Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **409,** 169–183 (2010) doi:10.1111/j.1365-2966.2010.17311.x

LoCuSS: connecting the dominance and shape of brightest cluster galaxies with the assembly history of massive clusters

Graham P. Smith,[1,2⋆] Habib G. Khosroshahi,[3,4⋆] A. Dariush,[1,5] A. J. R. Sanderson,[1] T. J. Ponman,[1] J. P. Stott,[3] C. P. Haines,[1] E. Egami,[6] and D. P. Stark[7]

[1]*School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT*
[2]*California Institute of Technology, Mail Code 105-24, Pasadena, CA 91125, USA*
[3]*School of Astronomy, Institute for Research in Fundamental Sciences (IPM), PO Box 19395-5746, Tehran, Iran*
[4]*Astrophysics Research Institute, Liverpool JMU, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD*
[5]*Cardiff School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA*
[6]*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[7]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*

Accepted 2010 July 6. Received 2010 July 2; in original form 2009 October 21

ABSTRACT

We study the luminosity gap, Δm_{12}, between the first- and second-ranked galaxies in a sample of 59 massive ($\sim 10^{15}\,\mathrm{M}_{\odot}$) galaxy clusters, using data from the Hale Telescope, the *Hubble Space Telescope*, *Chandra* and *Spitzer*. We find that the Δm_{12} distribution, $p(\Delta m_{12})$, is a declining function of Δm_{12} to which we fitted a straight line: $p(\Delta m_{12}) \propto -(0.13 \pm 0.02)\Delta m_{12}$. The fraction of clusters with 'large' luminosity gaps is $p(\Delta m_{12} \geq 1) = 0.37 \pm 0.08$, which represents a 3σ excess over that obtained from Monte Carlo simulations of a Schechter function that matches the mean cluster galaxy luminosity function. We also identify four clusters with 'extreme' luminosity gaps, $\Delta m_{12} \geq 2$, giving a fraction of $p(\Delta m_{12} \geq 2) = 0.07^{+0.05}_{-0.03}$. More generally, large luminosity gap clusters are relatively homogeneous, with elliptical/discy brightest cluster galaxies (BCGs), cuspy gas density profiles (i.e. strong cool cores), high concentrations and low substructure fractions. In contrast, small luminosity gap clusters are heterogeneous, spanning the full range of boxy/elliptical/discy BCG morphologies, the full range of cool core strengths and dark matter concentrations, and have large substructure fractions. Taken together, these results imply that the amplitude of the luminosity gap is a function of both the formation epoch and the recent infall history of the cluster. 'BCG dominance' is therefore a phase that a cluster may evolve through and is not an evolutionary 'cul-de-sac'. We also compare our results with semi-analytic model predictions based on the Millennium Simulation. None of the models is able to reproduce all of the observational results on Δm_{12}, underlining the inability of the current generation of models to match the empirical properties of BCGs. We identify the strength of active galactic nucleus feedback and the efficiency with which cluster galaxies are replenished after they merge with the BCG in each model as possible causes of these discrepancies.

Key words: gravitational lensing: strong – galaxies: clusters: general – galaxies: elliptical and lenticular, cD – galaxies: haloes – X-rays: galaxies.

1 INTRODUCTION

Numerical simulations and large-scale redshift surveys both indicate that we live in a hierarchical universe, i.e. one in which the large-scale structure of the universe grows from the bottom up with smaller objects forming earlier than larger objects. This picture rests on the matter content of the universe being dominated by collisionless dark matter particles, smoothly distributed at early times and seeded with small density perturbations. Exploring this picture observationally in the non-linear regime of gravitational collapse, i.e. within collapsed dark matter haloes that host individual galaxies through to massive clusters of galaxies, complements the statistical analysis of the linear regime probed by galaxy redshift surveys. Indeed, a promising route to fleshing out our understanding of hierarchical structure formation is to measure observable quantities that

⋆E-mail: gps@star.sr.bham.ac.uk (GPS); habib@ipm.ir (HGK)

are sensitive to the age and/or assembly history of dark matter haloes and thus in principle to test the hierarchical paradigm by comparing the observed and predicted distributions. Any discrepancies found between observation and theory may ultimately point to modifications to the theoretical model including, for example, the properties of the dark matter particle and the distribution of initial density fluctuations (e.g. Komatsu et al. 2009). Quantities discussed in the literature that may be useful probes of the age and assembly history of dark matter haloes include the luminosity gap between the first- and second-ranked galaxies in a group or cluster (often expressed as the difference between their magnitudes, $\Delta m_{12} = m_1 - m_2$; e.g. Dariush et al. 2007), the concentration of dark matter haloes (e.g. Neto et al. 2007; Okabe et al. 2010a) and the sub-halo population of dark matter haloes (e.g. Taylor & Babul 2004; Zentner et al. 2005).

The luminosities of the first- and second-ranked galaxies in clusters were first studied, as far as we are aware, by Sandage & Hardy (1973, see also Geller & Peebles 1976; Ostriker & Hausman 1977; Tremaine & Richstone 1977; Oergerle & Hoessel 1989). More recently the luminosity gap, Δm_{12}, was studied in the context of galaxy groups, the term 'fossil' being coined to describe virialized systems with $\Delta m_{12} \geq 2$ (Ponman et al. 1994; Jones, Ponman & Forbes 2000; Jones et al. 2003). $L^\star$ galaxies are absent from fossil groups, which were thus interpreted as having formed at early times, with dynamical friction then having sufficient time to cause the $L^\star$ galaxy population to merge and form the brightest group galaxy (BGG). Fossil groups are expected to be more common than fossil clusters, at least in part because the probability of galaxy–galaxy merging is anticorrelated with galaxy velocity and thus with cluster mass. Nevertheless, two clusters with masses of $\sim 10^{14}\,\mathrm{M}_\odot$ have been found with $\Delta m_{12} > 2$ (Cypriano, Mendes de Oliveira & Sodre 2006; Khosroshahi et al. 2006a; Mendes de Oliveira, Cypriano & Sodre 2006). These two objects (RX J1416.4+2315, and RX J15552.2+2013) raise the interesting question of whether the most massive ($\gtrsim 10^{15}\,\mathrm{M}_\odot$) clusters might host similarly dominant brightest cluster galaxies (BCGs). Theoretical studies suggest that this is the case; for example, Milosavljević et al. (2006) and Dariush et al. (2007) predict that ∼1–3 and ∼5 per cent of $10^{15}\,\mathrm{M}_\odot$ clusters have $\Delta m_{12} \geq 2$ respectively.

The concentration parameter of a dark matter halo describes the shape of its density profile following the so-called universal profile proposed by Navarro, Frenk & White (1997, hereafter NFW) and variants thereon. Haloes with smaller concentrations have a flatter density profile, while larger concentrations imply a steeper density profile. Bullock et al. (2001) analysed numerical simulations of cold dark matter (CDM) universes finding a weak dependence of concentration on halo mass: $c \propto M^\alpha$ with $\alpha \simeq -0.1$ (see also Dolag et al. 2004; Neto et al. 2007; Duffy et al. 2008). This relationship arises from the relative timing of the formation of dark matter haloes as a function of mass. On average, less massive haloes form at earlier times than more massive haloes in a hierarchical universe. The universe was denser at earlier times than at later times, and thus the central regions of less massive haloes are relatively dense, leading to higher concentration parameters, than for more massive haloes. Some observational studies have reported very high concentration parameters in individual systems; for example, one of the fossil groups studied by Khosroshahi, Jones & Ponman (2004) and Khosroshahi et al. (2006a) was found to have $c > 50$, based on modelling of X-ray observations. Lensing studies of several individual strong-lensing clusters have also obtained very high concentrations of $c \sim 10$ in contrast to the theoretical prediction of $c \sim 5$ (Gavazzi et al. 2003; Kneib et al. 2003; Broadhurst et al. 2005; Limousin et al. 2007). More recently, observational studies have begun to study larger samples and thus to constrain the concentration–mass relation itself and to probe the general population rather than a small number of potentially extreme objects (e.g. Buote et al. 2007; Okabe et al. 2010a).

Theoretically, substructures within dark matter haloes, i.e. the sub-halo population, are also sensitive to the assembly history of the host dark matter halo (e.g. Taylor & Babul 2004; Zentner et al. 2005). Observationally, substructures in galaxy clusters can be identified via detailed modelling of the observed gravitational lensing signal (Smith et al. 2005, 2009; Richard et al. 2010a,b). Specifically, group- and galaxy-scale perturbers are required to achieve statistically acceptable fits to the strong-lensing data. The contribution of these structures to the total cluster mass is quantified via the 'substructure fraction', $f_{\rm sub}$, defined as the amount of mass within the adopted cluster-centric radius that is assigned to substructures divided by the total cluster mass within the same aperture. Smith & Taylor (2008) combined Smith et al.'s (2005) observational measurements of $f_{\rm sub}$ for 10 X-ray luminous galaxy clusters with Taylor & Babul's (2004) semi-analytic model of structure formation to explore the interpretation of lensing-based measurements of $f_{\rm sub}$. The main conclusion was that $f_{\rm sub}$ depends on both when the cluster formed and the level of recent mass assembly, each defined as the lookback time when each cluster had acquired 50 and 90 per cent of their observed mass, respectively. For example, clusters at $z = 0.2$ with $f_{\rm sub} < 0.1$ formed on average at $z \gtrsim 0.8$ and suffered $\leq$10 per cent mass growth since $z = 0.4$; in contrast, clusters at $z = 0.2$ with the highest substructure fractions ($f_{\rm sub} \gtrsim 0.4$) formed on average since $z \simeq 0.4$ and acquired $\gtrsim$10 per cent of their mass between $z = 0.25$ and $z = 0.2$, i.e. a time interval of just 500 Myr.

A complementary view of hierarchical merging within galaxy clusters is available from BCG morphology. Based on their isophotal shapes, elliptical galaxies have been classified as discy or boxy (Bender et al. 1989), with positive and negative fourth-order Fourier coefficients, respectively. The interpretation of boxy and discy isophotes in terms of the details of galaxy merger histories is a controversial subject (Faber et al. 1997; Naab & Burkert 2003; Khochfar & Burkert 2005). In this study, we will sidestep these difficulties and concentrate simply on discy/boxy isophotes as an indicator of the presence of gas that has dissipated and settled into a disc-like structure, either because the last massive galaxy to merge with the BCG was gas rich or because gas has been accreted by the merger product from its environment, e.g. by a BCG in a cool core cluster. BCGs are the most massive early-type galaxies and are generally expected to have boxy isophotes consistent with formation via mergers of early-type (gas-poor) galaxies (Lin & Mohr 2004; Khosroshahi, Ponman & Jones 2006b). However, BGGs in some fossil groups are as bright as BCGs and do *not* have boxy isophotes (Khosroshahi et al. 2006b), suggesting that (i) fossil BGGs may form early from the mergers between gas-rich spiral galaxies and (ii) some fossil BGGs may have subsequently evolved into BCGs.

The main aim of this paper is to combine measurements of the luminosity gap, cool core strength, concentration, substructure fraction and BCG isophotal shape for a large sample of clusters to assemble an empirical picture of the hierarchical assembly of clusters and their BCGs. We present the first observational measurement of the distribution of the luminosity gap statistic of $10^{15}\,\mathrm{M}_\odot$ clusters and compare this distribution with the other probes of hierarchical assembly discussed above. This allows us to build an empirical

picture with which to assess the usefulness of the respective measurements for estimating the age of clusters. We also investigate how well the current generations of galaxy formation models can reproduce the observed luminosity gap distribution. The data used for this study are drawn from the Local Cluster Substructure Survey (LoCuSS; PI: Smith; http://www.sr.bham.ac.uk/locuss). A summary of LoCuSS is provided in Section 2, together with a description of the data used in this paper. The analysis and results are then presented in Section 3 and compared with theoretical predictions in Section 4. The main conclusions are summarized and discussed in Section 5. We assume $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$ and $\Omega_\Lambda = 0.7$ throughout. In this cosmology, 1 arcsec corresponds to a physical scale of 3.3 kpc, at $z = 0.2$. All photometric measurements are relative to Vega.

2 DATA

2.1 Sample selection and observing strategy

LoCuSS is a morphologically unbiased multiwavelength survey of X-ray luminous galaxy clusters at $0.15 \leq z \leq 0.3$. The overall aim of the survey is to measure the cluster–cluster scatter in key observables, such as the X-ray temperature, the Y_X parameter (Smith et al. 2005; Zhang et al. 2008; Okabe et al. 2010b), the Sunyaev–Zeldovich effect Y parameter (Marrone et al. 2009) and the obscured and unobscured star formation activity (Haines et al. 2009a,b, 2010; Pereira et al. 2010; Smith et al. 2010), and to correlate this scatter with the structure and thus hierarchical assembly history of the clusters. The backbone of the survey is the gravitational lensing analysis of *Hubble Space Telescope* (*HST*; Smith et al. 2005; Richard et al. 2010b; Hamilton-Morris et al., in preparation; May et al., in preparation) and Subaru (Oguri et al. 2010; Okabe et al. 2010a) imaging data, because the lensing-based mass maps can be used to infer the likely assembly history of the clusters (Smith & Taylor 2008).

The parent sample for this study comprises 115 clusters satisfying $-27^\circ \leq \delta \leq 70^\circ$, $0.15 \leq z \leq 0.3$, $n_\mathrm{H} \leq 7 \times 10^{20}\,\mathrm{cm^{-2}}$ drawn from the *ROSAT* All-Sky Survey catalogues (Ebeling et al. 1998, 2000; Böhringer et al. 2004). The cut at $\delta = -27^\circ$ ensures that the clusters are observable from Palomar Observatory at elevations above 30° (an airmass of $\sec z \leq 2$). The clusters span a decade in X-ray luminosity in the 0.1–2.4 keV band: $2 \times 10^{44} \lesssim L_\mathrm{X} \lesssim 20 \times 10^{44}\,\mathrm{erg\,s^{-1}}$ (Fig. 1), which corresponds to a mass range of $5 \times 10^{14} \lesssim M_\mathrm{virial} \lesssim 3 \times 10^{15}\,\mathrm{M_\odot}$ (Reiprich & Böhringer 2002), i.e. well matched to $10^{15}\,\mathrm{M_\odot}$.

The radius at which the mean enclosed density of a $10^{15}\,\mathrm{M_\odot}$ dark matter halo at the median redshift of the cluster sample ($z = 0.22$) is $\langle\rho(< r)\rangle = 200\rho_\mathrm{crit}$ is $r_{200} = 1.9$ Mpc. To ensure that our results are comparable with previous studies of the luminosity gap statistic, data that probe out to $\sim 0.5 r_{200}$ ($\sim$4.5 arcmin) are required. This requirement is met by the Wide-field Infrared Camera (WIRC) on the 200-inch Hale Telescope at Palomar Observatory (Section 2.2). Traditionally, the luminosity gap statistic has been studied at optical wavelengths. In contrast, working in the near-infrared permits the use of $(J - K)$ colours as a surrogate for a photometric redshift estimate of cluster galaxies (Section 3.1), taking advantage of the relative insensitivity of near-infrared colours to spectral type (e.g. Mannucci et al. 2001). This is vital, in the absence of exhaustive spectroscopic catalogues, to weed out non-cluster members when calculating the luminosity gap.

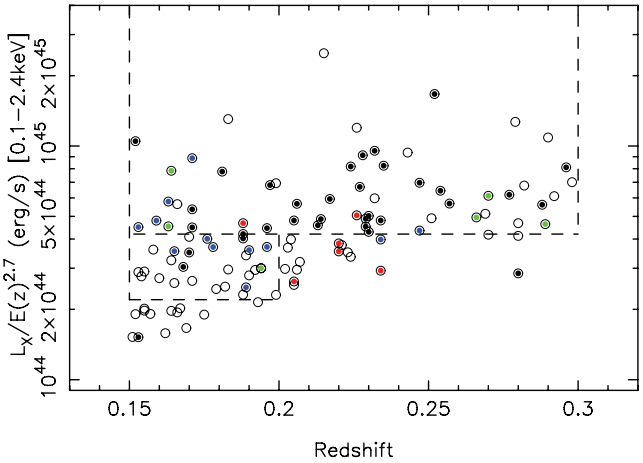

Figure 1. The distribution of the parent sample of 115 clusters in the L_X–redshift plane. Filled data points indicate clusters that we observed in acceptable conditions (see Section 2.2) with WIRC on the 200-inch Hale Telescope and are colour-coded as follows: black – also observed with both *HST* and *Chandra*; blue – also observed with *HST*, but not with *Chandra*; green – also observed with *Chandra*, but not with *HST*; red – observed with neither *HST* nor *Chandra*. Open data points were not observed with WIRC at Palomar and therefore do not form part of the sample studied in this paper. The dashed lines delineate the volume-limited samples against which the observed sample is compared statistically in Section 2.5. The absence of clusters from the parent sample to the lower right is caused by the flux limit of the *ROSAT* All-Sky Survey.

2.2 Ground-based near-infrared data

The parent sample of 115 clusters was used as a back-up observing programme during observing runs with WIRC (Wilson et al. 2003) on the 200-inch Hale Telescope[1] spanning from 2004 April to 2005 July. Data were acquired when the full width at half-maximum (FWHM) of point sources exceeded 1 arcsec. In total, data were obtained on 78 clusters, with no pre-selection on cluster properties other than the X-ray selection described above (Section 2.1). Each cluster was observed with a single 8.7×8.7 arcmin^2 WIRC pointing. BCGs at $z \simeq 0.2$ have a typical angular extent of $\sim$1 arcmin; the individual exposures were therefore dithered within a box of a full width of 80 arcsec to minimize the inclusion of BCG flux in sky-flats constructed from the science data. Each cluster was observed for a total of 600 s per filter, split into five dither positions.

The data were reduced in a uniform and standard manner using an automated pipeline of IRAF tasks to dark-subtract, flat-field, align and co-add the individual frames at the telescope. Data acquired in conditions worse than FWHM = 1.5 arcsec suffered strongly variable transparency and/or a non-uniform background and were therefore excluded from the analysis, leaving a total of 59 clusters with good quality data (Table 1). Astrometric and photometric calibration were achieved by reference to the Two-Micron All-Sky Survey (2MASS) catalogues, to root mean square (rms) precisions of 0.1 arcsec and 0.1 mag, respectively (Stott et al. 2008). The results described in this paper are insensitive to the uncertainty on the photometric calibration. An example $(J - K)$–K colour–magnitude diagram is shown in Fig. 2. The typical depth reached by the data is $K \simeq 17$; an $L^\star$ galaxy has $K \simeq 15$ and a typical BCG has $K \sim 12$–13 at $z \simeq 0.2$.

[1] The Hale Telescope at Palomar Observatory is owned and operated by the California Institute of Technology.

Table 1. The observed sample of clusters.

Cluster	α, δ [J2000]	Redshift	$M_{K,\mathrm{BCG}}{}^{a}$	$\Delta m_{12}{}^{a}$	*HST* PID	Also known as
A68	00 37 05.28 +09 09 10.8	0.255	−26.65	0.25	8249	
A115	00 55 50.65 +26 24 38.7	0.197	−26.29	0.39	11312	
A141	01 05 37.17 −24 40 49.7	0.230	−26.54	0.40	10881	RXC J0105.5−2439
ZwCl0104.4+0048	01 06 49.50 +01 03 22.1	0.255	−26.37	0.52	11312	Z348
A209	01 31 53.00 −13 36 34.0	0.206	−26.76	0.87	8249	RXC J0131.8−1336
A267	01 52 48.72 +01 01 08.4	0.230	−26.56	1.43	8249	RXC J0152.7+0100
A291	02 01 43.11 −02 11 48.1	0.196	−25.49	1.79	8301	RXC J0201.7−0212
A383	02 48 02.00 −03 32 15.0	0.188	−26.24	1.76	8249	RXC J0248.0−0332
RXC J0331.1−2100	03 31 05.87 −21 00 32.7	0.188	−27.07	0.96	10881	
A521	04 54 06.88 −10 13 24.6	0.247	−26.47	0.00	11312	RXC J0454.1−1014
A586	07 32 20.42 +31 37 58.8	0.171	−26.31	0.51	8301	
ZwCl0740.4+1740	07 43 23.16 +17 33 40.0	0.189	−26.26	1.62	11312	Z1432
A611	08 00 55.92 +36 03 39.6	0.288	−26.90	1.28	9270	
A665	08 30 57.36 +65 51 14.4	0.182	−25.81	0.66		
ZwCl0839.9+2937	08 42 56.06 +29 27 25.7	0.194	−26.45	0.67	11312	Z1883
ZwCl0857.9+2107	09 00 36.86 +20 53 40.0	0.235	−25.82	0.34	8301	Z2089
A750	09 09 12.74 +10 58 29.1	0.163	−26.52	0.95	11312	
A773	09 17 54.00 +51 42 57.6	0.217	−26.68	0.47	8249	
ZwCl0923.6+5340	09 27 10.69 +53 27 30.9	0.205	−25.86	0.55		Z2379
ZwCl0949.6+5207	09 52 47.52 +51 53 27.6	0.214	−26.45	1.24	8301	Z2701
A901	09 56 26.40 −10 04 12.0	0.163	−25.90	0.08	10395	RXC J0956.4−1004
RX J1000.5+4409	10 00 31.16 +44 08 42.5	0.153	−25.72	0.35	10881	
A963	10 17 01.20 +39 01 44.4	0.205	−26.42	1.73	8249	
A1201	11 12 54.61 +13 26 08.2	0.169	−26.25	1.37	8719	
A1204	11 13 20.55 +17 35 39.1	0.171	−25.80	0.95	8301	
A1246	11 23 58.83 +21 28 45.4	0.190	−25.92	0.42	8301	RXC J1123.9+2129
A1423	11 57 17.43 +33 36 38.6	0.213	−26.16	1.77	8719	
A1553	12 30 48.95 +10 32 45.6	0.165	−26.82	0.94		
ZwCl1231.4+1007	12 34 17.45 +09 45 58.1	0.229	−26.21	0.71	8719	Z5247
A1634	12 54 01.84 −06 42 14.4	0.196	−25.82	0.61		RXC J1254.0−0642
A1682	13 06 47.89 +46 33 32.5	0.226	−26.96	0.09	8719	
ZwCl1309.1+2216	13 11 46.15 +22 01 36.8	0.266	−26.18	2.17		Z5768
A1704	13 14 24.38 +64 34 31.0	0.220	−26.34	1.41		
A1758	13 32 44.47 +50 32 30.5	0.280	−26.26	0.22		
A1763	13 35 16.32 +40 59 45.6	0.228	−26.84	1.59	8249	
A1835	14 01 02.40 +02 52 55.2	0.253	−27.32	2.44	8249	
A1914	14 25 59.78 +37 49 29.1	0.171	−26.62	1.33	8301	
A1961	14 44 31.85 +31 13 34.3	0.234	−26.47	0.31		
A1994	14 56 13.48 −05 48 56.6	0.220	−26.45	0.55		RXC J1456.3−0549
MS 1455.0+2232	14 57 15.23 +22 20 34.0	0.258	−26.06	0.11	8301	ZwCl1454.8+2233, Z7160
A2009	15 00 19.63 +21 22 08.9	0.153	−25.99	0.17	8301	
ZwCl1459.4+4240	15 01 23.13 +42 20 39.6	0.290	−26.28	0.07		Z7215
A2111	15 39 40.51 +34 25 27.0	0.229	−25.58	0.37		
A2146	15 56 09.05 +66 21 33.1	0.234	−26.07	0.41	8301	
A2163	16 15 34.10 −06 07 26.0	0.169	−25.46	0.49		
A2204	16 32 46.94 +05 34 31.3	0.152	−25.82	0.11	8301	
A2218	16 35 52.80 +66 12 50.4	0.171	−26.12	0.32	5701	
A2219	16 40 22.56 +46 42 21.6	0.228	−26.62	1.18	6488	
A2254	17 17 45.96 +19 40 48.0	0.178	−26.05	0.88	8301	
RX J1720.1+2638	17 20 10.14 +26 37 30.9	0.164	−26.37	1.76	11312	
A2261	17 22 27.24 +32 07 56.7	0.224	−26.34	2.32	8301	
RXC J2102.1−2431	21 02 09.98 −24 32 01.8	0.188	−26.82	2.04		
A2345	21 27 13.73 −12 09 46.1	0.176	−26.67	1.09	11312	RXC J2127.1−1209
RX J2129.6+0005	21 29 40.02 +00 05 20.9	0.235	−26.78	1.93	8301	
A2390	21 53 36.72 +17 41 31.2	0.233	−26.21	1.50	5352	
RXC J2211.7−0350	22 11 45.95 −03 49 45.3	0.270	−26.35	1.71		
A2485	22 48 31.13 −16 06 25.6	0.247	−26.59	0.00	11312	RXC J2248.5−1606
A2537	23 08 23.20 −02 11 31.0	0.297	−26.19	0.63	9270	RXC J2308.3−2011
A2631	23 37 39.82 +00 16 16.9	0.278	−26.35	0.64	11312	RXC J2337.6+0016

[a]Uncertainties on $M_{K,\mathrm{BCG}}$ and Δm_{12} are dominated by the uncertainties on the photometric calibration, which is ∼0.1 mag in J and K bands.

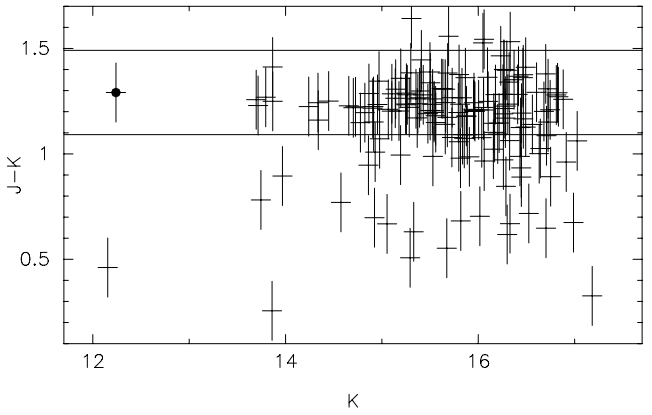

Figure 2. The $(J - K)/K$ colour–magnitude relation for A1763. The horizontal lines show the region within which likely cluster galaxies were selected. The filled circle denotes the BCG.

2.3 *Hubble Space Telescope* observations

HST[2] imaging data are available through a broad red filter (F606W, F702W and/or F814W) for 45 of the 59 clusters (Table 1) of which 13 are drawn from new LoCuSS Advanced Camera for Surveys (ACS; PID:10881) and Wide Field Planetary Camera 2 (WFPC2; PID:11312) observations. The reduction of the data on 10 clusters observed under PID:5701, PID:6488 and PID:8249 is described by Smith et al. (2005). Of the remaining 35 clusters, 18 clusters with WFPC2 data (PIDs:5352, 8301, 8719, 11312) were all reduced on to a 0.1-arcsec pixel scale using WFIXUP, WMOSAIC, IMSHIFT and CRREJ tasks within IRAF to clean, register and combine the individual exposures. Details of the reduction of the remaining clusters observed with ACS are described by Hamilton-Morris et al., (in preparation).

2.4 *Chandra* X-ray observations

Chandra X-ray observations are available for 41 of the total sample of 59 clusters. The reduction and analysis of these data are described in detail by Sanderson, Edge & Smith (2009). In brief, for each cluster, an annular spectral profile was extracted and used to deproject the X-ray emission to measure the gas density and temperature in spherical shells. The phenomenological cluster model of Ascasibar & Diego (2008) was then jointly fitted to the temperature and density profiles to determine the mass profile, assuming hydrostatic equilibrium, following the procedures described in Sanderson & Ponman (2010). The model is based on a Hernquist (1990) density profile, which yields larger scale radii (and correspondingly lower mass concentrations) than the commonly used NFW profile. Following Sanderson et al. (2009), we also use the logarithmic slope of the gas density profile at $0.04r_{500}$ (α; Vikhlinin et al. 2007) as an indicator of cool core strength, which has also been shown to correlate with the substructure fraction of cluster cores, based on strong lens models (Richard et al. 2010b). A more negative value of α indicates a steeper central gas density profile, and thus a stronger cool core, and vice versa.

[2] Based on observations with the NASA/ESA *HST* obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

Table 2. Statistical comparison of sub-samples.

Sample	$N_{\rm clus}$	$\langle \log_{10}(L_{\rm X}) \rangle^a$
All clusters observed with WIRC	59	44.71 ± 0.03
Clusters observed with WIRC and *HST*	45	44.71 ± 0.03
Clusters observed with WIRC and *Chandra*	41	44.76 ± 0.04
Clusters observed with WIRC and *Spitzer*	39	44.73 ± 0.03
Mean of 100 000 samples drawn randomly from a volume-limited sample	59	44.71 ± 0.03

[a]The uncertainties are errors on the mean X-ray luminosity of each sample, with the exception of the last row, in which we quote the standard deviation of the 100 000 samples around the mean luminosity of all of these randomly drawn samples.

2.5 Statistical comparison of sub-samples

Incomplete coverage of the parent sample of 115 clusters with WIRC and heterogeneous coverage of the WIRC-observed clusters with other facilities (Fig. 1) may introduce subtle biases into our results. We therefore compare statistically the various observed sub-samples, including for completeness the sub-sample for which *Spitzer* data are available (Section 3.2). Specifically, the cluster X-ray luminosities are compared, after correction for the modest redshift evolution within the sample due to the expansion of the universe: $L_{{\rm X},z} = L_{\rm X} E(z)^{-2.7}$, where $E(z) = H(z)/H_0 = [\Omega_{\rm M}(1 + z)^3 + \Omega_\Lambda]^{0.5}$ following Evrard et al. (2002).

The mean X-ray luminosity of the full sample of 59 clusters is statistically indistinguishable from the mean luminosity of the sub-samples observed at other wavelengths (Table 2). We also draw 100 000 samples of 59 clusters at random from the combined volume-limited samples defined by $0.15 < z < 0.2$, $2.2 \times 10^{44}\,{\rm erg\,s^{-1}} \leq L_{{\rm X},z} \leq 4.2 \times 10^{44}\,{\rm erg\,s^{-1}}$ and $0.15 < z < 0.3$, $L_{{\rm X},z} \geq 4.2 \times 10^{44}\,{\rm erg\,s^{-1}}$ (see Fig. 1). The average X-ray luminosity of the observed sample of 59 clusters is well within 1 standard deviation of the average X-ray luminosity of these randomized samples (Table 2). We therefore conclude that the $L_{\rm X}$ distributions of the full sample of 59 clusters, the sub-samples observed with other telescopes and the volume-limited sample defined above are all statistically indistinguishable from each other. We therefore expect any biases to be negligible and that our results can be treated as comparable with those that would be achieved with a volume-limited sample. We also take care to double-check that the Δm_{12} distributions of the various observational sub-samples are statistically indistinguishable from each other in Section 3.

3 ANALYSIS AND RESULTS

3.1 Source detection and photometry

The J- and K-band frames were analysed with SExtractor (Bertin & Arnouts 1996), extracting all objects subtending >25 pixels at a signal-to-noise ratio of $>2.5\,{\rm pixel^{-1}}$. The resulting catalogues were matched using a search radius comparable with the seeing disc, and point sources were excluded based on the stellarity index calculated by SExtractor. In the absence of spectroscopic redshift information, we rely on the red ridge line of galaxies seen in the $(J - K)/K$ colour–magnitude diagrams for each cluster (e.g. Fig. 2) to isolate likely cluster galaxies. A simple model based on redshifting local galaxy spectral templates (King & Ellis 1985) confirms that the $(J - K)$ colour of galaxies varies by $\lesssim 0.2$ mag between E/S0 and Scd spectral types. We therefore selected galaxies within ± 0.2 mag

of the BCG colour in each cluster as likely cluster members (see horizontal lines in Fig. 2).

The extended envelope of the BCGs typically spans a diameter of ∼1 arcmin in the *HST* frames. In contrast, BCGs typically span just ∼20–30 arcsec in the near-infrared frames. The difference is due to the brighter sky in the near-infrared relative to the optical. We therefore use the deep F702W *HST*/WFPC2 data available for 10 clusters in our sample (Smith et al. 2005) to estimate the K-band flux lost due to the bright K-band sky, under the assumption that the $(R_{702} - K)$ colour of BCGs does not vary significantly with radius on large scales. This assumption introduces negligible systematic uncertainty into our results because colour gradients in elliptical galaxies are measured to be $\mathrm{d}(R-K)/\mathrm{d}(\log r) \sim 0.3$–$0.4$ (La Barbera et al. 2004, 2010), which translates into a possible ∼0.1 mag systematic error on the factor of 2 radial corrections to the K-band photometry estimated below.

After masking out other galaxies from the data, the BCG R_{702}- and K-band light distributions are modelled using ELLIPSE in the STSDAS package in IRAF. The R_{702}-band model is then used to extrapolate the K-band light distribution out to 2σ above the mean local background. The same procedure was applied to a sample of $L^\star$ galaxies detected in the WFPC2 frame of each of these 10 clusters. This analysis revealed that reliance on solely K-band data causes the total flux of BCGs to be under-estimated by ∼0.3–0.7 mag, with a median of ∼0.45 mag. This effect is much less severe for non-BCGs, with the total flux being under-estimated by ∼0.07–0.15 mag, with a median of 0.1 mag. We fit a straight line to these data: $\Delta K = \alpha + \beta K$, obtaining $\alpha = -1.67 \pm 0.43$ and $\beta = 0.11 \pm 0.03$. The correction, ΔK, was then applied to all galaxies within our sample. The amplitude of this systematic correction to the luminosity gap statistic measurements is therefore $\beta \Delta m_{12}$ and is typically in the range of ∼0–0.3 mag with an uncertainty of ∼25 per cent, both of which are smaller than the bin width in our subsequent analysis. Our results are therefore not significantly affected by the uncertainties on this correction.

3.2 Luminosity gap statistic of 10^{15} $\mathrm{M}_\odot$ clusters

In the absence of models of the mass distribution and thus measurements of r_{200} for all clusters in the sample, we adopt a fixed projected physical radius of $R = 640$ kpc within which to calculate Δm_{12} for each cluster. This aperture fits comfortably within the observed field of view for all clusters and corresponds to $\sim 0.4 r_{200}$ for a $M_{\mathrm{virial}} \simeq 10^{15}\,\mathrm{M}_\odot$ cluster at $z = 0.2$. The distribution of the luminosity gap statistic is shown in Fig. 3; $p(\Delta m_{12})$ is a declining function of Δm_{12}. We therefore fit a straight line to the data: $p(\Delta m_{12}) = A + B\Delta m_{12}$, weighting the data points by σ^{-2} where σ is the Poisson uncertainty on Δm_{12} in each bin. The best-fitting parameter values are $A = 0.36 \pm 0.03$ and $B = -0.13 \pm 0.02$. We also measure the fraction of 'fossil clusters': a total of four clusters have $\Delta m_{12} \geq 2$, yielding a fraction of $10^{15}\,\mathrm{M}_\odot$ clusters satisfying this selection of $p(\Delta m_{12} \geq 2) = 0.07^{+0.05}_{-0.03}$, where the error bar is at 1σ using binomial statistics (Gehrels 1986).

Following Dariush et al. (2007), we also show in Fig. 3 the Δm_{12} distribution derived from a Monte Carlo simulation in which galaxies were drawn at random from a Schechter function with $M^\star = -24.5$ and $\alpha = -1.2$, adopted from a fit of the Schechter function to the K-band galaxy luminosity function of the Millennium semi-analytic catalogue and is also consistent with observed luminosity functions (e.g. Lin, Mohr & Stanford 2004). This simulation allows us to identify whether the Δm_{12} distribution presents any excess probability over random statistical sampling of a common underlying luminosity function. Excess probability over random is only found at $\Delta m_{12} \gtrsim 1$. We measure the observed probability of a cluster of having a luminosity gap of $\Delta m_{12} \geq 1$ to be $p(\Delta m_{12} \geq 1) = 0.37 \pm 0.08$, compared with the estimated probability based on the Monte Carlo simulation of $p_{\mathrm{MC}}(\Delta m_{12} \geq 1) = 0.13$. We therefore detect an excess probability over random sampling at $\Delta m_{12} \geq 1$ of ∼0.24 at $\sim 3\sigma$ significance and conclude that the Δm_{12} distribution at $\Delta m_{12} > 1$ has a physical origin.

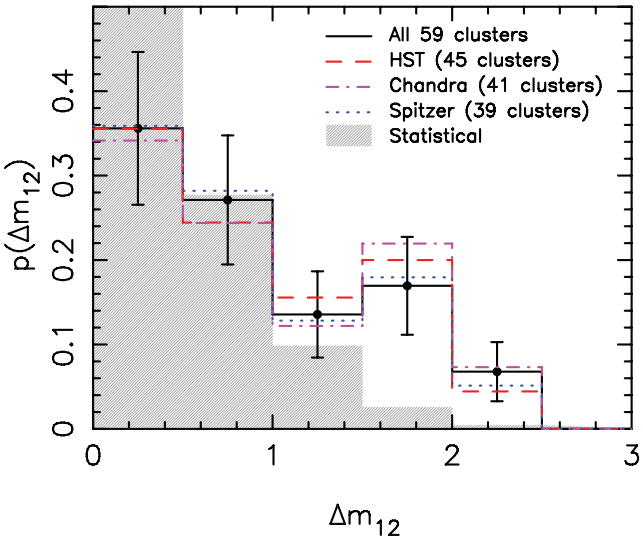

Figure 3. Distribution of the observed luminosity gap (black points and solid line). The grey filled histogram is the expected distribution if the galaxies are drawn at random from a Schechter function following Dariush et al. (2007) (see Section 3.2 for more details). The dashed, dot–dashed and dotted histograms show the Δm_{12} distributions of the sub-samples of clusters for which *HST*, *Chandra* and *Spitzer* data are available.

The Δm_{12} distributions of the sub-samples of clusters for which *HST*, *Chandra* and *Spitzer* data are available are statistically consistent with that of the full sample of 59 clusters (Fig. 3). Two sample Kolmogorov–Smirnov (KS) tests that compare the *HST*, *Chandra* and *Spitzer* sub-samples in turn with the full sample confirm that the probability of the respective sub-samples being drawn from a different underlying Δm_{12} distribution than the full sample is $P \leq 0.1$ per cent in all cases, with the largest difference between the cumulative distributions being $D = 0.0748$, between the *Chandra* sub-sample and the full sample.

We also look at how the absolute magnitudes of the first- and second-ranked cluster galaxies vary with Δm_{12} (Fig. 4). The luminosity of the first-ranked galaxy increases very slowly with Δm_{12}, remaining in the range $-27 \lesssim M_K \lesssim -26$ across the full range of Δm_{12}. In contrast, the luminosity of the second-ranked galaxy declines from $M_K \sim -26$ at $\Delta m_{12} \sim 0$ to $M_K \sim -24$ at $\Delta m_{12} \sim 2$. We characterize these trends by fitting the following relations to the data: $M_{K,1} = \alpha_1 + \beta_1 \Delta m_{12}$ and $M_{K,2} = \alpha_2 + \beta_2 \Delta m_{12}$, where the numerical subscripts denote the first- and second-ranked galaxies, respectively. The best-fitting values are $\alpha_1 = -26.13 \pm 0.02$, $\beta_1 = -0.22 \pm 0.02$ and $\alpha_2 = -26.13 \pm 0.02$, $\beta_2 = 0.78 \pm 0.02$. Empirically large luminosity gap statistics are therefore due to both an over-bright BCG, $M_{K,1}(\Delta m_{12} = 2) - M_{K,1}(\Delta m_{12} = 0) \simeq -0.4$, and an under-bright second-ranked galaxy, $M_{K,2}(\Delta m_{12} = 2) - M_{K,2}(\Delta m_{12} = 0) \simeq 1.6$. The relative faintness of second-ranked galaxies in large luminosity gap clusters supports the idea that the growth of dominant BCGs is driven by the merging of luminous cluster galaxies with the BCG. Indeed, the current star

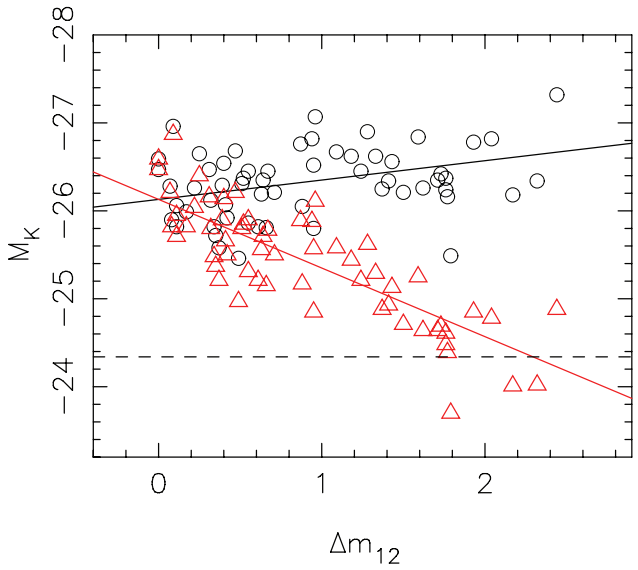

Figure 4. Absolute K-band magnitude of the first-ranked (black circles) and second-ranked (red triangles) galaxies as a function of luminosity gap. The solid black and red lines show the best-fitting straight line to the data (see Section 3.2 for more details). The horizontal dashed line is at $M_K = -24.34$, the absolute magnitude of an $L^\star$ galaxy, taken from Lin et al. (2004).

formation rate (SFR) of BCGs discussed above lends additional support – the BCG in a cluster with a luminosity gap of $\Delta m_{12} = 2$ is six times more luminous and has a stellar mass of $\sim 10^{11}\,\mathrm{M}_\odot$ more than the second-ranked galaxy. Just two of the four clusters with $\Delta m_{12} \gtrsim 2$ in Fig. 5 host an active BCG. The most active of these, A1835, is forming stars at SFR $= 125\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ (Egami et al. 2006) and the other, RX J2129.6+0005, is forming stars at SFR $= 14\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ (Quillen et al. 2008). These two BCGs would therefore

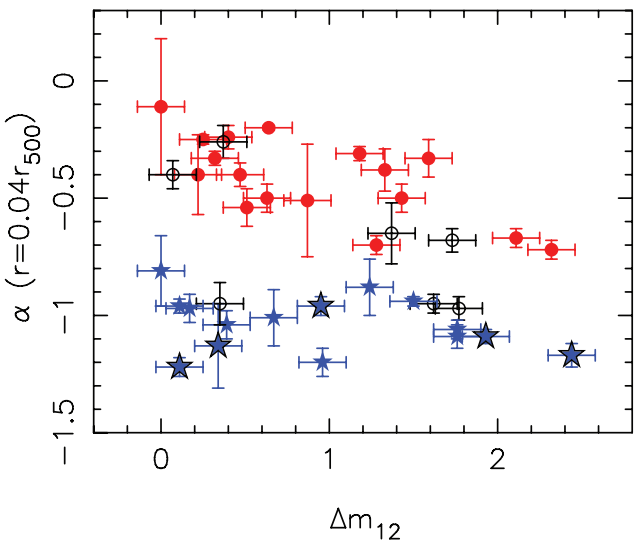

Figure 5. The gradient of the logarithmic gas density profile at $0.04r_{500}$ versus luminosity gap for 41 clusters that have also been observed with *Chandra* and studied by Sanderson et al. (2009). Blue stars correspond to clusters with an Hα emitting BCG (see Sanderson et al. 2009); blue stars with a black outline have also been identified as hosting a BCG that is forming stars at SFR $\gtrsim 10\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ using *Spitzer*/MIPS observations; filled red circles denote clusters with BCGs that are not Hα emitters and are forming stars at SFR $< 10\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$; open black circles indicate clusters that have not been observed with *Spitzer*.

have to form stars continuously at this rate for $\sim 10^9$ and $\sim 10^{10}$ yr, respectively, for their large luminosity gap to be caused exclusively by gas cooling and consequent star formation.

Finally, we note that on average second-ranked galaxies in clusters with $\Delta m_{12} > 2$ have $\langle M_{K,2} \rangle = -24.6 \pm 0.6$ where the uncertainty is the rms scatter around the mean. Lin et al. (2004) measured $M_K = -24.34 \pm 0.01$ for $L^\star$ cluster galaxies at $z \leq 0.1$, in agreement with similar studies of field galaxies and of higher redshift clusters (De Propris et al. 1999; Cole et al. 2001). The distribution of luminosities of second-ranked cluster galaxies in clusters with $\Delta m_{12} > 2$ is therefore statistically consistent with them being $L^\star$ galaxies. This contrasts with low-mass $\Delta m_{12} > 2$ systems, i.e. fossil groups, in that $L^\star$ galaxies are absent from low-mass systems. This difference is probably due, at least in part, to the relative inefficiency of galaxy merging in massive clusters.

3.3 Comparing luminosity gap with cool core strength

To explore further the physical origin of large luminosity gaps we plot Δm_{12} versus α, the slope of the logarithmic gas density profile at $0.04r_{500}$, for 41 clusters that have also been observed with *Chandra* in Fig. 5. The measurements of α are based on Sanderson et al.'s (2009) analysis of the *Chandra* data (Section 2.4). At $\Delta m_{12} \simeq 0$, the clusters span the full range of cool core strengths: $-1.2 \lesssim \alpha \lesssim -0.1$. This dynamic range shrinks to just $-1.2 \lesssim \alpha \lesssim -0.6$ at $\Delta m_{12} \gtrsim 2$ – the clusters with large luminosity gaps also host relatively strong cool cores. We also identify star-forming BCGs in Fig. 5. It has long been known that the Hα emission from the BCG is closely associated with the presence of significant central cooling in the cluster core (e.g. Heckman 1981; Crawford et al. 1999). More recently, Sanderson et al. (2009) found in their sample of 65 clusters that Hα emitting BCGs occur exclusively in those clusters with the most cuspy inner gas density profiles ($\alpha < -0.85$) and where the projected offset between the X-ray centroid and the BCG is $\leq 0.02r_{500}$. The same is true of the five BCGs with an SFR of $\gtrsim 10\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, based on mid-infrared observations with *Spitzer*/MIPS – this SFR corresponds to a flux of ~ 1 mJy from a BCG at $z \simeq 0.2$. These measurements are drawn from the literature (Egami et al. 2006; Quillen et al. 2008) and from our own measurements using data from cycle 4 (PID:40827, PI: Smith; PID: 41011, PI: Egami), the details of which will be published elsewhere (Egami et al., in preparation). Fig. 5 therefore confirms that cool core clusters tend to host actively star-forming BCGs (e.g. Edge et al. 1999; Egami et al. 2006; Quillen et al. 2008). However, cool core clusters ($\alpha \lesssim -1$) with active BCGs (SFR $\gtrsim 10\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ and/or Hα emission) are found across the full range of Δm_{12} in Fig. 5.

These results are consistent with the interpretation of large luminosity gap clusters as objects that formed relatively early and subsequently developed a large luminosity gap through the merging of bright cluster galaxies with the BCG. A similarly long period of time – a few Gyr – is required to form a strong cool core following cluster formation. Conversely, if all clusters with smaller luminosity gaps formed more recently than those with larger gaps, and thus have had insufficient time to form a large luminosity gap and a cool core, then they should all host relatively weak cool cores. However, this is not the case. This can be understood if the so-called fossil status of a large luminosity gap cluster is not the endpoint of its evolution. If bright ($L > L^\star$) galaxies fall into a cool core 'fossil' cluster, then that cluster would move immediately from lower right to lower left in Fig. 5. As the in-falling system (presumably a group) reaches the cluster core ~ 1 Gyr later, it may disrupt partially

or fully the cooling of gas on to the BCG and cause the cluster to move vertically in the Δm_{12}–α plane. This scenario naturally explains the triangular distribution of points in Fig. 5 and is consistent with hierarchical infall (i.e. mergers) playing a role in regulating cooling in cluster cores.

To place this discussion on a more quantitative footing, we adopt a strategy that we return to often in Section 3 – we split the sample into low- ($\Delta m_{12} < 1$) and high-Δm_{12} ($\Delta m_{12} > 1$) sub-samples and perform a two-sample KS test on the cumulative distribution of the other variable, in this case α. The hypothesis that high-Δm_{12} clusters are drawn from the same underlying α distribution as low-Δm_{12} clusters is rejected at just 74 per cent confidence, i.e. slightly over 1σ significance, based on a maximum difference between the cumulative α distributions of $D = 0.3102$. In the absence of a decisive test, we therefore divide the sample at $\Delta m_{12} = 1.5$, i.e. a more extreme value of Δm_{12}, attempting to identify roughly the luminosity gap at which the α distribution diverges from that of lower Δm_{12} clusters. This time the two-sample KS test rejects the null hypothesis at 95 per cent confidence – i.e. 2σ – based on a maximum difference between the respective cumulative α distributions of $D = 0.472$.

3.4 Comparing the luminosity gap with BCG morphology

We use the high angular resolution *HST* imaging observations of the 45 clusters discussed in Section 2.3 to measure the isophotal shape of the BCGs in these clusters. The ELLIPSE task in IRAF was used to measure the fourth Fourier coefficient (B_4) of the light distribution. This coefficient indicates whether the galaxy has a discy or boxy shape (Bender 1988). In Fig. 6, we show the B_4 profile of four BCGs to illustrate the diversity within the sample. Following Bender et al. (1989), we tried to use the extremum value of B_4 (i.e. $B_{4,\rm ext}$ in Table 3) to classify galaxies as either discy ($B_{4,\rm ext} > 0$) or boxy ($B_{4,\rm ext} < 0$). If the B_4 profile passes through a stationary point, then the extremum is obtained by finding the maximum or minimum value of B_4 in the radial range enclosed by three times the FWHM of point sources and the effective radius derived from a de Vaucouleurs profile fit. In the absence of a stationary point, the extremum value of B_4 is the value at the effective radius, under the assumption that B_4 is a monotonic function of radius. However B_4 is in general not a monotonic function of radius for BCGs in our sample, even for those that have, on average, boxy and discy isophotes (Fig. 6). For these reasons, the isophotal shapes of 22 out of 45 BCGs cannot be classified based on $B_{4,\rm ext}$. We also find some clusters (e.g. A1204; see Fig. 6) in which B_4 is consistent with zero across the full radial range of the data.

We therefore implement a modified scheme, in which we calculate the error-weighted mean value of B_4 in the same radial range as above, with no weighting of the bins to account for the variation of the bin solid angle as a function of radius. BCGs with $\langle B_4 \rangle$ consistent with zero within the uncertainties were classified as elliptical, otherwise BCGs are classified as boxy or discy if $\langle B_4 \rangle < 0$ or $\langle B_4 \rangle > 0$, respectively. Finally, a BCG is 'unclassified' if the error on $\langle B_4 \rangle$ is comparable with the dynamic range of the data, i.e. $\geq$0.01. BCG morphologies derived under both Bender et al.'s 'extremum' scheme and our own 'mean' scheme are listed in Table 3 along with the boxy/discy/elliptical/unclassified classification based on each method. The respective methods agree on morphological classification for 16 of the 22 BCGs for which classification was possible under both methods. However, only three of the six discrepant BCGs have $\langle B_4 \rangle$ and $B_{4,\rm ext}$ values that formally disagree between the methods within the quoted uncertainties – A521, A750 and ZwCl0949.6+5207. The important advantage of our method is that classification is possible for an additional 15 BCGs that were unclassifiable under the Bender et al. scheme. We therefore adopt $\langle B_4 \rangle$ as our measure of BCG morphology for all clusters with *HST* data for the reasons outlined above regarding the general absence of clearly defined stationary points and monotonic behaviour of the B_4 profiles.

In summary, out of 45 clusters, 10 are classified as boxy, 13 as discy, 14 as elliptical and eight are unclassified. In Fig. 7, we plot $\langle B_4 \rangle$ versus Δm_{12}, the most striking feature of which is the lack of clusters with large Δm_{12} and negative $\langle B_4 \rangle$, i.e. boxy BCGs appear not to live in large luminosity gap clusters. As in Section 3.3, we split the clusters into low-Δm_{12} ($\Delta m_{12} \leq 1$) and high-Δm_{12} ($\Delta m_{12} > 1$) samples and perform a two-sample KS test. The low- and high-Δm_{12} samples contain 27 and 18 clusters, respectively, with a maximum difference between their cumulative $\langle B4 \rangle$ distributions of $D = 0.3567$. The hypothesis that the low- and high-Δm_{12} samples are drawn from the same underlying $\langle B_4 \rangle$ distribution is therefore disfavoured at 91 per cent confidence, i.e. 1.7σ. Unlike the situation for the analysis of the α distributions of high- and low-Δm_{12} clusters in Section 3.3, the significance with which the null hypothesis is rejected does not increase if the sub-samples are redefined by splitting the full sample at $\Delta m_{12} = 1.5$. This is obvious from a comparison of Figs 5 and 7 and suggests that the relationship between Δm_{12} and BCG morphology is stronger than between Δm_{12} and cool core strength.

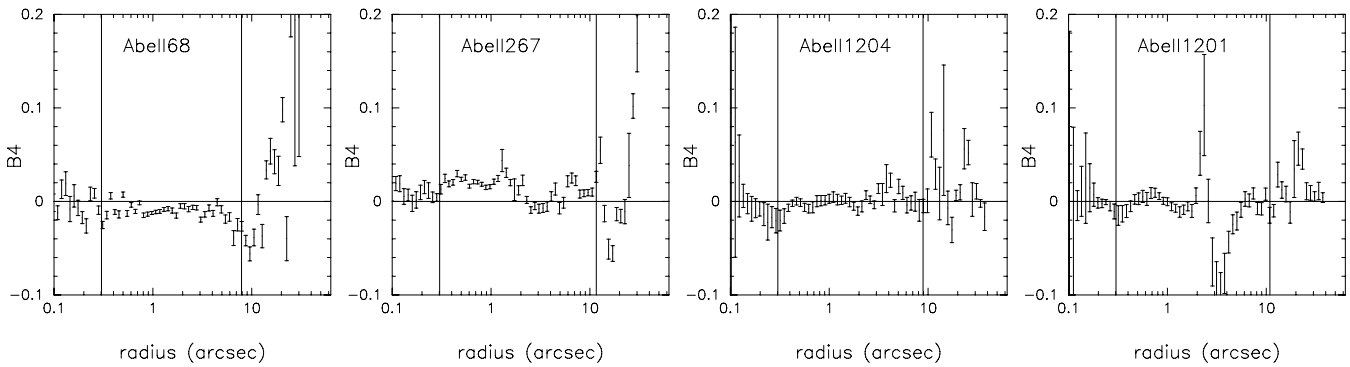

Figure 6. Example isophotal shape profiles. From left to right, the BCGs are classified as boxy, discy, pure ellipse and unclassified. The vertical line at the left is set to three times the FWHM of point sources and the one at the right indicates the half-light radius of the BCG. Note that to keep the analysis simple and conservative, no flux was masked out of the *HST* data. So, for example, the BCG in A1201 was unclassified because of the impact of the gravitational arc at a BCG-centric radius of ∼2 arcsec (Edge et al. 2003) on the isophotal analysis.

Table 3. Results from isophotal analysis of BCGs using *HST* data.

Cluster	Effective radius (arcsec)	Extremum method $100B_{4,\mathrm{ext}}$	Classification	Mean method $100\langle B_4\rangle$	Classification
A68	7.9	-2.0 ± 0.4	Boxy	-0.99 ± 0.19	Boxy
A115	4.1	–	Unclassified	$+0.96 \pm 0.22$	Discy
A141	3.6	$+3.4 \pm 0.9$	Discy	$+1.09 \pm 0.19$	Discy
ZwCl0104.4+0048	4.1	–	Unclassified	-0.14 ± 0.80	Elliptical
A209	7.4	$+1.8 \pm 0.5$	Discy	$+0.45 \pm 0.07$	Discy
A267	11.6	$+3.5 \pm 1.2$	Discy	$+0.99 \pm 0.17$	Discy
A291	7.3	$+2.1 \pm 0.6$	Discy	$+0.67 \pm 0.24$	Discy
A383	8.3	–	Unclassified	-0.20 ± 0.97	Unclassified
RXC J0331.1−2100	3.1	–	Unclassified	-2.69 ± 7.50	Unclassified
A521	3.9	-3.1 ± 1.0	Boxy	-0.44 ± 0.79	Elliptical
A586	7.7	–	Unclassified	-0.16 ± 0.64	Elliptical
ZwCl0740+1740	5.5	$+0.1 \pm 0.2$	Elliptical	-0.11 ± 0.06	Boxy
A611	3.5	–	Unclassified	-0.21 ± 0.15	Boxy
ZwCl0839.9+2937	3.5	–	Unclassified	$+0.84 \pm 0.50$	Discy
ZwCl0857.9+2107	4.6	-4.0 ± 1.1	Boxy	-1.49 ± 0.61	Boxy
A750	4.0	$+5.0 \pm 1.5$	Discy	$+0.11 \pm 1.45$	Unclassified
A773	6.8	-2.1 ± 0.4	Boxy	-0.50 ± 0.18	Boxy
ZwCl0949.6+5207	7.1	$+2.4 \pm 0.7$	Discy	$+0.22 \pm 0.30$	Elliptical
A901	3.2	$+1.1 \pm 0.3$	Discy	$+0.50 \pm 0.10$	Discy
RX J1000.5+4409	3.1	–	Unclassified	-0.77 ± 0.18	Boxy
A963	14.3	–	Unclassified	$+0.46 \pm 0.19$	Discy
A1201	10.8	–	Unclassified	-0.26 ± 0.48	Elliptical
A1204	8.9	0.0 ± 0.7	Elliptical	-0.10 ± 0.21	Elliptical
A1246	8.6	–	Unclassified	-0.14 ± 1.78	Unclassified
A1423	7.4	–	Unclassified	-0.91 ± 60.6	Unclassified
ZwCl1231.4+1007	4.4	-1.3 ± 0.5	Boxy	-0.56 ± 0.19	Boxy
A1682	6.2	0.0 ± 0.5	Elliptical	-0.43 ± 0.27	Boxy
A1763	8.2	0.0 ± 0.6	Elliptical	$+0.04 \pm 0.08$	Elliptical
A1835	6.8	–	Unclassified	$+1.10 \pm 0.60$	Discy
A1914	12.2	0.0 ± 0.5	Elliptical	$+0.25 \pm 0.15$	Discy
MS 1455.0+2232	5.0	–	Unclassified	-0.47 ± 0.25	Boxy
A2009	9.3	–	Unclassified	$+0.14 \pm 0.19$	Elliptical
A2146	6.3	–	Unclassified	-0.20 ± 0.50	Elliptical
A2204	7.9	–	Unclassified	$+0.08 \pm 1.20$	Unclassified
A2218	8.1	$+2.3 \pm 0.8$	Discy	$+0.83 \pm 0.29$	Discy
A2219	8.5	–	Unclassified	$+0.83 \pm 0.19$	Discy
A2254	9.6	–	Unclassified	-0.03 ± 0.36	Elliptical
RXJ 1720.1−2638	5.5	–	Unclassified	$+0.28 \pm 1.39$	Unclassified
A2261	7.2	–	Unclassified	$+0.39 \pm 7.20$	Unclassified
A2345	7.8	0.0 ± 0.5	Elliptical	$+0.03 \pm 0.09$	Elliptical
RX J2129.6+0005	9.6	–	Unclassified	-0.16 ± 0.60	Elliptical
A2390	3.5	$+3.5 \pm 0.5$	Discy	$+1.42 \pm 0.32$	Discy
A2485	3.9	-0.5 ± 0.6	Elliptical	$+0.19 \pm 0.63$	Elliptical
A2537	5.2	0.0 ± 0.3	Elliptical	-0.11 ± 0.18	Elliptical
A2631	6.0	0.0 ± 0.5	Elliptical	-0.24 ± 0.12	Boxy

3.5 Comparing the luminosity gap with cluster concentration

To investigate the possibility that high-Δm_{12} clusters formed at earlier times than low-Δm_{12} clusters, we explore the relationship between Δm_{12} and the shape of the cluster dark matter haloes via the concentration parameter. In Fig. 8 we plot Δm_{12} versus concentration, c_{500} for the 41 clusters with the available *Chandra* data (Sanderson et al. 2009; Section 2.4). The c_{500}–Δm_{12} distribution is similar to the $\langle B4\rangle$–Δm_{12} distribution in that the lower right of both plots is empty and that clusters with $\Delta m_{12} < 1$ span the full dynamic range in the vertical axis. To quantify this we again perform a two-sample KS test, on the $\Delta m_{12} < 1$ and $\Delta m_{12} > 1$ sub-samples. In this case the low- and high-Δm_{12} samples contain 24 and 17 clusters, respectively, with a maximum difference between their cumulative c_{500} distributions of $D = 0.2574$. Acceptance/rejection of the null hypothesis that low- and high-Δm_{12} clusters are drawn from the same underlying c_{500} distribution therefore has roughly equal probability. However if we modify the definition of the low- and high-Δm_{12} sub-samples by splitting the full sample at $\Delta m_{12} = 1.5$, we are able to reject the null hypothesis at $\sim 1.7\sigma$. We therefore conclude that the c_{500}–Δm_{12} plane qualitatively supports the interpretation of the $\langle B4\rangle$–Δm_{12} plane; however, statistically this is not decisive. Specifically, clusters with a large luminosity gap tend to have a relatively large concentration parameter, although there is a curious deficit of clusters with $\Delta m_{12} \gtrsim 1.8$ and $c_{500} \gtrsim 1.5$. Clusters with lower luminosity gaps plausibly comprise both

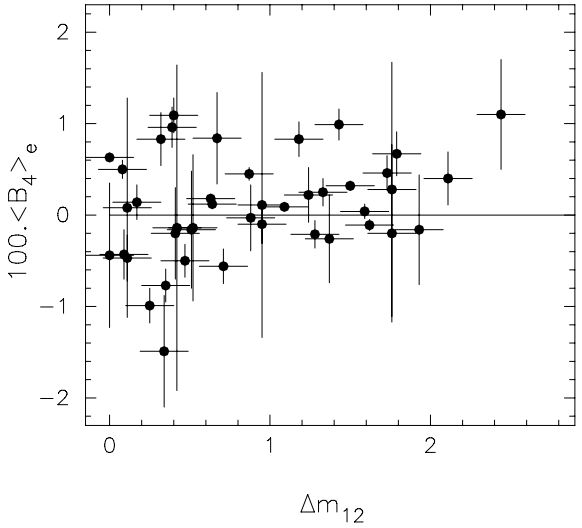

Figure 7. Luminosity gap statistic (Δm_{12}) versus error-weighted mean fourth Fourier component of the BCG light distribution ($\langle B_4 \rangle$). Positive values of $\langle B_4 \rangle$ correspond to discy BCGs, negative values correspond to boxy BCGs and values consistent with zero are consistent with elliptical isophotes. Clusters with $\Delta m_{12} \lesssim 1$ host BCGs with both boxy and discy isophotes. In contrast, clusters with $\Delta m_{12} \gtrsim 1$ host only non-boxy (i.e. elliptical or discy BCGs).

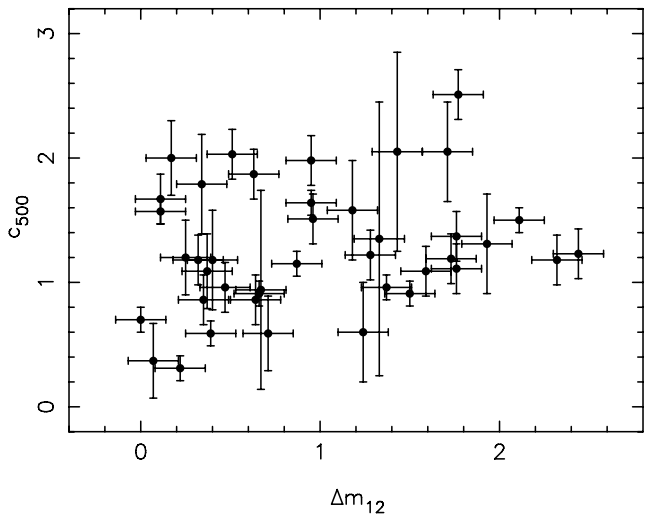

Figure 8. Concentration c_{500} versus luminosity gap for 41 clusters for which the X-ray-based mass profiles are available from Sanderson et al.'s (2009) analysis of archival *Chandra* data.

clusters that formed more recently than clusters with large gaps – and thus have lower concentration parameters – and clusters that used to have a large luminosity gap, and thus formed early, and have a higher concentration parameter but that then suffered infall of bright ($L > L^\star$) galaxies. Put another way, the existence of clusters in the top-left corner of Fig. 8 is consistent with the time-scale on which the concentration parameter of a cluster may be reset following a cluster–cluster merger being long compared with the infall time-scale of $\sim$1 Gyr.

3.6 Comparing the luminosity gap with cluster substructure

Measurements of the substructure fraction ($f_{\rm sub}$), i.e. the fraction of the total cluster mass that resides in substructures, are available for

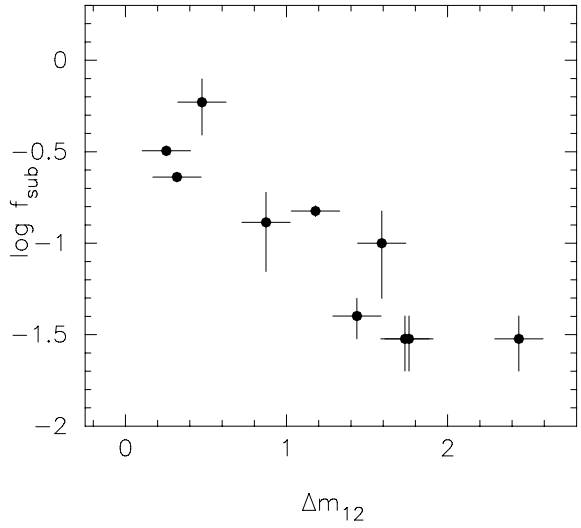

Figure 9. Luminosity gap versus substructure fraction measured within $R \leq 250\,h^{-1}$ kpc in 10 clusters from our sample by Smith et al. (2005).

10 of the clusters (Smith et al. 2005) from our sample of 59. Smith et al.'s gravitational lens models include mass components that account explicitly for substructures required to reproduce the observed positions of multiply imaged background galaxies – these substructures comprise both galaxy group and individual galaxy masses. We plot $f_{\rm sub}$ versus Δm_{12} for these 10 clusters in Fig. 9, revealing a relationship between these quantities in the sense that clusters with simpler gravitational potentials (low $f_{\rm sub}$) have more dominant BCGs (high Δm_{12}) and vice versa. To quantify this relationship, we fit a simple model to the data: $\log f_{\rm sub} = \mu + \nu \Delta m_{12}$, and obtain best-fitting parameters of $\mu = -0.29 \pm 0.15$ and $\nu = -0.58 \pm 0.11$. This result is consistent with that found by Richard et al. (2010b), despite the smaller aperture of 250 kpc used in their study. This consistency arises because the typical projected separation of the first- and second-ranked galaxies in our sample is $\lesssim$250 kpc. We also double-check that the Δm_{12} distribution of the 10 clusters in Fig. 9 is consistent with that of the full sample, finding a maximum difference between the cumulative Δm_{12} distributions of $D = 0.2414$, indicating roughly equal probability of rejection/acceptance of the hypothesis that the two samples are drawn from different underlying populations. These results are qualitatively consistent with those of Ramella et al. (2007).

3.7 Summary

We now summarize the comparison of our luminosity gap measurements with other probes of the structure, and thus the age and assembly history of clusters, and discuss the interpretation of these results.

The clearest empirical relationship found is between Δm_{12} and $f_{\rm sub}$ in the sense that clusters with a dominant BCG ($\Delta m_{12} > 1$) have a lower substructure fraction ($f_{\rm sub} < 0.1$) and vice versa. The strong correlation between Δm_{12} and $f_{\rm sub}$ is in stark contrast with the triangular distributions of clusters in the Δm_{12}–α, Δm_{12}–$\langle B_4 \rangle$ and Δm_{12}–c_{500} planes. A simple physical interpretation of the Δm_{12}–$f_{\rm sub}$ relation is that both quantities are sensitive to the same thing. As galaxies and groups of galaxies fall into clusters, the light emitted by the galaxies will cause Δm_{12} to either decrease or stay the same, depending on how bright the infalling galaxies are. At the same time, the total mass of these galaxies and the group-scale haloes within which they may be embedded causes $f_{\rm sub}$ to increase. Early

studies discussed the idea that galaxy groups with $\Delta m_{12} \geq 2$ may have formed at earlier times than groups with $\Delta m_{12} < 2$. However, more recently, a variety of studies have shown that both Δm_{12} and $f_{\rm sub}$ are correlated with *both* the formation epoch of the host dark matter halo *and* the recent hierarchical assembly history of the halo (Dariush et al. 2007, 2010; Smith & Taylor 2008). Therefore, both theoretical and observational studies across a broad range of dark matter halo mass are converging on the view that 'fossil' status is not an endpoint in the evolution of galaxy systems that formed early. Rather it is a phase that a galaxy system can evolve through if it formed early and then suffered minimal hierarchical infall after the formation of a bright massive central galaxy. The triangular distributions of clusters in the Δm_{12}–α, Δm_{12}–$\langle B_4 \rangle$ and Δm_{12}–c_{500} planes are all consistent with this interpretation and inconsistent with the idea that fossil galaxy systems are evolutionary cul-de-sacs. Specifically, if a cluster forms early and then sufficient time elapses for a large luminosity gap to form via merging of $L^\star$ gas-rich galaxies to form the BCG, and for a cool core to form, then this cluster will reside in the top-right corner of Figs 7 and 8 and the bottom right of Fig. 5. If a $\gtrsim L^\star$ galaxy then falls into the cluster, either on its own or in a group, then the cluster would move leftwards in all of Figs 5, 7 and 8 as soon as the infalling galaxy system crosses the aperture within which Δm_{12} is measured (in our case $0.4r_{200}$). Several Gyr later the infalling structure will reach the centre of the cluster, and its merger with the cluster may be sufficiently energetic to modify the strength of the cluster cool core, the shape of the BCG and the concentration of the cluster dark matter halo. In this way, clusters can move vertically in Figs 5, 7 and 8 and produce the observed triangular distribution of clusters.

The interpretation of non-boxy morphologies ($\langle B_4 \rangle \geq 0$) of BCGs in clusters with large luminosity gaps is an important element of the above discussion. Khochfar & Burkert (2005) showed that the morphology of early-type galaxies is sensitive to the morphology (indicative of gas content) of their progenitors and subsequent gas infall. The straightforward interpretation of the observables is therefore that dominant BCGs formed from mergers of gas-rich (presumably spiral) galaxies and/or have accreted gas since the last major merger in their assembly history. Formation of dominant BCGs from gas-rich progenitors is consistent with the early formation of these BCGs as discussed above, because at earlier times the galaxies from which BCGs formed would have been more gas rich than at later times.

To disentangle the relative contribution of gas-rich mergers and accretion of gas to the discy shape of some BCGs we plot α in Fig. 10, the slope of the logarithmic gas density profile at $0.04r_{500}$ versus $\langle B_4 \rangle$. If BCG morphology is strongly influenced by gas cooling on to the BCG, then one would expect a relationship between α and $\langle B_4 \rangle$ in the sense that discy BCGs ($\langle B_4 \rangle > 1$) would live in clusters with a steep central ($\alpha < -0.5$) gas density profile. This is because clusters with steep central gas density profiles host a cool core – i.e. a central positive temperature gradient, absence of an entropy core and a cooling time-scale short compared with the age of the universe. However, we do not find any strong relationship between $\langle B_4 \rangle$ and α in Fig. 10. We divide the 41 clusters with *Chandra* data into those with the strongest cool cores – $\alpha < -0.9$ – and the rest. A two-sample KS test on these two sub-samples yields a maximum difference between the cumulative $\langle B_4 \rangle$ distributions of $D = 0.209$, indicating roughly equal probability of accepting/rejecting the hypothesis that the two distributions are drawn from the same underlying distributions.

In the absence of a strong relationship between cool core strength and BCG morphology, we therefore conclude that BCG morphology

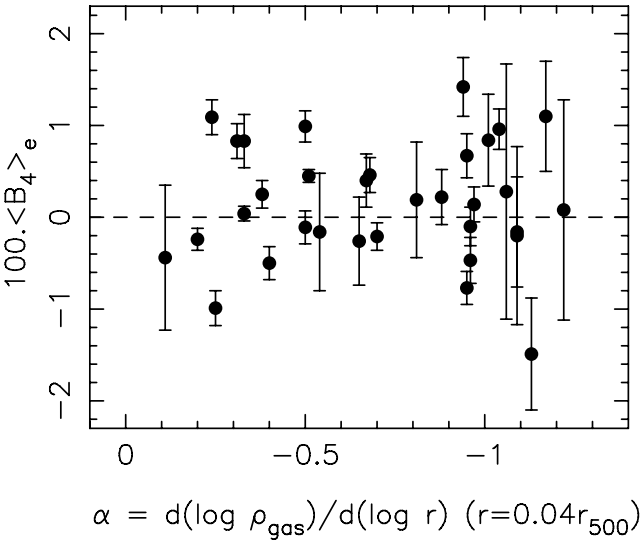

Figure 10. Strength of the cool core in each cluster, as measured by α, the slope of the logarithmic gas density profile at $0.04r_{500}$ from Sanderson et al. (2009) versus the $\langle B_4 \rangle$. The absence of a relationship between α and $\langle B_4 \rangle$ suggests that discy BCG isophotes are more likely caused by such BCGs being formed from mergers between gas-rich galaxies than by cooling of gas on to the BCG. The typical error bar on α is $\lesssim 0.1$.

is more sensitive to the gas content of the galaxies that merged to form it than to the subsequent gas accretion history of the BCG. This view is consistent with the comparison of α, Δm_{12} and BCG activity in Fig. 5 and with the discussion of the dependence of $M_{K,2}$ on Δm_{12} in Section 3.2, the key point being that the merging of luminous cluster galaxies to form the BCG appears to have a much stronger influence on the luminosity gap than gas cooling and subsequent star formation within BCGs.

4 COMPARISON WITH THEORETICAL PREDICTIONS

Modern galaxy formation and evolution models contain physical prescriptions for many physical processes relevant to the formation and evolution of galaxies, including dynamical friction, conversion of cold gas into stars during galaxy mergers and AGN feedback. These processes are particularly important in the centres of galaxy clusters where, for example, AGN feedback is thought to regulate the cooling of gas on to, and thus star formation in, BCGs. However, the models were not constrained by the luminosity gap distribution; our observational results can therefore provide a strong test of the models.

We compare our observations with the Bower et al. (2006), Croton et al. (2006) and de Lucia & Blaizot (2007) semi-analytic models, all of which are based on the Millennium Simulation[3] – a cosmological numerical simulation of dark matter in a volume spanning $500\,h^{-1}$Mpc containing $\sim 10^{10}$ particles. An important difference between the models is that the Bower et al. model implements 'quasar' mode AGN feedback, whereas the Croton et al. and de Lucia & Blaizot models implement 'radio' mode AGN feedback. We also note that de Lucia & Blaizot compared their model

[3] The Millennium Simulation used in this paper was carried out by the Virgo Supercomputing Consortium at the Computing Centre of the Max-Planck Society in Garching. The semi-analytic galaxy catalogue is publicly available at http://www.mpa-garching.mpg.de/galform/agnpaper.

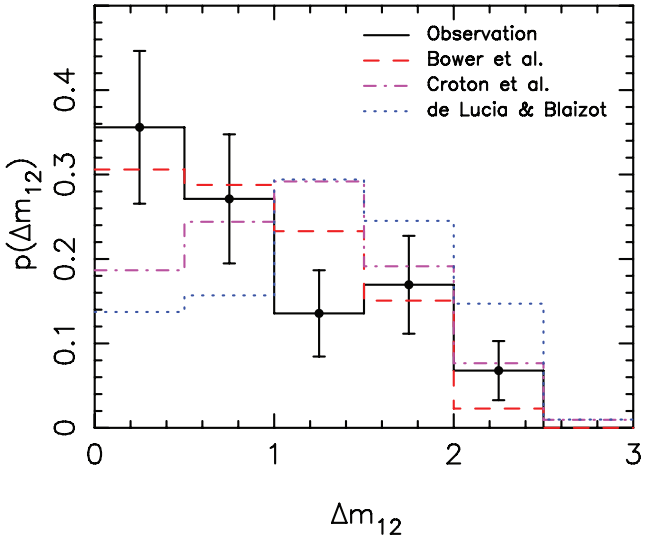

Figure 11. Distribution of the observed luminosity gap (black points – see also Fig. 3) compared with the same for clusters with $M_{\rm virial} \geq 5 \times 10^{14}\,{\rm M}_{\odot}$, measured within a projected BCG-centric radius of 640 kpc using the Millennium Simulation-based semi-analytic galaxy formation models of Bower et al. (2006), Croton et al. (2006) and de Lucia & Blaizot (2007). The error bar on each bin in the theoretical histograms is comparable with the observational errors.

predictions with the observed properties of BCGs; however, they did not compare with observed luminosity gaps.

First, we select dark matter haloes from the Millennium dark matter friends-of-friends catalogue. Within the whole simulated volume, 209 haloes were found with masses greater than $5 \times 10^{14}\,{\rm M}_{\odot}$, i.e. above the mass threshold of the observed sample. We then extracted galaxies in these 209 haloes from the semi-analytic galaxy catalogues based on each of the three models. The K-band luminosity gap was computed for each halo within a projected cluster-centric radius of 640 kpc. The predicted luminosity gap statistic distributions are over-plotted on the observed distribution in Fig. 11.

The observed Δm_{12} distribution is consistent within the uncertainties with a monotonically declining function of Δm_{12} (Section 3.2). The Bower et al. model matches this observational result well, and the predicted fraction of clusters with the most extreme luminosity gaps is $p(\Delta m_{12} \geq 2) = 0.02^{+0.02}_{-0.01}$, just $\sim$1.3σ below the observed fraction of $p(\Delta m_{12} \geq 2) = 0.07^{+0.05}_{-0.03}$ (Section 3.2). In contrast, the Croton et al. model peaks at $\Delta m_{12} \sim$ 1–1.5, i.e. it does not predict a monotonic decline of $p(\Delta m_{12})$; however, it predicts $p(\Delta m_{12} \geq 2) = 8.6^{+2.4}_{-2.0}$ per cent which is in excellent agreement with the observations. The de Lucia & Blaizot model predicts a yet more prominent peak at $\Delta m_{12} \sim$ 1–1.5 and a yet higher fraction of clusters with extreme Δm_{12}, $p(\Delta m_{12} \geq 2) = 16.7^{+4.5}_{-3.8}$ per cent, that disagrees with the observations at $\sim 2\sigma$.

Following the same approach as in Section 3.2, we also decompose the predicted Δm_{12} distributions into the predicted absolute magnitudes of the first-ranked ($M_{K,1}$) and second-ranked ($M_{K,2}$) galaxies (Fig. 12). The most striking feature of this figure is that the slopes of $M_{K,1}$ versus Δm_{12} and $M_{K,2}$ versus Δm_{12} are much steeper and shallower than the observations, respectively, in the Bower et al. model. In contrast, the Croton et al. and de Lucia & Blaizot models succeed much better in reproducing the observed trends. Interestingly, the discrepant trends in $M_{K,1}$–Δm_{12} and $M_{K,2}$–Δm_{12} within the Bower et al. model conspire to produce a distribution of Δm_{12} in Fig. 11 that is in good agreement with observations.

The absolute magnitudes of BCGs span $\sim$2 mag in the Bower et al. model, in contrast to the observed range of $\sim$1 mag. As BCGs grow, the largest increase in luminosity from purely ingesting another galaxy is a brightening by 0.75 mag, i.e. a merger between the two brightest galaxies in a cluster with $\Delta m_{12} = 0$. The very large spread in M_K for BCGs in the Bower et al. model therefore indicates that the conversion of cold gas into stars is too efficient in their model. In the model, most of the mergers that form BCGs are between gas-poor galaxies. The main source of gas for formation of new stars is that which cools from the intracluster medium. The steep relationship between $M_{K,1}$ and Δm_{12} therefore implies that AGN feedback in BCGs is too weak in the Bower et al. model. An important caveat on this interpretation is that we showed in Sections 3.4 and 3.7 that clusters with large luminosity gaps ($\Delta m_{12} \geq 1$) have non-boxy isophotes and therefore likely formed from mergers of gas-rich galaxies, i.e. probably at higher redshift than the BCGs in the model.

The shallow slope of the relationship between $M_{K,2}$ and Δm_{12} in the Bower et al. model implies that the replenishment of the supply of cluster galaxies that are ingested into their respective BCGs is too efficient in this model. Specifically, the difference in the slopes of $M_{K,2}$–Δm_{12} between the Bower et al. model and the other two models could arise from differing treatments of the merging of galaxies in the respective models following the time at which individual galaxy haloes lose their identity following ingestion into the parent cluster halo. We also comment, more generally, that the galaxies in Bower et al.'s model tend to be less luminous than the observed galaxies by $\sim$0.5 mag and those in Croton et al.'s

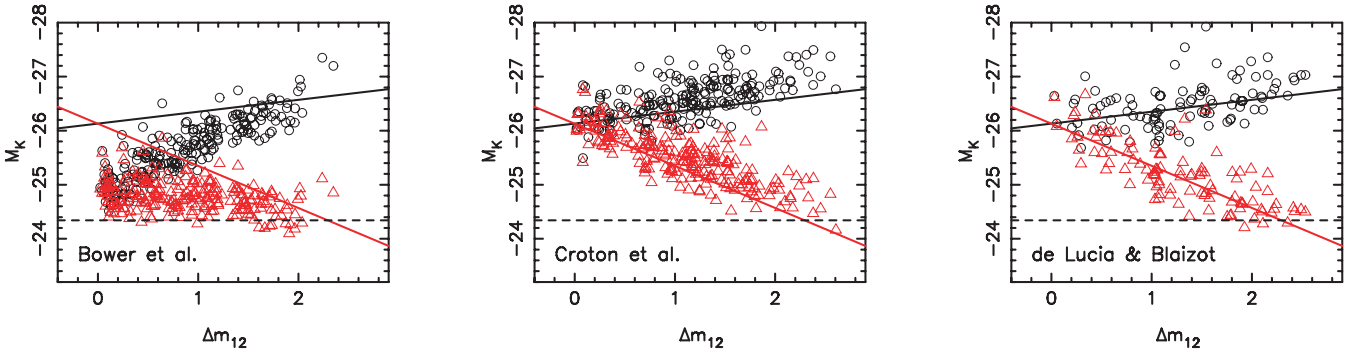

Figure 12. Absolute K-band magnitude of the first-ranked (black circles) and second-ranked (red triangles) galaxies as a function of luminosity gap from the three semi-analytic galaxy formation models discussed in Section 4. The solid black and red lines show the best-fitting straight line to the observational data shown in Fig. 4. The horizontal dashed line in each panel is at $M_K = -24.34$, the absolute magnitude of an $L^{\star}$ galaxy, taken from Lin et al. (2004).

model tend to be over-luminous by ∼0.3 mag. This suggests that the strength of feedback in the general cluster population may be too strong in the former and too weak in the latter model.

For completeness, we also compare our measurement of the fraction of $10^{15}\,M_{\odot}$ clusters that satisfy $\Delta m_{12} \geq 2$ with predictions from Milosavljević et al.'s (2006) analytic model. Our measurement of $0.07^{+0.05}_{-0.03}$ is well within 2σ of Milosavljević et al.'s prediction of 0.03. The most obvious difference between their model and our observations is that the prediction is calculated within the cluster virial radii, in contrast to our calculation within a projected cluster-centric radius of $\sim 0.4 r_{200}$. The larger volume within each cluster probed by Milosavljević et al. will reduce the probability of finding clusters with large luminosity gap statistics. The same authors also estimate the fraction of $10^{15}\,M_{\odot}$ clusters with $\Delta m_{12} \geq 2$ using data from the Sloan Digital Sky Survey (SDSS; Miller et al. 2005), obtaining a similar fraction to their prediction. The possible disagreement between this estimate and our own is harder to understand because both use a similar physical aperture for the calculation of Δm_{12}. We note, however, that the two observed cluster samples are selected in different ways; our sample is X-ray selected whilst SDSS is optically selected.

5 CONCLUSIONS

We have combined wide-field near-infrared imaging from the WIRC camera on the 200-inch Hale Telescope with *HST*, *Chandra* and *Spitzer* observations of 59 massive galaxy clusters at $z \simeq 0.2$ to explore the connections between the formation histories of BCGs and the galaxy clusters that they inhabit. This large statistical sample is intended to be representative of the underlying population of massive X-ray luminous clusters. Extensive tests confirm that results based on this sample can be regarded as statistically compatible with those from a complete volume-limited sample. Our main empirical results are as follows.

(i) We have made the first observational measurement of the distribution of the luminosity gap statistic, Δm_{12}, of massive $\sim 10^{15}\,M_{\odot}$ clusters. The probability distribution of the luminosity gap statistic is a monotonically declining function of Δm_{12}, well described by the relation $p(\Delta m_{12}) = A + B\Delta m_{12}$ with $A = 0.41 \pm 0.03$ and $B = -0.13 \pm 0.02$.

(ii) Following Dariush et al. (2007), we used Monte Carlo simulations to quantify the fraction of clusters with large luminosity gaps expected from random sampling of a Schechter function. The observed distribution exceeds the statistical distribution derived from the Monte Carlo simulation at $\Delta m_{12} \geq 1$ at $\sim 3\sigma$ significance, confirming that the most extreme luminosity gaps have a physical origin and are not statistical flukes.

(iii) Four of our sample of 59 clusters have extreme luminosity gaps of $\Delta m_{12} \geq 2$ – ZwCl1309.1+2216, A1835, A2261 and RXC J2102.1−2431, which equates to a fraction of $10^{15}\,M_{\odot}$ clusters that have $\Delta m_{12} \geq 2$ of $p(\Delta m_{12} \geq 2) = 0.07^{+0.05}_{-0.03}$.

(iv) The morphology of 45/59 BCGs was measured by analysing the shape of the BCG isophotes in archival and new *HST* observations of the cluster cores. The split between boxy, elliptical and discy isophotes is 22, 32 and 29 per cent, respectively, with 17 per cent unclassified.

(v) A strong correlation is found between Δm_{12} and $f_{\rm sub}$, the fraction of mass in the cluster cores associated with group- and galaxy-scale dark matter haloes, the latter coming from published gravitational lens models of the cluster cores (Smith et al. 2005). The relationship between Δm_{12} and $f_{\rm sub}$ is parametrized; thus, $\log f_{\rm sub} = \mu + \nu\Delta m_{12}$, with best-fitting parameters $\mu = -0.29 \pm 0.15$ ad $\nu = -0.58 \pm 0.11$.

(vi) Clusters with large luminosity gaps, $\Delta m_{12} \gtrsim 1$–1.5, have cuspy gas density profiles and thus relatively strong cool cores ($\alpha \leq -0.6$, where α is the logarithmic gas density profiles at $0.04 r_{500}$), elliptical or discy BCGs ($\langle B_4 \rangle \geq 0$, where B_4 is the fourth-order Fourier coefficient of the optical isophotes), concentrated dark matter density profiles [$c_{500} \gtrsim 1$, where c_{500} is based on a Hernquist (1990) model fit to the *Chandra* data] and small substructure fraction ($f_{\rm sub} \lesssim 0.1$, where $f_{\rm sub}$ is based on strong lens modelling of the mass distribution).

(vii) In contrast, clusters with small luminosity gaps, $\Delta m_{12} \lesssim 1$, span the full range of observed cool core strengths ($-1.3 \lesssim \alpha \lesssim 0$); span the full range of boxy, elliptical and discy BCG morphologies ($+0.015 \lesssim \langle B_4 \rangle \lesssim +0.015$); span the full range of concentrations ($c_{500} \sim 0$–2.5) and have large substructure fractions ($f_{\rm sub} \gtrsim 0.1$).

Clusters with $\Delta m_{12} \gtrsim 1$ are therefore a more homogeneous population than clusters with $\Delta m_{12} \lesssim 1$. The stronger cool cores, more concentrated mass distribution and non-boxy BCGs, all point towards high-Δm_{12} clusters forming at early times. Such early formation is required to allow sufficient time to pass for the BCG to ingest (aided by dynamical friction) the $gsL^{\star}$ galaxy population in order to develop the large luminosity gap and for the establishment of the cool core. The formation of more concentrated dark matter haloes at earlier times than less concentrated haloes is a generic prediction of CDM theory (e.g. Neto et al. 2007). The interpretation of discy BCGs is less straightforward; however, such morphologies can plausibly be interpreted as evidence for the last major mergers in a BCG's formation history comprising gas-rich galaxies – the presence of gas thus leading to the establishment of a disc-like structure in the BCG. This gas-rich merger scenario for BCG formation is consistent with the early formation of large-Δm_{12} clusters.

How can the heterogeneous population of low-Δm_{12} clusters and, more specifically, the fact that some low-Δm_{12} clusters have strong cool cores, non-boxy BCGs and high concentrations be interpreted within the context of the early formation of high-Δm_{12} clusters? The most natural explanation is that large-Δm_{12} clusters can evolve into low-Δm_{12} clusters when the supply of $gsL^{\star}$ galaxies is replenished by episodes of hierarchical infall of smaller galaxy systems, such as galaxy groups. Such infall would depress Δm_{12} and increase $f_{\rm sub}$ immediately that the group entered the measurement aperture (in this case, a clustercentric radius of $\sim 0.4 r_{200}$) and would modify other cluster properties such as the cool core strength, BCG morphology and concentration of the mass distribution on longer time-scales of several Gyr. The observed heterogeneity of low-Δm_{12} clusters can therefore be explained by these clusters comprising both (i) clusters that have formed more recently, and thus have a low concentration, haven't had time to develop a large luminosity gap and cool core, and have a BCG formed from relatively gas-poor mergers, and (ii) clusters that formerly had a large luminosity gap, and have suffered hierarchical infall in the previous few Gyr. We therefore conclude that a large luminosity gap (and large substructure fraction) is a phase through which a cluster can evolve if sufficient time elapses between episodes of hierarchical merging of other galaxies and groups of galaxies with the cluster. The large scatter seen in the theoretical age–Δm_{12} and age–$f_{\rm sub}$ relationships (Dariush et al. 2007, 2010; Smith & Taylor 2008) lends further weight to the view that both the age *and* the recent merger history of a cluster contribute to the observed values of Δm_{12} and $f_{\rm sub}$.

We also compare our observational results with predictions from Millennium Simulation-based semi-analytic models of galaxy

evolution. We find that none of the models can successfully reproduce the observations in their entirety. Bower et al. (2006) succeed best at reproducing the monotonically declining $p(\Delta m_{12})$; however, they predict a relationship between the BCG luminosity and Δm_{12} that is far too steep. In contrast, both Croton et al. (2006) and de Lucia & Blaizot (2007) predict that $p(\Delta m_{12})$ peaks at $\Delta m_{12} \sim 1$–1.5, in disagreement with the observations, with de Lucia & Blaizot predicting the more prominent peak. de Lucia & Blaizot also predict $p(\Delta m_{12} \geq 2) \sim 0.17$, in contrast to the observed value of $p(\Delta m_{12} = 0.07^{+0.05}_{-0.03})$. Nevertheless, both Croton et al. and de Lucia & Blaizot match the observed slope of the relationship between the BCG luminosity and Δm_{12} very well. We discuss the possible causes of these disagreements and suggest that Bower et al.'s model may be too efficient at converting cold gas to stars in BCGs and may also to be too efficient at replenishing the supply of galaxies in clusters.

We also note that semi-analytic galaxy evolution models also fail to reproduce observational results on high-redshift BCGs (Collins et al. 2009; Stott et al. 2010). Our new results add to this picture of the inability of models to reproduce observations of BCGs. An important strength of our results is that we do not rely on calculations of the stellar mass of BCGs and thus are insensitive to possible systematic uncertainties in stellar mass estimates for observed BCGs arising from alternative stellar population models.

Our future work on the hierarchical assembly of clusters at $z \simeq 0.2$ will take advantage of the wide-field multiwavelength data set that we are assembling, including mid/far-infrared observations with *Spitzer* and *Herschel*, joint strong/weak-lens modelling of the cluster mass distributions, our spectroscopic redshift survey of cluster galaxies with MMT /Hectospec and X-ray observations with *XMM–Newton* and *Chandra*.

ACKNOWLEDGMENTS

We acknowledge helpful comments from the anonymous referee. We thank our LoCuSS collaborators, in particular Alastair Edge, Victoria Hamilton-Morris, Jean-Paul Kneib, Yuying Zhang and Nobuhiro Okabe, for encouragement, assistance and many stimulating discussions. HGK, GPS, AJRS, TJP and JPS acknowledge support from PPARC and latterly from STFC. GPS acknowledges support from the Royal Society. GPS thanks Andrew Benson, Richard Bower, Gabriella de Lucia and Malcolm Bremer for helpful discussions and comments; Kevin Bundy, Brad Cenko, Chris Conselice, Richard Ellis, Avishay Gal-Yam, Sean Moran, David Sand and Keren Sharon for assistance with acquiring some of the near-infrared data presented in this paper; and Rick Burruss and Jeff Hickey for their support at Palomar Observatory.

REFERENCES

Ascasibar Y., Diego J. M., 2008, MNRAS, 383, 369
Bender R., 1988, A&A, 193, 7
Bender R., Surma P., Doebereiner S., Moellenhoff C., Madejsky R., 1989, A&A, 217, 35
Bertin E., Arnouts S., 1996, A&A, 117, 393
Böhringer H. et al., 2004, A&A, 425, 367
Bower R. G. et al., 2006, MNRAS, 370, 645
Broadhurst T. et al., 2005, ApJ, 619, 143
Bullock J. S. et al., 2001, MNRAS, 321, 559
Buote D. A., Gastaldello F., Humphrey P. J., Zappacosta L., Bullock J. S., Brighenti F., Mathews W. G., 2007, ApJ, 664, 123
Cole S., Norberg P., Baugh C. M., Frenk C. S., Bland-Hawthorn J., Bridges T., Cannon R., 2001, MNRAS, 326, 255
Collins C. A. et al., 2009, Nat, 458, 603
Crawford C. S., Allen S. W., Ebeling H., Edge A. C., Fabian A. C., 1999, MNRAS, 306, 857
Croton D. J. et al., 2006, MNRAS, 365, 11
Cypriano E. S., Mendes de Oliveira C., Sodre L., Jr, 2006, AJ, 132, 514
Dariush A. A., Khosroshahi H. G., Ponman T. J., Pearce F., Raychaudhury S., Hartly W., 2007, MNRAS, 382, 433
Dariush A. A., Raychaudhury S., Ponman T. J., Khosroshahi H. G., Benson A. J., Bower R. G., Pearce F., 2010, MNRAS, 405, 1873
de Lucia G., Blaizot J., 2007, MNRAS, 375, 2
de Propris R., Stanford S. A., Eisenhardt P. R., Dickinson M., Elston R., 1999, AJ, 118, 719
Dolag K., Bartelmann M., Perrotta F., Baccigalupi C., Moscardini L., Meneghetti M., Tormen G., 2004, A&A, 416, 853
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, 64
Ebeling H., Edge A. C., Bohringer H., Allen S. W., Crawford C. S., Fabian A. C., Voges W., Huchra J. P., 1998, MNRAS, 301, 881
Ebeling H., Edge A. C., Allen S. W., Crawford C. S., Fabian A. C., Huchra J. P., 2000, MNRAS, 318, 333
Edge A. C., Ivison R. J., Smail I., Blain A. W., Kneib J.-P., 1999, MNRAS, 306, 599
Edge A. C., Smith G. P., Sand D. J., Treu T., Ebeling H., Allen S. W., van Dokkum P. G., 2003, ApJ, 599, 69
Egami E. et al., 2006, ApJ, 647, 922
Evrard A. E. et al., 2002, ApJ, 573, 7
Faber S. M. et al., 1997, AJ, 114, 1771
Gavazzi R. et al., 2003, A&A, 403, 11
Gehrels N., 1986, ApJ, 303, 336
Geller M. J., Peebles P. J. E., 1976, ApJ, 206, 939
Haines C. P. et al., 2009a, MNRAS, 396, 1297
Haines C. P. et al., 2009b, ApJ, 704, 126
Haines C. P., Smith G. P., Pereira M. J., Egami E., Moran S. M., Hardegree-Ullman E., Rawle T. D., Rex M., 2010, A&A, 518, L19
Heckman T. M., 1981, ApJ, 250, 59
Hernquist L., 1990, ApJ, 356, 359
Jones L. R., Ponman T. J., Forbes D. A., 2000, MNRAS, 312, 139
Jones L. R., Ponman T. J., Horton A., Babul A., Ebeling H., Burke D. J., 2003, MNRAS, 343, 627
Khochfar S., Burkert A., 2005, MNRAS, 359, 1379
Khosroshahi H. G., Jones L. R., Ponman T. J., 2004, MNRAS, 349, 1240
Khosroshahi H. G., Maughan B., Ponman T. J., Jones L. R., 2006a, MNRAS, 369, 1211
Khosroshahi H. G., Ponman T. J., Jones L. R., 2006b, MNRAS, 372, L68
King C. R., Ellis R. S., 1985, ApJ, 288, 456
Kneib J.-P. et al., 2003, ApJ, 598, 804
Komatsu E. et al., 2009, Astronomy, 2010, 158
La Barbera F., Merluzzi P., Busarello G., Massarotti M., Mercurio A., 2004, A&A, 425, 797
La Barbera F., de Carvalho R. R., de la Rosa I. G., Gal R. R., Swindle R., Lopes P. A. A., 2010, AJ, preprint (arXiv:1006.4056)
Limousin M., Kneib J. P., Bardeau S., Natarajan P., Czoske O., Smail I., Ebeling H., Smith G. P., 2007, A&A, 461, 881
Lin Y.-T., Mohr J. J., 2004, ApJ, 617, 879
Lin Y.-T., Mohr J. J., Stanford S. A., 2004, ApJ, 610, 745
Mannucci F. et al., 2001, MNRAS, 321, 733
Marrone D. et al., 2009, ApJ, 701, L114
Mendes de Oliveira C., Cypriano E. S., Sodre L., Jr, 2006, AJ, 131, 158
Miller C. J. et al., 2005, AJ, 130, 968
Milosavljević M., Miller C. J., Furlanetto S. R., Cooray A., 2006, ApJ, 637, L9
Naab T., Burkert A., 2003, ApJ, 597, 893
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493 (NFW)
Neto A. F. et al., 2007, MNRAS, 381, 1450
Oegerle W. R., Hoessel J. G., 1989, AJ, 98, 1523
Oguri M., Takada M., Okabe N., Smith G. P., 2010, MNRAS, 405, 2215
Okabe N., Takada M., Umetsu K., Futamase T., Smith G. P., 2010a, PASJ, 62, 811

Okabe N., Zhang Y.-Y., Finoguenov A., Takada M., Smith G. P., Umetsu K., Futamase T., 2010b, ApJ, 721, 875
Ostriker J. P., Hausman M. A., 1977, ApJ, 217, 125
Pereira M. et al., 2010, A&A, 518, 40
Ponman T. J., Allan D. J., Jones L. R., Merrifield M., MacHardy I. M., 1994, Nat, 369, 462
Quillen A. C. et al., 2008, ApJS, 176, 39
Ramella M. et al., 2007, A&A, 470, 39
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Richard J. et al., 2010a, MNRAS, 402, 44
Richard J. et al., 2010b, MNRAS, 405, 325
Sandage A., Hardy E., 1973, ApJ, 183, 743
Sanderson A. J. R., Ponman T. J., 2010, MNRAS, 402, 65
Sanderson A. J. R., Edge A. C., Smith G. P., 2009, MNRAS, 398, 1698
Smith G. P., Taylor J. E., 2008, ApJ, 682, L73
Smith G. P. Kneib J., Smail I., Mazzotta P., Ebeling H., Czoske O., 2005, MNRAS, 359, 417
Smith G. P. et al., 2009, ApJ, 707, 163
Smith G. P. et al., 2010, A&A, 518, 18
Stott J. P., Edge A. C., Smith G. P., Swinbank A. M., Ebeling H., 2008, MNRAS, 384, 1502
Stott J. P. et al., 2010, ApJ, 718, 23
Taylor J. E., Babul A., 2004, MNRAS, 348, 811
Tremaine S. D., Richstone D. O., 1977, ApJ, 212, 311
Vikhlinin A., Burenin R., Forman W. R., Jones C., Hornstrup A., Murray S. S., Quintana H., 2007, in Böhringer H., Pratt G. W., Finoguenov A., Schuecker P., eds, ESO Astrophys. Symp., Heating Versus Cooling in Galaxies and Clusters of Galaxies. Springer-Verlag, Berlin, p. 48
Wilson J. C. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 451
Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., Wechsler, R. H., 2005, ApJ, 624, 505
Zhang Y.-Y., Finoguenov A., Böhringer H., Kneib J.-P., Smith G. P., Kneissl R., Okabe N., Dahle H., 2008, A&A, 482, 451

This paper has been typeset from a TEX/LATEX file prepared by the author.