample has the same distribution of $m_{\text{Nuc}} - m_{\text{Host}}$ as the quasars. This underscores our warning against interpreting the merger fractions as absolute major merger fractions, rather than only as relative to one another.

**APPENDIX B**

**POINT SOURCE-SUBTRACTED IMAGES AND SENSITIVITY TO FAINT FEATURES**

As discussed in the main text, the presence of a nuclear point source and its subtraction necessarily affect sensitivity to interaction signatures. For relative comparisons of the inferred merger fractions between the two samples—as in Sections 5.2 and 6 above—it is sufficient that the samples be subjected to the same sensitivity limitations, i.e., that we added synthetic point sources to the inactive comparison sample, with the same distribution of nuclear-host contrast ratios as the quasar sample. Despite absolute merger fraction being a concept with no consistent definition and fraught with problems, some discussion of how the sensitivity constraints differentially affect hosts with distinct morphologies is warranted. Figure 10 shows all 19 quasar hosts and 84 inactive galaxies in consensus sequence order, ranked by the co-authors from the most to least evidence for major mergers. For inactive galaxies, we show both the images with point source residuals as provided to the co-authors for ranking (left panels), and the original CANDELS $F160W$ images for comparison (right panels).

False positive identifications do not appear to be significant i.e., point source residuals have not been mis-identified as merger features. Most of the galaxies in the latter half of the sequence (ranks $>50$), identified as having the least evidence for mergers, are indeed compact and symmetric in the original CANDELS images. However, there are a few galaxies that have stronger evidence for interactions in their original image than is visible in the image used for ranking—e.g., ranks 45, 57, 60, 66, and 74. These are disk-like galaxies with a single nucleus and comparatively low surface brightness, but with some large-scale asymmetric features. The asymmetric features are difficult to pick out with the degraded sensitivity after point source subtraction.

Low surface brightness features remain detectable as long as they extend sufficiently far from the removed point source. For example, extended low surface brightness emission is present in SDSS J124949.65+593216.9 (rank 7), SDSS J131535.42+253643.9 (rank 13), and COSMOS 28565 (rank 16), all of which rank highly due to such emission that forms bridges to moderately bright neighboring galaxies. GOODS-N 30283 (rank 24) appears to be a canonical late-stage major merger with double tidal features—still visible even in the point source-subtracted image—although it ranks slightly lower since the double nucleus is not directly visible in the subtracted image. The average surface brightness of the lower tail is 23.8 mag arcsec$^{-2}$, roughly half as bright as the low surface brightness bridges in the luminous, distorted $z \approx 2$ merger described in Van Dokkum & Brammer (2010) and Ferreras et al. (2012).

Thus, while the point source-subtraction process does reduce surface brightness sensitivity, it does not wholly conceal the features that are used to identify mergers. For the inactive galaxies, symmetrical galaxies are ranked as having the least evidence for mergers, train-wreck clump chain galaxies are ranked as having the most, and those with slight asymmetries or minor companions fall in the middle, with only a handful of exceptions. Furthermore, any biases against correctly ranking certain morphologies applies equally to both the quasar and inactive galaxy samples. The quasar hosts do not seem to show
Figure 10. (a) Point source-subtracted images of the 19 quasar hosts and 84 inactive galaxies, ordered by consensus rank from most distorted to least distorted. All images use the same asinh stretch, 0.060 pixel scale, and 5" x 5" field of view as the images in Figure 1. Left panels: point source-subtracted images presented to the co-authors for ranking. Right panels: original CANDELS F160W images without point source residuals. Quasar hosts have no such image and thus have the words “Quasar Host” in its place. The number in the upper left of each image is its consensus sequence rank (lower numbers are more distorted). Objects which were flagged as non-detections by at least half the individuals were excluded from statistical calculations (see Section 5.1), and are annotated with the white text “Non-detection” at the bottom of the image. The object names below each image are either the SDSS quasar designation as in Table 1 (for quasar hosts), or the field name and galaxy ID number from the 3D-HST catalog (for inactive galaxies). (b)-(e) Consensus sequence.
Figure 10. (Continued.)
Figure 10. (Continued.)
Figure 10. (Continued.)

Quasar Host

SDSS J085117.41+301838.7

AEGIS 17644

COSMOS 19071

AEGIS 26649

COSMOS 20983

GOODS-N 2295

AEGIS 9128

GOODS-N 23187

UDEV 25630

COSMOS 25581

AEGIS 36574

COSMOS 11494

AEGIS 2016

COSMOS 16419

UDS 4701

Quasar Host

SDSS J102719.13+584114.3

COSMOS 8443

Quasar Host

SDSS J123011.84+401442.9

UDS 32892

Quasar Host

SDSS J135851.73+540805.3

UDS 20529
a preference for any particular morphology compared to the inactive galaxies.

We do not attempt to precisely calibrate an absolute major merger fraction for either sample. However, given the relatively small number of inactive galaxies with wholly obscured merger evidence, we are confident in our assertion that the underlying merger fraction for the quasar hosts is indeed lower than the ≈0.8 fraction found in reddened quasar studies (see Section 6.4).

REFERENCES

Allen, J. T., Hewett, P. C., Maddox, N., Richards, G. T., & Belokurov, V. 2011, MNRAS, 410, 860
Antonucci, R. 1993, ARA&A, 31, 473
Arrow, K. J. 1950, J. Political Econ., 58, 328
Assef, R. J., Eisenhardt, P. R. M., Stern, D., et al. 2015, ApJ, 804, 27
Bahcall, J. N., Kirhakos, S., Saxe, D. H., & Schneider, D. P. 1997, ApJ, 492, 547
Bély, P. Y., Hasan, H., & Miebach, M. 1993, Orbital Focus Variations in the Galaxy, Baltimore, MD: STScI
Biretta, J. 2012, TinyTIM Modeling of WFC3/IR Images, Instrument Science Report SESD-1997-01 Baltimore, MD: STScI
Bohm, A., Wisotzki, L., Ebel, E. F., et al. 2004, ApJL, 604, L89
Bournaud, F., Juneau, S., Le Floc’h, E., et al. 2012, ApJS, 196, 29
Böhm, A., Wisotzki, L., Bell, E. F., et al. 2013, A&A, 549, A6
Bouma, F., Juneau, S., Le Floc’h, E., et al. 2012, ApJ, 757, 81
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Bhattacharjee, M., Gilli, R., Lotz, J. M., & Norman, C. 2015, ApJ, 806, 147
Bély, P. Y., Hasan, H., & Miebach, M. 1993, Orbital Focus Variations in the Galaxy, Baltimore, MD: STScI
Bok, A., Wisotzki, L., Bell, E. F., et al. 2013, A&A, 549, A6
Bournaud, F., Juneau, S., Le Floc’h, E., et al. 2012, ApJ, 757, 81
Brüggen, M., van Dijk, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Canalizo, G., & Stockton, A. 2001, ApJ, 555, 719
Chiaberge, M., Jiang, L., Brandt, W. N., et al. 2010, ApJ, 710, 28
Cisternas, M., Jahnke, K., Jasker, J. K., et al. 2011, ApJ, 726, 57
Cox, C., & Niemi, S.-M. 2011, Evaluation of a temperature-based HST focus model, Instrument Science Report TEL 2011-01 Baltimore, MD: STScI
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 435, 642
Disney, M. J., Boyce, P. J., Blakes, J. C., et al. 1995, Nature, 376, 150
Donley, J. L., Rieke, G. H., Pérez-González, P. G., & Barro, G. 2008, ApJ, 687, 111
Donlup, J. S. 2001, in QSO Hosts and their Environments, Vol. XVII (Granada: Kluwer/Plenum), 376
Donlup, J. S., McLeure, R. J., et al. 2003, MNRAS, 340, 1095
Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, AJ, 135, 1877
Evans, I. N., Primini, F. A., Glottfelter, K. J., et al. 2010, ApJS, 189, 37
Ferreras, I., Pasquali, A., Khocharf, S., et al. 2012, AJ, 144, 47
Floyd, D. J. E., Kukula, M. J., Dunlop, J. S., et al. 2004, MNRAS, 355, 196
Foreman-Mackey, D. 2016, corner.py: Scatterplot matrices in Python, The Journal of Open Source Software, doi:10.21105/joss.00024
Gabor, J. M., Impey, C. D., Jahnke, K., et al. 2009, ApJ, 691, 705
Gelman, A., & Rubin, D. B. 1992, J. Stat. Sci., 7, 457
Gibson, R. R., Jiang, L., Brandt, W. N., et al. 2009, ApJ, 692, 758
Glikman, E., Helfand, D. J., White, R. L., et al. 2007, ApJ, 667, 673
Glikman, E., Simmons, B., Mailly, M., et al. 2015, ApJ, 806, 218
Glikman, E., Urrutia, T., Lacy, M., et al. 2010, ApJ, 757, 51
Grogan, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Haring, N., & Rix, H.-W. 2004, ApJL, 604, L109
Hastings, W. K. 1970, Biometrika, 57, 97
Hennawi, J. F., Strauss, M. A., Oguri, M., et al. 2006, AJ, 131, 1
Hershey, J. L. 1998, Modelling HST Focal-length Variations, Instrument Science Report SESD-1997-01 Baltimore, MD: STScI
Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356
Hibbert, O., McCracken, H. J., Le Févre, O., et al. 2013, A&A, 556, A55
Jahnke, K., Kuhlebrot, B., & Wisotzki, L. 2004, MNRAS, 352, 399
Jahnke, K., & Macciò, A. V. 2011, ApJ, 734, 92
Jahnke, K., Sanchez, S. F., Wisotzki, L., et al. 2004b, ApJ, 614, 568
Jiang, L., Fan, X., Ivezic, Z., et al. 2007, ApJ, 656, 680
Jiang, L., Fan, X., Brandt, W. N., et al. 2010, Natur, 464, 380
Jurić, M., & Mihos, J. C. 2013, ApJ, 779, 104
Kocevski, D. D., Faber, S. M., Mozena, M., et al. 2012, ApJ, 744, 148
Koekemoer, A. M., Ellis, R. S., McLeure, R. J., et al. 2013, ApJ, 709, 3
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Koekemoer, A. M., Fruchter, A. S., Hook, R. N., & Hack, W. 2002, in 2002 HST Calibration Workshop, ed. S. Arribas, A. Koekemoer, & B. Whitmore (Baltimore, MD: STScI), 337
Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511
Kukula, M. J., Dunlop, J. S., McLeure, R. J., et al. 2001, MNRAS, 326, 1533
Lackner, C. N., Silverman, J. D., Salvato, M., et al. 2014, AJ, 148, 137
Lauer, T. R., Tremaine, S., Richstone, D., & Faber, S. M. 2007, ApJ, 670, 249
Lottz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, MNRAS, 404, 590
Lottz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010b, MNRAS, 404, 575
Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013, ApJ, 777, 86
Martini, P., & Weinberg, D. H. 2001, ApJ, 547, 12
McLure, R. J., & Dunlop, J. S. 2004, MNRAS, 352, 1390
McLure, R. J., Kukula, M. J., Dunlop, J. S., et al. 1999, MNRAS, 308, 377
Mechtley, M. 2014, PhD thesis, Arizona State Univ.
Mechtley, M., Windhorst, R. A., Ryan, R. E., et al. 2012, ApJL, 756, L38
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, JChPh, 21, 1087
Myers, A. D., Richards, G. T., Brunner, R. J., et al. 2008, ApJ, 678, 635
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Patil, A., Huard, D., & Fonnesbeck, C. J. 2010, J. Stat. Softw., 35, 1
Peng, C. Y. 2007, ApJ, 671, 1008
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766
Ridgway, S. E., Heckman, T. M., Calzetti, D., & Lehnert, M. 2001, ApJ, 550, 122
Ryan, R. E., Jr., McCarthy, P. J., Cohen, S. H., et al. 2012, ApJ, 749, 53
Sánchez, S. F., Jahnke, K., Wisotzki, L., et al. 2004, ApJ, 614, 586
Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988, ApJ, 325, 74
Schawinski, K., Treister, E., Urry, C. M., et al. 2011, ApJL, 727, L31
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102
Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360
Schulze, A., Bongiorno, A., Gavignaud, I., et al. 2015, MNRAS, 447, 2085
Schulze, A., & Wisotzki, L. 2011, A&A, 535, A87
Schulze, A., & Wisotzki, L. 2014, MNRAS, 438, 3422
Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169
Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45
Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, AJ, 133, 2222
Silverman, J. D., Kampczyk, P., Jahnke, K., et al. 2011, ApJ, 743, 2
Silverman, J. D., Kováč, K., Knobel, C., et al. 2009, ApJ, 695, 171
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
The Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Trump, J. R., Barro, G., Juneau, S., et al. 2014, ApJ, 793, 101
Ueda, Y., Watson, M. G., Stewart, I. M., et al. 2008, ApJS, 179, 124
Urrutia, T., Lacy, M., & Becker, R. H. 2008, ApJ, 674, 80
Urrutia, T., Lacy, M., Spoon, H., et al. 2012, ApJ, 757, 125
Van Dokkum, P. G., & Bramer, G. 2010, ApJL, 718, L73
Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549
Vanden Berk, D. E., Shen, J., Yip, C.-W., et al. 2006, AJ, 131, 84
Vestergaard, M., & Osmer, P. S. 2009, ApJ, 699, 800
Villforth, C., Hamann, F., Rosario, D. J., et al. 2014, MNRAS, 439, 3342
White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, ApJ, 475, 479
Wild, V., Heckman, T., & Charlot, S. 2010, MNRAS, 405, 933
Xu, C. K., Zhao, Y., Scoville, N., et al. 2012, ApJ, 747, 85
Yoon, I., Weinberg, M. D., & Katz, N. 2011, MNRAS, 414, 1625
Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965
Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2006, AJ, 132, 1496