University of Central Lancashire UCLan

# Central Lancashire Online Knowledge (CLoK)

| | |
|---|---|
| Title | A 3D Voronoi+Gapper Galaxy Cluster Finder in Redshift Space to z ~ 0.2. II. An Abundant Cluster Population Dominated by Late-type Galaxies Unveiled |
| Type | Article |
| URL | https://clok.uclan.ac.uk/id/eprint/25734/ |
| DOI | https://doi.org/10.3847/1538-4357/aaeff3 |
| Date | 2018 |
| Citation | Campusano, Luis E., Marinello, Gabriel, Clowes, Roger G orcid iconORCID: 0000-0001-8370-465X, Haines, Christopher P., Pereira, Sebastián, Pizarro, Daniel, Hitschfeld-Kahler, Nancy and Söchting, Ilona K. (2018) A 3D Voronoi+Gapper Galaxy Cluster Finder in Redshift Space to z ~ 0.2. II. An Abundant Cluster Population Dominated by Late-type Galaxies Unveiled. The Astrophysical Journal, 869 (2). p. 145. ISSN 0004-637X |
| Creators | Campusano, Luis E., Marinello, Gabriel, Clowes, Roger G, Haines, Christopher P., Pereira, Sebastián, Pizarro, Daniel, Hitschfeld-Kahler, Nancy and Söchting, Ilona K. |

It is advisable to refer to the publisher's version if you intend to cite from the work. https://doi.org/10.3847/1538-4357/aaeff3

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

The Astrophysical Journal, 869:145 (27pp), 2018 December 20 https://doi.org/10.3847/1538-4357/aaeff3
© 2018. The American Astronomical Society.
OPEN ACCESS

# A 3D Voronoi+Gapper Galaxy Cluster Finder in Redshift Space to $z \sim 0.2$. II. An Abundant Cluster Population Dominated by Late-type Galaxies Unveiled

Luis E. Campusano[1], Gabriel Marinello[1,2], Roger G. Clowes[3], Christopher P. Haines[4], Sebastián Pereira[5,6], Daniel Pizarro[5], Nancy Hitschfeld-Kahler[5], and Ilona K. Söchting[7]
[1] Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile; luis@das.uchile.cl
[2] Atacama Large Millimeter/submillimeter Array, Joint ALMA Observatory, Alonso de Córdova 3107, Vitacura 763-0355, Santiago, Chile
[3] Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK
[4] INAF—Osservatorio Astronomico di Brera, via Brera 28, I-20121 Milano, Italy
[5] Departamento de Ciencias de la Computación, Universidad de Chile, Casilla 2777, Santiago, Chile
[6] INRIA Chile, Av. Apoquindo 2827, Las Condes, Santiago, Chile
[7] Wärtsilä Corporation, P.O. Box 196, Helsinki, Finland
*Received 2018 January 12; revised 2018 October 27; accepted 2018 October 29; published 2018 December 20*

## Abstract

We identify 1901 galaxy clusters ($N_g \geqslant 2$) with the VoML+G algorithm (Paper I) on the Two-Degree Field Galaxy Redshift Survey. We present the 341 clusters with at least 10 galaxies that are within $0.009 < z < 0.14$ (the Catalog), of which 254 (∼75%) have counterparts in the literature (NED), with the remainder (87) plausibly "new" because of incompleteness of previous searches or unusual galaxy contents. The 207 clusters within $z = 0.04$–$0.09$ are used to study the properties of the galaxy systems in the nearby universe, including their galaxy contents parameterized by the late-type galaxy fractions ($f_L$). For this nearly complete cluster subsample, we find the following: (i) 63% are dominated by early-type galaxies (i.e., the late-type-poor clusters, $f_L < 0.5$) with corresponding mean multiplicity and logarithmic virial mass (in units of $M_\odot$) of $22 \pm 1$ and $12.91 \pm 0.04$, respectively; and (ii) 37% are dominated by late-type galaxies (i.e., the late-type-rich clusters, $f_L \geqslant 0.5$) with corresponding mean multiplicity and logarithmic virial mass (in units of $M_\odot$) of $15.7 \pm 0.9$ and $12.66 \pm 0.07$, respectively. The statistical analysis of the late-type fraction distribution supports, with a $3\sigma$ confidence level, the presence of two population components. It is suggested that the late-type-poor galaxy systems reflect and extend the class of Abell-APM-EDCC clusters and that the late-type-rich systems (∼one-third of the total) belong to a new, previously unappreciated class. The late-type-rich clusters, on average high mass-to-light ratio systems, appear to be more clustered on large scales than the late-type-poor clusters. A class of late-type-rich clusters is not predicted by current theory.

*Key words:* cosmology: observations – galaxies: clusters: general – large-scale structure of universe – methods: data analysis

*Supporting material:* machine-readable table

## 1. Introduction

Clusters of galaxies are the largest gravitationally bound systems known. They are detectable over a large fraction of the age of the universe, thus providing a testing ground for cosmology (Henry et al. 2009; Vikhlinin et al. 2009; Mantz et al. 2010; Rozo et al. 2010; Clerc et al. 2012; Benson et al. 2013), for the history of the agglomeration of matter (Jing et al. 1998; Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Zheng et al. 2005, 2009), and for the description of galaxy evolution (e.g., Dressler 1980; Lewis et al. 2002; Boselli & Gavazzi 2006; Haines et al. 2015). They also constitute one of the key probes of dark energy for ongoing and upcoming major photometric surveys (e.g., DES; Flaugher 2005; LSST Science Collaboration et al. 2009).

Nowadays galaxy clusters are searched by algorithms (e.g., Pereira et al. 2017, hereafter Paper I; Wen et al. 2018) applied on large databases coming from optical, X-ray, or millimeter-band observations. It is generally agreed that clusters are gravitationally bound systems of galaxies, which, in optical surveys, are often separated into clusters and groups of galaxies (e.g., Bahcall 2000, for a description). On a historical vein, we note that Abell (1962) has remarked that "clusters of galaxies range from rich aggregates to relatively poor groups, if these can be classed as clusters." In fact, we adopt in this work the practice by some algorithm-based searches of galaxy systems (e.g., Eke et al. 2004; Cucciati et al. 2010) of considering all the detections in a single selection procedure without an a priori separation between clusters and groups. The latter is physically justified given the evidence that dark matter dominates the mass budget of clusters, the claims of systems on different scales with extremely high mass-to-light ratios (e.g., Simon & Geha 2007; Proctor et al. 2011), and the possibility that dark matter clusters on various scales (Nussinov & Zhang 2018). Moreover, it is of interest to identify a sample of galaxy systems, without *priors* on richness or galaxy content and open to mass determinations. In this work we try to build one such cluster sample in the low-$z$ universe.

Algorithms that minimize the number of assumptions to find optical clusters may lead to new insights and constraints for cluster physics. This approach was adopted by Pereira et al. (2017), who used a two-step algorithm (VoML+G) that starts with the identification of galaxy overdensities, unconstrained to the presence of BCGs, by a purely geometrical method

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

(Voronoi) and then proceeds with the cluster identifications through a radial velocity isolation criterion. The algorithm, which does not use cluster parameterization and identifies members within $R_{200}$,[8] was tested and optimized through a mock Two-Degree Field Galaxy Redshift Survey (Colless et al. 2001; hereafter 2dFGRS) derived from the Millennium cosmological simulations (Springel et al. 2005). Another important result from Paper I was that the velocity dispersions computed for clusters with 10 or more galaxies implied masses compatible with corresponding mock clusters.

This paper reports the detection of 1901 galaxy clusters (the full catalog) resulting from the application of the VoML+G algorithm to the real 2dFGRS but focuses on the analyses of clusters with at least 10 galaxies that are thus open to mass estimates. Because of its depth ($z \sim 0.2$) and width, blue magnitude limit, known selection functions, and database containing a proxy for galaxy type, the 2dFGRS is well fitted to the application of the VoML+G algorithm and allows detailed comparisons between the resulting clusters and previous identifications done by several teams using a similar or identical database. The expected completeness and purity rates for our cluster catalog are ∼75% and ∼90%, respectively (Paper I).

The three main achievements of this work are (i) the identification of a volume-limited subsample of clusters with 10 or more member galaxies, and thus with mass estimates, uniformly detected over redshift; (ii) the realization of a census of cluster properties, including the late-type fraction, over this volume; and (iii) the finding that the cluster population of the subsample can be described by two components, one with galaxy membership dominated by early-type galaxies (about 2/3 of the total), and one with membership dominated by late-type galaxies (about 1/3 of the total). Existing major studies on the galaxy content of clusters have used samples consisting of several dozen cD or regular clusters (De Propris et al. 2004; Fasano et al. 2012), which are known to be dominated by early-type galaxies. Therefore, their measurements are related to the cluster population component dominated by early-type galaxies that we find. As for the second component, there have been previous indications of a cluster population dominated by late-type galaxies, notably by Oemler (1974); however, it is in this paper that we establish their abundant presence in the relatively local universe.

An exhaustive comparison between our cluster detections and previous systematic searches in the related region, which includes the recovery of Abell-like clusters, was conducted to reassess the population of clusters to $z \sim 0.10$. For the application of the VoML+G cluster finder we chose an identical 2dFGRS database to the one used by Eke et al. (2004) for their automatic search for 2PIGG groups. This allowed both a general comparison of the 2PIGG groups with our clusters and the comparison of individual clusters in a few regions of the sky. We show that almost all of the VoML+G clusters have counterparts in the 2PIGG catalog, albeit with quite different characterization.

De Propris et al. (2002, hereafter DP02) were the first to search for the Abell-type clusters within the 2dFGRS galaxy database. They used a collation of clusters from the following catalogs: Abell and ACO clusters (Abell et al. 1989), the Automated Plate Measuring (Dalton et al. 1997; APM), and the Edinburgh-Durham Cluster Catalogue (Lumsden et al. 1992; EDCC). Based on catalog entries, they identified cluster members within 1 Abell radius ($R_A$)[9] using a "double-gapping filter" technique, with a velocity gap of 1000 km s$^{-1}$. Many of these Abell-like clusters, discovered through visual inspection, have been found to be nearly virialized structures. VoML+G rediscovers a large fraction of the DP02 clusters, whose characterization agrees closely when compared with *expanded versions* of the VoML+G clusters, from $R_{200}$ to 1 $R_A$.

Later, Robotham et al. (2010) reviewed the existing 2dF-based catalogs of galaxy groups. They considered the 2PIGG catalog of Eke et al. (2004), which is the largest sample of groups within the 2dFGRS automatically generated through a friends-of-friends (FoF) algorithm, together with the group catalogs by Yang et al. (2005), Merchán & Zandivarez (2002), and Tago et al. (2006). Although Robotham et al. (2010) found differences between catalogs, as there is not a single *correct* determination of the boundary of a galaxy group, they concluded that both the 2PIGG and Yang et al. (2005) catalogs describe the same galaxy base population for a given dark matter halo, but with the 2PIGG including a more tenuously associated population.

The structure of the paper is as follows. We specify in Section 2 the adopted data set. Section 3 describes the basics of the VoML+G cluster finder. In Section 4 we discuss the full catalog of VoML+G clusters and show their consistency with the 2PIGG groups. Section 5 describes the cluster catalog (subset of $N_g \geqslant 10$ and $z < 0.14$ clusters) and considers the recovery of REFLEX and DP02 Abell-like clusters together with the correspondences with NED. In Section 6 the properties of the catalog clusters are analyzed. Section 7 presents the determination method of the fraction of late-type galaxies and defines the statistical subsample of clusters. Section 8 analyzes the properties of the clusters in the statistical subsample and shows that they can be described by two population components. In Section 9 we discuss the results and possible interpretations. Section 10 contains our conclusions.

A cosmology with $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ is adopted throughout this paper.

## 2. The 2dFGRS Data

We use a subset of the final data release of the 2dFGRS. The main survey covers about 1800 deg$^2$ and is made out of two continuous decl. strips, one in the southern Galactic hemisphere spanning 90° × 15° crossing the south Galactic pole (the SGP section), and the other in the northern Galactic hemisphere spanning 75° × 10° along the celestial equator (the north Galactic pole [NGP] section), having a median depth of $z = 0.11$ (Colless et al. 2001, 2003). Hereafter, unless explicit, we refer indistinctly to clusters or groups to represent physical associations of any number of galaxies.

We adopted the data set used by Eke et al. (2004) in their 2PIGG search of galaxy groups, containing 191,440 galaxies. Their galaxy data set comes from the final 2dFGRS release after removal of all the 2dFs having a redshift completeness less than 70% and then of all sectors (areas by the overlap of 2dFGRS fields) with completeness inferior to 50%. Figure 1 shows wedge diagrams corresponding to the selected galaxy sample and where the particularities of the large-scale structure (LSS) for the north and south strips of the survey can be appreciated.

[8] $R_{200}$ = radius of a sphere whose mean density is 200 times the mean density of the universe.

[9] $R_A = 1.7$ arcmin$/z \sim 2.14\, h_{70}^{-1}$ Mpc.

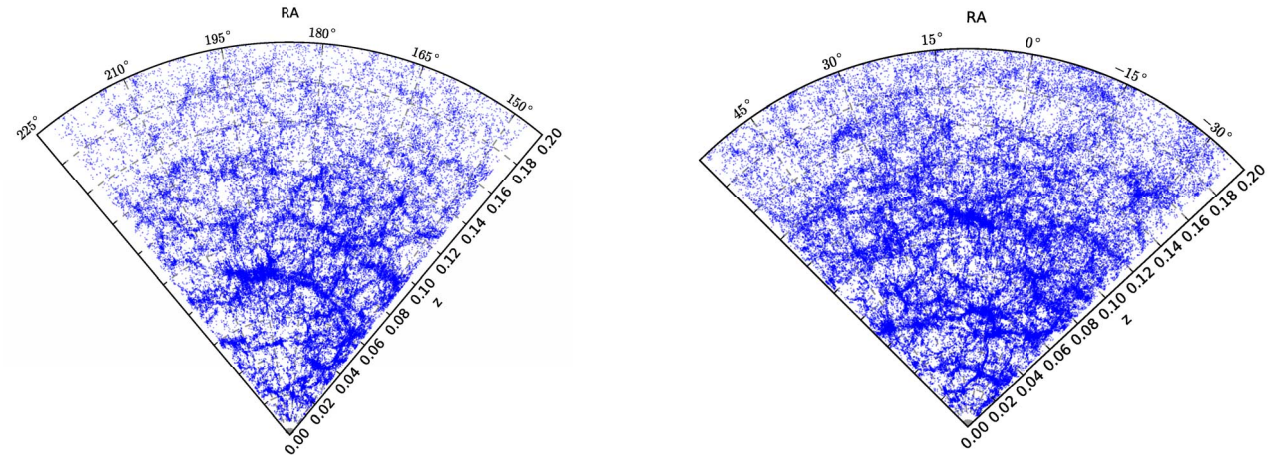


**Figure 1.** Wedge diagrams representing the projected spatial distribution for the north section (left panel) and the south section (right panel) of the 2dFGRS galaxies used in this work. The magnitude-limited character of the survey causes the decrease in galaxy density above $z \sim 0.12$.

Colless et al. (2001, 2003) provide a quality parameter $Q$ in the integer range of 1–5 for all the 2dFGRS survey redshift identifications; $Q \geqslant 3$ redshifts are 98.4% reliable and have an rms uncertainty of 85 km s$^{-1}$.

Galaxies included in the 2dFGRS have their nominal extinction-corrected magnitude $b_J$[10] within the 13.5–19.45 interval. The faint magnitude limit actually fluctuates from field to field, the effective median limit over the area of the survey being estimated at $b_j \sim 19.3$ (Colless et al. 2001). Our input database contains only galaxies with $Q \geqslant 3$, and we adopt a lower limit of $z = 0.009$ for the selected data set since at smaller redshifts the 2dF contains mostly galaxies belonging to the Local Supercluster. The magnitudes we use from the database are the final $b_J$ magnitudes with extinction correction.

The redshift completeness ($C_z$) is the ratio between the number of galaxies with redshifts of quality $Q \geqslant 3$, within an area of interest, and the respective number of galaxies in the parent catalog. The overall redshift completeness for galaxies with $Q \geqslant 3$ is 91.8%, but this varies with $b_J$ magnitude from 99% for the brightest galaxies to 90% for objects at the survey magnitude limit. The cluster's $C_z$ are determined using information from the 2dF database within areas 1° in radius centered on the cluster's centroids. Valid statistical comparison of $C_z$ values requires the definition of a common sampling area, and we have followed the choice by the 2dF team (Colless et al. 2003).

The default 2dFGRS spectral classification parameter $\eta$ was introduced by Madgwick et al. (2002), which identifies the average emission- and absorption-line strength in the galaxy rest-frame spectrum. Madgwick (2003) showed that the early-type (elliptical, S0) and late-type (spiral, irregular) galaxy populations can be separated through a particular $\eta$ threshold. The $\eta$ parameter is listed in the 2dFGRS database for nearly all (∼99%) of the galaxies.

## 3. The Algorithm

The VoML+G (Pereira et al. 2017) algorithm consists of two highly decoupled stages and does not parameterize the galaxy systems. In its first stage VoML+G finds the density peaks in the 2dFGRS galaxy database using a Voronoi Tessellation (VT), and then through a maximum likelihood method (MLE), adopted from a 2D version (Allard & Fraley 1997), it identifies structures consisting of at least two galaxies. The VT-MLE is locally adaptive, as the local density used for the generation of structures is computed within spheres $35\,h_{70}^{-1}$ Mpc in radius centered on the density peaks. The structures cannot share galaxies, thus biasing the selection toward more isolated systems, or exceed $2R_A$ in size. The centroids of the identified galaxy structures are the input for the second stage (Gapper$R_{200}$) of the algorithm, which is the one that identifies the gravitationally bound systems that are called clusters independently of their membership. The upper limit to the number of clusters is set by the first stage. At the start, Gapper$R_{200}$ selects the galaxies within $0.5\,h^{-1}$ Mpc and 4500 km s$^{-1}$ of the centroids of the structures and that are separated from each other by less than an initial velocity gap, which are filtered by an iterative process designed to compute $R_{200}$. The aforementioned parameters were determined in Paper I through an optimization process that involved running the VoML+G code multiple times over a mock 2dFGRS and comparing the resulting clusters against the "mock clusters" whose dark matter halo masses ranged across the $1.0 \times 10^{12}$–$1.0 \times 10^{15}\,M_\odot h^{-1}$ interval. Interestingly, their adopted optimal velocity gap (1000 km s$^{-1}$), crucial for the determination of gravitational bound systens, is coincident with the one used by DP02 to recover semiautomatically Abell-APM-EDCC clusters within the 2dFGRS.

The Gapper$R_{200}$ iterative process ends in either of two cluster detection cases. A case $a$, —or "$R_{200}$"—detection occurs when the process converges to a system with at least two member galaxies confined within an $R_{200}$. The case $a$ clusters are likely virialized systems. When the iterative process ends with a single galaxy, the candidate cluster is discarded unless at the start of the iteration the cylinder $0.5\,h^{-1}$ Mpc in radius contained at least five member galaxies. The latter is a case $b$ —or "$0.5\,h^{-1}$ Mpc"—detection. The case $b$ clusters are likely to be unvirialized systems. So, case $a$ solutions of the Gapper$R_{200}$ stage of the algorithm occur when consistent centroids, velocity dispersions, and virial radii are found, while the case $b$ solutions occur when the iteration fails to converge into a single centroid, suggesting that these clusters may present substructure. In Sections 5.1 and 8.5 we analyze the observational properties of the case $a$ and case $b$ clusters.

[10] Based on a IIIa-J photographic emulsion exposed through a GG395 filter.

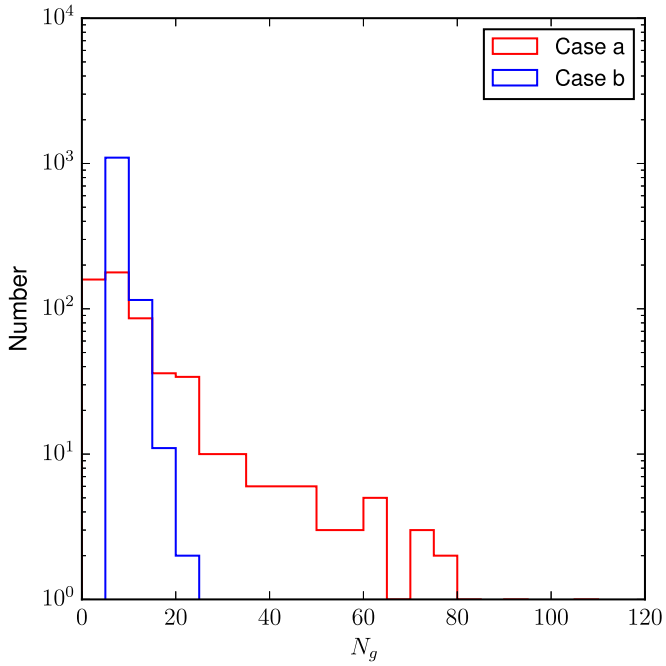


**Figure 2.** Histograms showing the distribution of multiplicities for clusters in the VoML+G full catalog ($N_g \geqslant 2$, $z < 0.14$) by detection case; case *a* in red and case *b* in blue. Nearly all the bound systems with 20 or more members are case *a*.

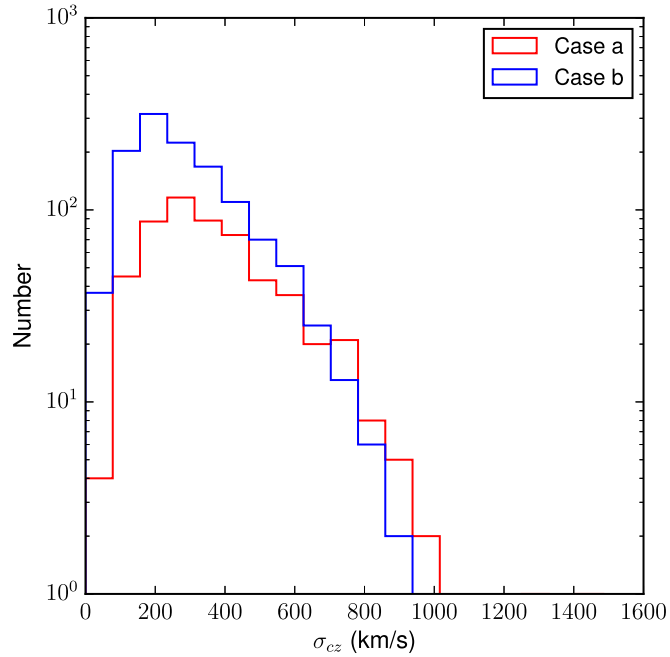


**Figure 3.** Histogram showing the distribution of cluster velocity dispersions for clusters in the VoML+G full catalog ($N_g \geqslant 2$, $z < 0.14$) by detection case, case *a* in red and case *b* in blue. The distributions are similar, although clusters with velocity dispersions beyond ~900 km s$^{-1}$ are all case *a* detections.

## 4. The Full VoML+G Cluster Catalog

The application of VoML+G to the selected 2dFGRS database detected 1901 clusters with $N_g \geqslant 2$ and $z \lesssim 0.22$. They include a total of 15,652 galaxies, that is, 8.2% of the galaxies in the data set. "Clusters" is the denomination used for all the bound galaxy systems identified by the algorithm. They may correspond, completely or partially, to entries in existing catalogs of groups and/or clusters of galaxies.

The number of case *a* clusters is 581 and of case *b* clusters is 1320.

The histogram of the number of galaxies (multiplicity) of the VoML+G clusters is shown in Figure 2. It can be seen that the majority of the clusters have between 2 and 25 galaxies. The histogram of velocity dispersions of the VoML+G output is shown in Figure 3. It can be seen that the bulk of $\sigma_{cz}$ spans from a few tens to 1000 km s$^{-1}$ and that the distribution peaks at ~250 km s$^{-1}$. There are two outlier clusters with $\sigma_{cz} > 1000$ km s$^{-1}$, possibly due to the presence of important substructure.

The projected spatial distribution to $z_{\rm lim} = 0.14$ of the VoML+G clusters is displayed in Figure 4 through wedge diagrams separately for the NGP and SGP sections of the 2dF survey. The information on the cluster's multiplicity and velocity dispersion is coded in Figure 4 through the sizes and color of the symbols representing the clusters, with the ranges involved being specified in the figure itself. Walls, superclusters of galaxies, filamentary structure, and voids are apparent in the figure. As a consequence of the brightness limit of the 2dF survey, few clusters are detected beyond $z \sim 0.12$, all of which have small multiplicities. The two rare superclusters noted by Baugh et al. (2004), one in NGP at $z \sim 0.08$ and especially the other in the SGP at $z \sim 0.11$, stand out because of their large cluster space density and the high multiplicity and velocity dispersion of the member clusters. The Pisces–Cetus is another SGP supercluster, at $z \sim 0.06$, easily recognized. Also in the SGP, we note an extended region devoid of clusters at $z \sim 0.05$, possibly related to a reported underdensity in the southern Galactic cap (De Propris et al. 2002; Whitbourn & Shanks 2014).

### 4.1. Correspondence with 2PIGG Groups

In this section it is shown that there are correspondences between the VoML+G clusters and the 2PIGG groups, albeit with characterizations that may differ considerably. Although the VoML+G algorithm was validated already (Paper I) through a mock 2dFGRS, such verification is a consistency check.

#### 4.1.1. Statistics and Spatial Distribution

VoML+G was applied to exactly the same 2dF galaxy database as the one used for the 2PIGG search for galaxy groups (Eke et al. 2004). Eke et al. (2004) generated the 2PIGG catalog that contains ~29,000 groups (to $z_{\rm lim} = 0.25$), which compose about 55% of the galaxy database; however, they estimate that ~40% of these groups may be interlopers. The VoML+G full sample is about 15 times smaller than the 2PIGG catalog and contains only 1/7 as many galaxies. Comparisons between the 2PIGG with other 2dF-based automatically generated catalogs, performed by Tago et al. (2006) and Robotham et al. (2010), have shown that 2PIGG contains a larger number of groups, an effect that Robotham et al. (2010) interpreted to be due to differences in the way group substructure was dealt and to the biasing in some algorithms toward more compact and gravitationally bound groups.

We searched for VoML+G clusters with correspondence with at least one 2PIGG cluster. The adopted matching criterion was an angular separation smaller than $2R_A$, and a

**Table 1**
Clusters for Comparison between 2PIGG and VoML+G Detections

| VoML+G Catalog No. | R.A. (deg) | Decl. (deg) | $\bar{z}$ | $N_g$ | $\sigma_{cz}$ (km s$^{-1}$) | Brightest Galaxy $M_{bj}$ (mag) | Close Cluster(s) (NED Search) |
|---|---|---|---|---|---|---|---|
| 82 | 37.68 | −33.05 | 0.077 | 47 | 657 | −21.8 | A3027 |
| 129 | 155.09 | −2.55 | 0.061 | 47 | 590 | −21.3 | LCLG-03 018 |
| 135 | 157.46 | −1.75 | 0.035 | 57 | 1276 | −21.3 | None |
| 33 | 10.58 | −28.55 | 0.108 | 60 | 735 | −21.98 | A2811 |
| 213 | 194.99 | −4.18 | 0.084 | 69 | 797 | −21.99 | A1651 |
| 123 | 153.46 | −0.91 | 0.045 | 74 | 706 | −22.24 | A0957 |

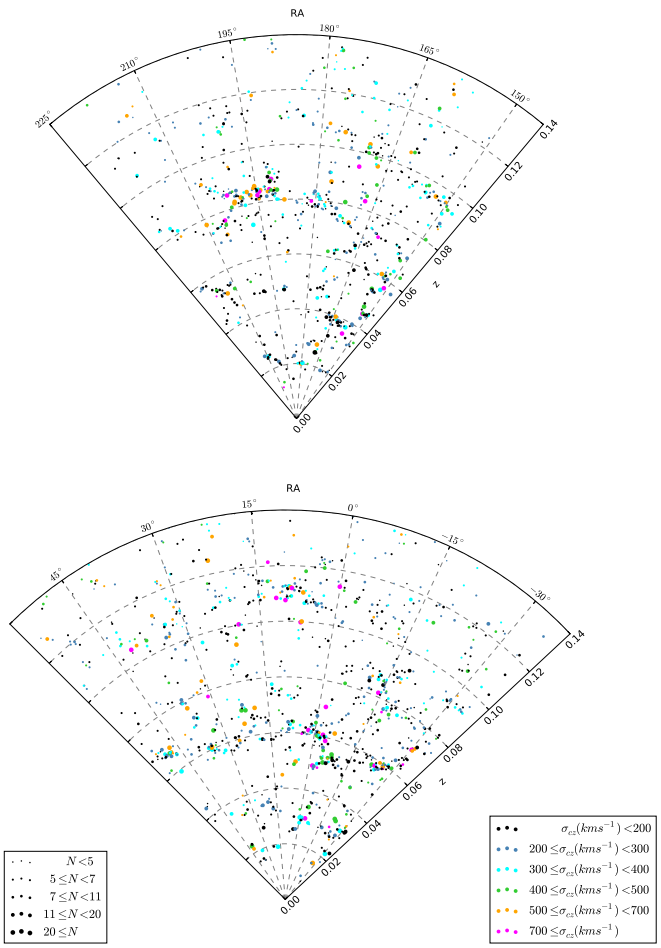


**Figure 4.** Projected spatial distribution of the clusters detected by the VoML+G algorithm with two or more galaxies, for the NGP (top) and SGP sections (bottom) of the 2dF survey. Cluster multiplicities and velocity dispersions are coded in the display and their ranges specified in the boxes next to the cones. Beyond $z \sim 0.12$, there are fewer clusters and with smaller multiplicities. Three rare superclusters are apparent, two in the SGP and one in the NGP.

radial velocity separation less than twice the largest cluster velocity dispersion of the pair under comparison. We find that ∼90% (1706 out of 1901 clusters) of the VoML+G clusters have a correspondence in the 2PIGG catalog, a fraction that increases to 100% when only clusters with 10 or more member galaxies are taken into account.

We then considered the statistics for clusters with at least five galaxies and with redshifts in a range where the 2dF galaxy density is large ($z$ = 0.05–0.10). It is found that there are 872 VoML+G clusters that include 8109 galaxies (9.3 galaxies per cluster), while 2PIGG has been reported by Robotham et al. (2010) to contain 1535 groups with $z$ = 0.05–0.10 that include 17,805 galaxies (11.6 galaxies per cluster). That is, for this redshift range, 2PIGG finds twice as many groups as the VoML+G algorithm, and they contain on average ∼25% more galaxies.

The spatial distribution of the VoML+G clusters is shown through wedge diagrams in Figure 4, where we code the cluster's properties with the same symbols and ranges used by Eke et al. (2004, their Figure 5) to display the space distribution of the 2PIGG groups. In spite of the differences between the VoML+G and 2PIGG algorithms, we find that the space distributions of the respective clusters/groups are qualitatively compatible, as they should be because they reflect the same galaxy LSS.

#### 4.1.2. Graphical Comparison of Detections

We have established that all the VoML+G clusters with $N_g \geqslant 10$ have at least one correspondence in the 2PIGG catalog (Section 4.1). In order to get some insight on the differences/similarities between the structures that these two algorithms find in a given region of space, we explore the groups found by 2PIGG in a region where a VoML+G cluster lies. We address this question by graphically considering sky regions centered on a few VoML+G clusters, with extent several times (4×) the cluster size. For each cluster, represented by its components and convex hull, the surrounding 2dF galaxies within the redshift interval of the member galaxies are displayed. Shown separately are the 2PIGG groups detected in the same sky areas. Six clusters with $N_g > 45$ (case *a* detections) were selected from the VoML+G catalog (see Section 5), four of which have correspondences with Abell clusters and one with a galaxy group. The clusters are listed in Table 1, ordered by increasing multiplicity, with the following information: catalog number, angular coordinate of centroid, mean redshift, number of members, velocity dispersion, absolute magnitude of the brightest cluster galaxy, and associated correspondence in the literature, if any.

The comparison between VoML+G clusters and 2PIGG detections in the same regions reveals how similar/different the solutions found by different algorithms can be. It has been remarked already that the 2PIGG algorithm uses an FoF method (Press & Schechter 1974) that is known to have a tendency to overestimate the cluster's multiplicities (see More et al. 2011).

The graphical comparisons are shown in Figures 5 and 6, where the right column displays the VoML+G clusters, whose member galaxies are represented by blue dots and the circles correspond to the $R_{200}$ radii, and the left column displays the detected 2PIGG groups in the same regions of the sky. The galaxies not belonging to the VoML+G cluster are represented by small black dots, and the bars in the lower right corner of the panels representing the VoML+G clusters correspond to the

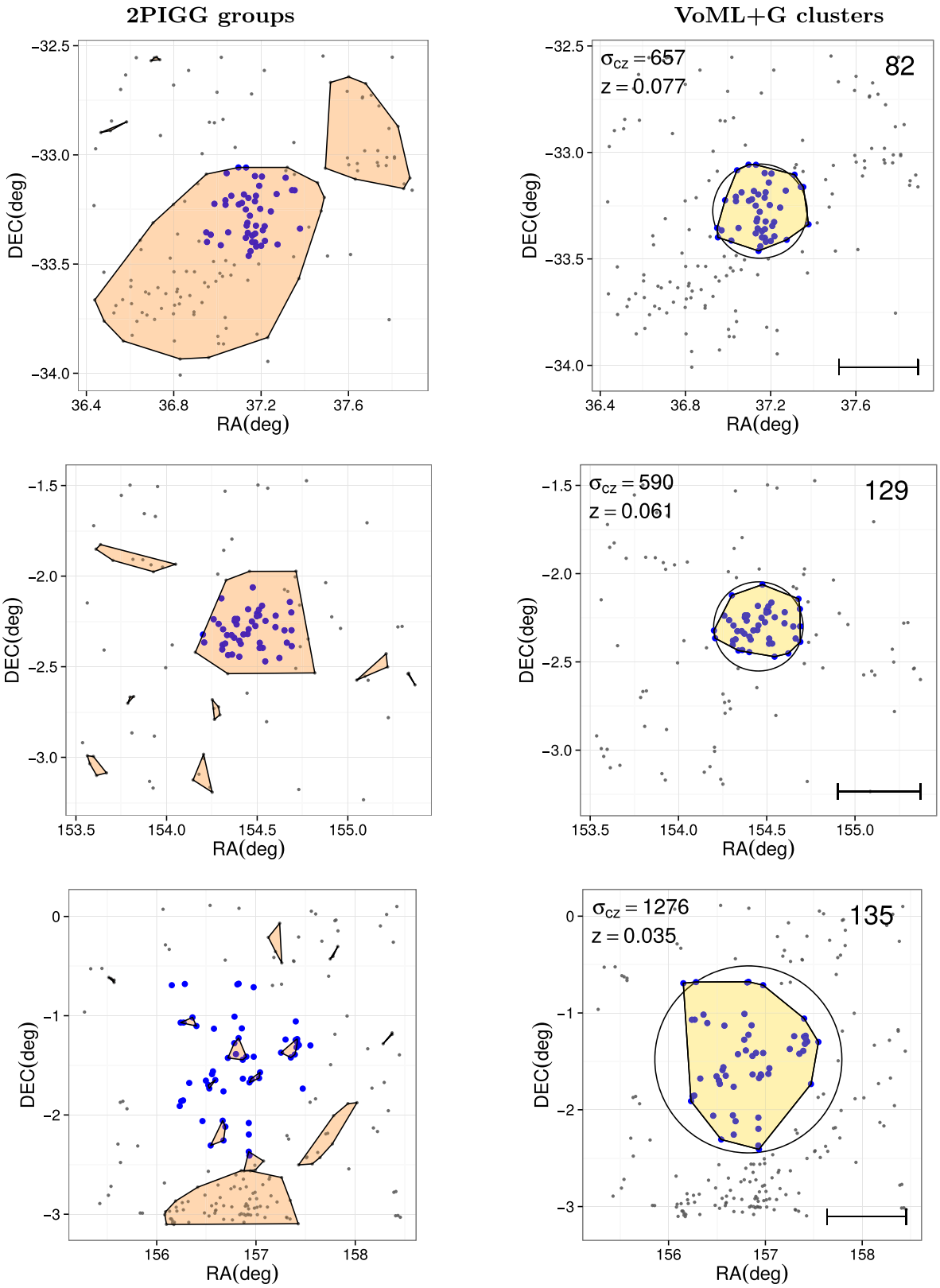


**Figure 5.** Right panels display the sky projection of the VoML+G cluster detections (all case *a*) listed in Table 1; large blue dots represent the identified member galaxies, and black dots represent the surrounding galaxies. The circles have $R_{200}$ radii, and the yellow-filled polygons are the convex hulls of member galaxies corresponding to each cluster. Left panels display the same sky fields and galaxies as the right panels, but the orange-filled polygons represent the galaxy groups detected therein by Eke et al. (2004) (2PIGG). The two algorithms detect overdensities in these example fields, although they are characterized differently. The bars in the lower right corner of the right panels correspond to the cluster's Abell radii.

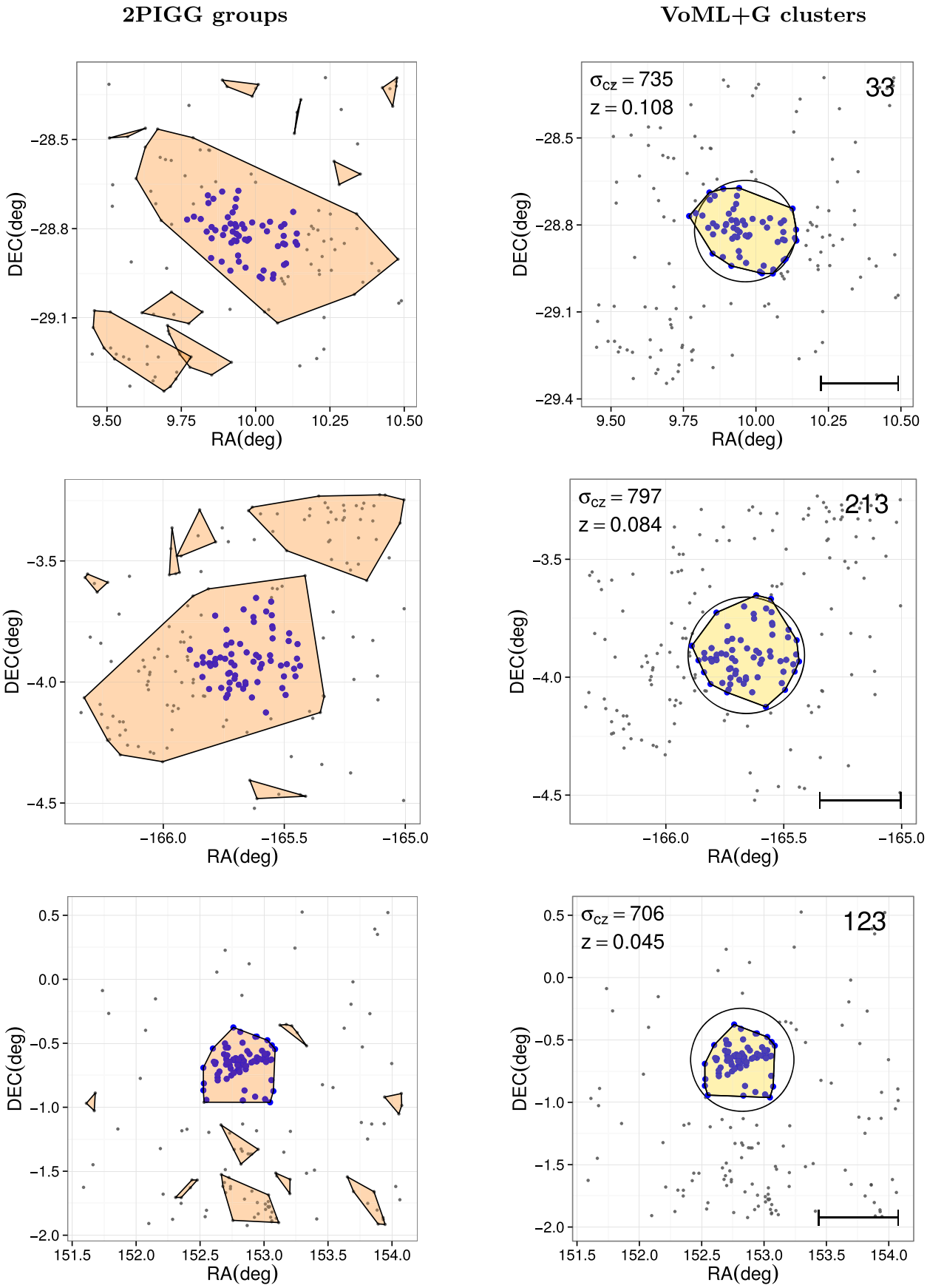


**Figure 6.** Right panels display, similarly to Figure 5, the VoML+G cluster Nos. 33, 213, and 123, which are counterparts to the A2811, A1651, and A0957 clusters, respectively. The VoML+G detections are densely packed within $R_{200}$ and reveal low ellipticities; their galaxy contents are later shown to be dominated by early-type galaxies. The left panels show that the 2PIGG counterparts for these Abell clusters are larger and more elongated than the identifications by the VoML+G algorithm.

**Table 2**
VoML+G Cluster Catalog ($N_g > 10$)

| ID | R.A. (deg) | Decl. (deg) | Redshift $\bar{z}$ | $N_g$ | $\sigma_{cz}$ (km s$^{-1}$) | $f_L^0$ | $f_L$ | $M_{bJ}^{\min}$ (mag) | $\overline{M}_{bJ}$ (mag) | $N_g'$ ($M_{bJ} \leqslant -18.5$) | $C_z$ | Correspondence | Case |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.43 | −27.48 | 0.028 | 17 | 306 | 0.71 | 0.75 | −20.9 | −18.4 | 8 | 0.92 | ABELL S1171 | *a* |
| 2 | 0.65 | −30.56 | 0.030 | 21 | 306 | 0.52 | 0.29 | −20.4 | −18.0 | 7 | 0.87 | ABELL S0001 | *b* |
| 3 | 0.71 | −27.13 | 0.067 | 13 | 245 | 0.23 | 0.20 | −21.3 | −19.2 | 10 | 0.92 | ABELL 2716 | *a* |
| 4 | 0.74 | −34.63 | 0.114 | 31 | 801 | 0.23 | 0.23 | −21.5 | −20.0 | 31 | 0.67 | ABELL 2715 | *a* |
| 5 | 0.79 | −27.89 | 0.065 | 19 | 426 | 0.58 | 0.57 | −21.2 | −19.3 | 14 | 0.91 | ABELL S0003 | *a* |
| 6 | 1.78 | −25.86 | 0.063 | 10 | 241 | 0.20 | 0.14 | −19.8 | −18.7 | 7 | 0.80 | None | *b* |
| 7 | 1.83 | −28.06 | 0.061 | 11 | 314 | 0.09 | 0.17 | −22.1 | −19.1 | 6 | 0.92 | ABELL 2726 | *a* |
| 8 | 2.12 | −25.62 | 0.064 | 10 | 169 | 0.40 | 0.67 | −21.2 | −19.4 | 6 | 0.69 | None | *b* |
| 9 | 2.58 | −32.31 | 0.026 | 14 | 248 | 0.79 | 0.00 | −20.8 | −17.8 | 1 | 0.92 | None | *b* |
| 10 | 2.86 | −28.87 | 0.061 | 77 | 739 | 0.14 | 0.12 | −21.4 | −18.9 | 50 | 0.93 | ABELL 2734 | *a* |

(This table is available in its entirety in machine-readable form.)

Abell radii at the cluster's redshifts. The shapes of the clusters are shown using the convex hull constructed from the respective member galaxies, colored pink for the 2PIGG groups and yellow for the VoML+G clusters. It is seen in the right panels of Figures 5 and 6 that these VoML+G clusters are approximately circular.

Figure 5 displays from top to bottom the sky regions corresponding to VoML+G cluster Nos. 82, 129, and 135. For cluster No. 82, corresponding to A3027, it is seen that the VoML+G detection picks out the virialized region out to $R_{200}$ while the 2PIGG version of the cluster seems to be including a connecting filament. The identifications by VoML+G and 2PIGG of cluster No. 129, corresponding to LCLG-03 018, are nearly identical, possibly because of its isolation and packed galaxies. Cluster No. 135, without correspondence in NED, is a special case because of its large radius, low galaxy surface density, and a large velocity dispersion (1276 km s$^{-1}$), possibly a consequence of substructure; 2PIGG does not find a large group in this region, but instead detects many small groups.

Analogously, Figure 6 displays from top to bottom VoML+G cluster Nos. 33, 213, and 123. The comparison of the VoML+G detections of the cluster Nos. 33 and 213, corresponding to A2811 and A1651, respectively, with the 2PIGG versions shows again, as for cluster No. 82, that VoML+G is picking up the virialized regions while 2PIGG includes connecting filaments. Cluster No. 213 corresponds to A1651 and also has a VoML+G detection with circular geometry, which is also detected by the 2PIGG group finder as a much larger oblate structure. Finally, for cluster No. 123, which corresponds to A0957, the VoML+G detection has circular geometry and possesses a very dense central region, while the 2PIGG finder assigns an identical membership.

From the comparison performed for the six sky regions containing VoML+G clusters with high multiplicity ($N_g > 45$), we may conclude that the 2PIGG algorithm generates nearly the same galaxy structure when the cluster has a dense central region and is surrounded by relatively few galaxies. Otherwise, these examples highlight a difficulty of 2PIGG, and of FoF methods in general, in the sense that when the overdensities are not quite isolated, the algorithm merges in connecting filaments with the cluster core. In contrast, when the VoML+G cluster corresponds to sparse galaxy overdensities, the 2PIGG algorithm splits these loose clusters into many small groups.

### 4.2. Detection of Abell-APM-EDCC Clusters

DP02 used a 2dF galaxy data release that included ∼90% of the final survey to identify and characterize (mean-$z$, $\sigma_{cz}$, and multiplicity) semiautomatically 450 Abell-APM-EDCC clusters that lie within the 2dF continuous strips of our study and that are contained in their Table 1 (hereafter the DP02 cluster sample). Using the same matching criteria as employed in the comparison with the 2PIGG groups, we determine how many of the clusters in the DP02 sample have counterparts in the VoML+G full sample. We find that 73% (330 out of 450) of the DP02 clusters have correspondence with a VoML+G cluster. When the DP02 sample is restricted to clusters with 10 or more galaxies, 82% (314 out of 385) have matches in the VoML+G full sample, while the percentage of clusters with 30 or more galaxies that have correspondences with VoML+G clusters is 97% (134 out of 138).

The above recovery rates are compatible with the completeness determined for the VoML+G catalogs (Paper I) using a mock 2dFGRS. A complete recovery is not expected because the DP02 study was based on cluster centroids from the Abell-APM-EDCC, in contrast to automatically determined ones by the VoML+G algorithm, and because the galaxy databases are not identical. A detailed comparison between the VoML+G identifications and the ones by DP02 is addressed in Section 5.

## 5. The Cluster Catalog

The catalog contains the galaxy systems identified through the VoML+G algorithm that fulfill the following conditions:

1. Identifications within either of the continuous NGP or SGP strips of the 2dF survey.
2. $N_g \geqslant 10$ within either $R_{200}$ (case *a*) or $R \leqslant 0.5\,h^{-1}$ Mpc (case *b*).
3. $\bar{z}$ within the 0.009–0.140 range.

The lower redshift limit is to avoid the inclusion of galaxies from the Local Supercluster, and the upper limit is set by the strong drop at $z \sim 0.15$ in galaxy density and cluster detection efficiency.

The catalog contains 341 clusters, 155 (45%) of them belonging to the NGP section and 186 (55%) to the SGP section. Further, 213 (63%) and 128 (37%) are case *a* and case *b* detections, respectively. All the clusters with $N_g > 25$ are case *a*.

Table 2 provides the basic properties for the Catalog clusters. Only the first 10 clusters are listed to show the structure of the

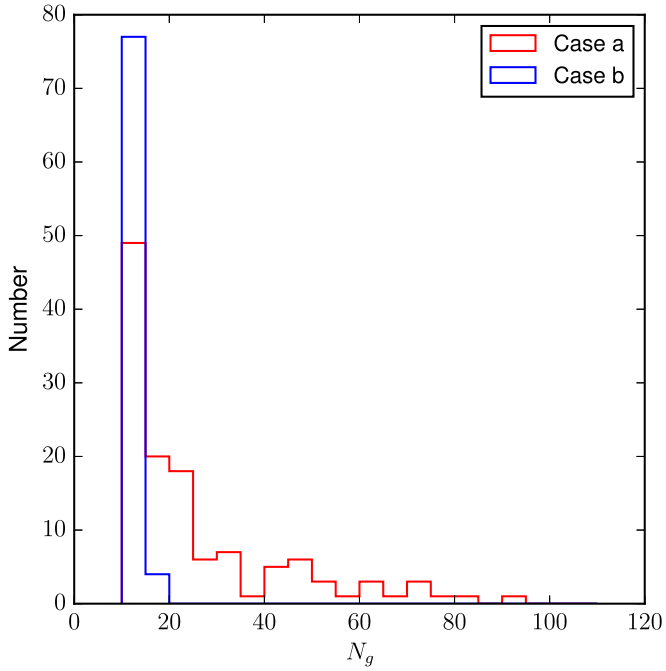


**Figure 7.** Multiplicity histograms for the VoML+G catalog clusters ($N_g \geqslant 10$) by detection case; case *a* in red and case *b* in blue. Nearly all the case *b* detections have between 10 and 15 member galaxies. The case *a* distribution also peaks at 10–15 members; however, it has a tail toward higher multiplicities that reaches up to $N_g \sim 90$.

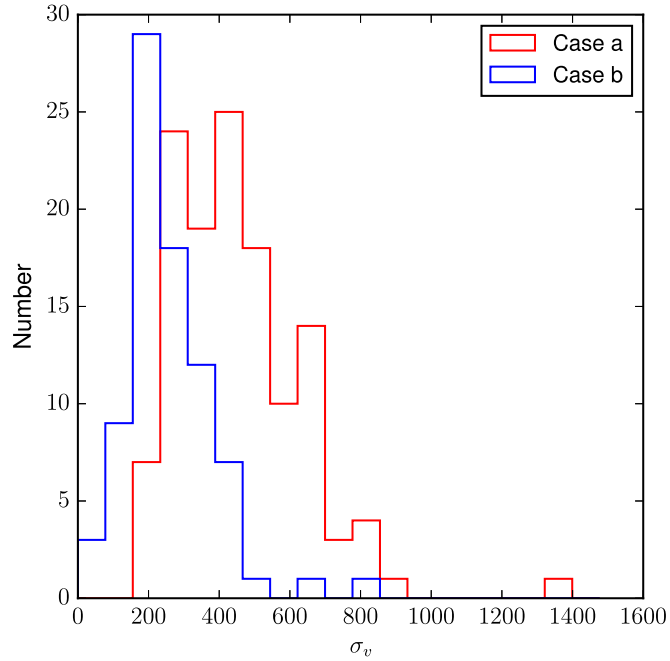


**Figure 8.** Velocity dispersion histograms for the VoML+G catalog clusters ($N_g \geqslant 10$) by detection case; case *a* in red and case *b* in blue. The distribution for case *b* detections peaks at $\sim$200 km s$^{-1}$. The case *a* detections become dominant for $\sigma_v \gtrsim 300$ km s$^{-1}$ (see the text).

table, as the complete list is provided online. The clusters are listed in order of increasing R.A., with the contents of each column being as follows: VoML+G identification number, coordinates $\alpha$ and $\delta$ (2000) of the centroid, mean redshift, cluster multiplicity, velocity dispersion, raw fraction of late-type galaxies, corrected fraction of late-type galaxies, absolute magnitude of the brightest galaxy, mean absolute magnitude of the member galaxies, multiplicity after absolute magnitude cut, redshift completeness $C_z$, NED correspondence with either clusters or groups of galaxies (with the closest correspondence being listed when they were several) as of 2018 July (see Section 5.2), and VoML+G cluster detection case (*a* or *b*). See Section 7 for the definition of the fractions of late-type galaxies in clusters.

### 5.1. Cluster Properties by Detection Case

The design of the VoML+G algorithm accepts the identification of either virialized or unvirialized systems: cases *a* or *b*. Their actual properties are examined below.

The multiplicity histograms for the catalog clusters, separately for case *a* (in red) and *b* (in blue) detections, are shown in Figure 7. Most of the case *b* detections have between 10 and 15 member galaxies, whereas about 2/3 of the case *a* detections have between 10 and 25 members, with the rest reaching up to $N_g \sim 90$. This indicates that case *b* detections tend to be low-multiplicity clusters consistent with their generally smaller sizes (upper limit of 1.0 $h^{-1}$ Mpc).

Figure 8 shows the distributions of cluster velocity dispersion for the cluster cases. The distribution for case *b* clusters extends over the 100–400 km s$^{-1}$ interval with a peak at $\sim$200 km s$^{-1}$. The distribution for case *a* clusters extends over the 200–800 km s$^{-1}$ interval with a peak at $\sim$400 km s$^{-1}$. For bins above $\sigma_v \gtrsim 300$ km s$^{-1}$ the majority of the clusters are case *a*, and for $\sigma_v \gtrsim 500$ km s$^{-1}$ nearly all of them are case *a*.

The histograms of absolute magnitudes of the brightest cluster galaxies, separately for case *a* (in red) and case *b* (in blue), are shown in Figure 9. The brightest cluster galaxies with the higher luminosities (Min($M_{b_J}$) $< -21.5$) belong almost exclusively to case *a* detections. However, for lower luminosities than this threshold, representing most of the clusters, both cluster detection cases are equally represented.

### 5.2. Cluster Cross-correlation with NED

Using a tolerance of $0.5R_A$ for the angular distance between centroids and allowing a maximum mean redshift difference of 0.0015, we find that 254 of the 341 catalog clusters ($\sim$75%) have counterparts with clusters or groups in the literature (NED). The remaining systems (87) not in NED can be accounted for partially by incompleteness of catalogs and plausibly by systems with properties that make them more difficult to detect.

The Catalog clusters without NED correspondences are denominated the "new" clusters to differentiate them from the others ("known"). Figure 10 shows the number of "known" and "new" clusters separated by detection case. It is seen that the "new" clusters correspond to either case *a* or case *b*, although they proportionally contribute more to the total number of case *b* detections.

### 5.3. Cluster Cross-correlation with REFLEX

We consider the correspondence with REFLEX (Böhringer et al. 2004) clusters.

There are 28 REFLEX clusters within the survey and the redshift limit of the VoML+G cluster catalog. Eighty-two percent of them (23 out of 28) have correspondences in the

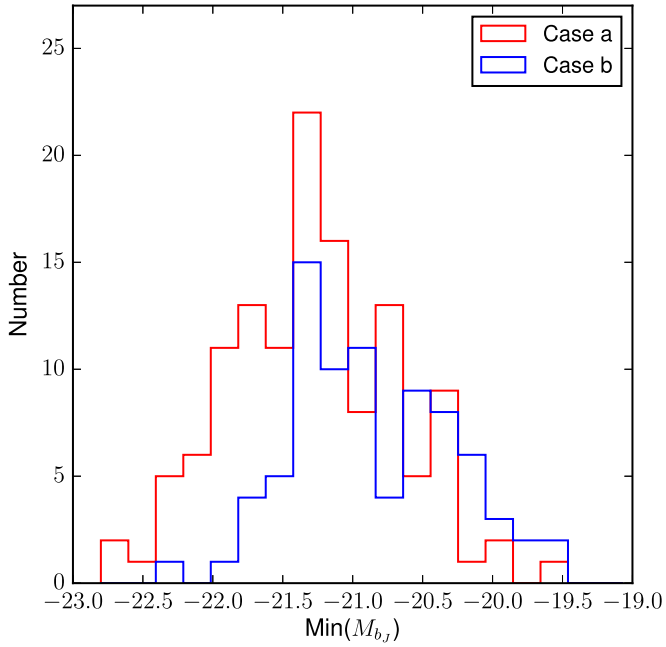


**Figure 9.** Distribution of the absolute magnitude of the brightest cluster galaxies (Min($M_{b_J}$)) for the VoML+G catalog clusters ($N_g \geqslant 10$) by detection case; case *a* in red and case *b* in blue. Few case *b* detections have Min($M_{b_J}$) under −21.5; however, both detection cases have similar distributions in the lower luminosity range.

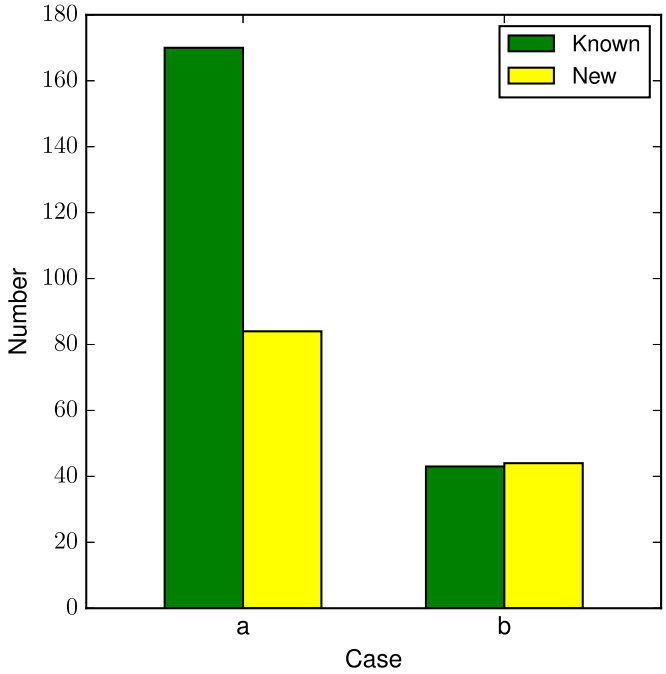


**Figure 10.** Number of "known" (in NED) and "new" (not in NED) clusters contained in the VoML+G catalog ($N_g \geqslant 10$) separated by detection case.

Cluster Catalog, a rate consistent with the completeness estimated in Paper I. We investigated whether special circumstances determined the nondetection of the remaining five clusters and found them to be either in regions of low 2dF completeness or close to a bright galaxy/star. These conditions presumably determined that some galaxies could not be included in the 2dFGRS database, thus affecting their identification by the algorithm.

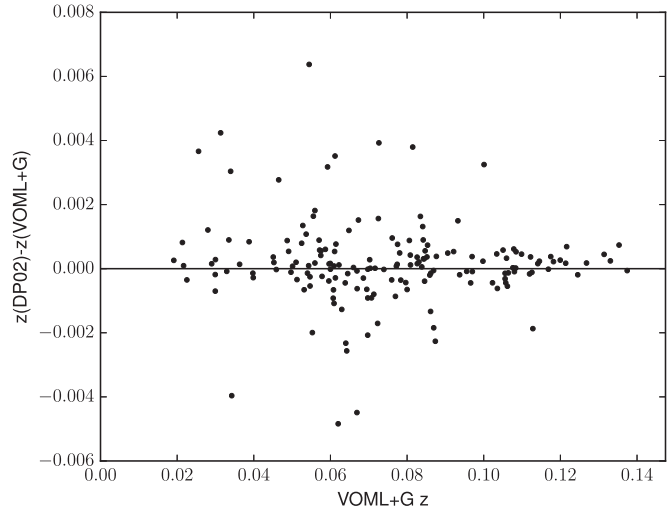


**Figure 11.** Difference in mean redshift between clusters in De Propris et al. (2002) and corresponding matches in the VoML+G catalog cluster ($N_g \geqslant 10$) vs. the VoML+G cluster mean redshift.

**Table 3**
REFLEX Clusters Not Detected by the Algorithm

| Reflex Name | Counterpart | Reason for Nondetection |
|---|---|---|
| RXC J0006.0-3443 | A2721 | Close bright star, low completeness (0.5) |
| RXC J0017.5-3509 | A2755 | Low completeness (0.4) |
| RXC J1309.2-0136 | MS1306.7 | Central bright galaxy |
| RXC J2149.1-3041 | A3814 | Bright star close to the center |
| RXC J2213.0-2753 | ⋯ | Galaxy group, small number of galaxies |

Table 3 specifies the most likely reason for each of the nondetections by the VoML+G algorithm as well as their corresponding optical cluster counterparts from the literature.

### *5.4. Rediscovery and Characterization of Abell-APM-EDCC Clusters*

The properties of the Catalog clusters with correspondence in the DP02 study are intercompared. The aim is to determine how close can the properties of the automatically identified VoML+G clusters be with respect to the 2dF-based study of Abell-APM-EDCC clusters. This is important because our algorithm was designed to detect galaxy systems in general, up to $2R_A$ in size, and thus the recovery of well-established galaxy associations like the Abell clusters is expected.

Using the same criteria for establishing correspondences with the 2PIGG groups, we find that 166 clusters (49%) from the Cluster Catalog have matches with Abell-APM-EDCC clusters included in the DP02 sample. For these galaxy systems we compare the mean-$z$, multiplicity, and $\sigma_{cz}$ computed from the VoML+G identifications with corresponding values determined by DP02. Note that the VoML+G multiplicities used for this comparison are for modified versions of our clusters, radii *expanded* to $1R_A$, such that our counts are computed over radii equal to those used by DP02. The mean-$z$ matching has an rms of 0.001 (3% relative rms), as seen in Figure 11, which is smaller than the maximum expected redshift difference due to the matching criterion (0.002). The VoML+G multiplicities have a very good agreement with the DP02 determinations, shown in Figure 12, with a highly significant Spearman correlation coefficient of

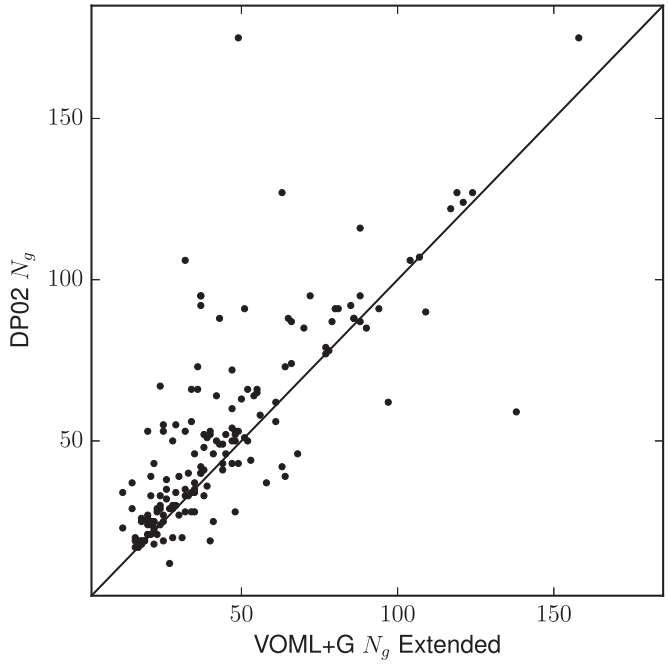


**Figure 12.** Scatter plot of multiplicities for clusters in De Propris et al. (2002) and corresponding matches in the VoML+G catalog cluster ($N_g \geqslant 10$). Note that the VoML+G multiplicities correspond to extended versions that include galaxies within $1R_A$ from the centroids (see the text).

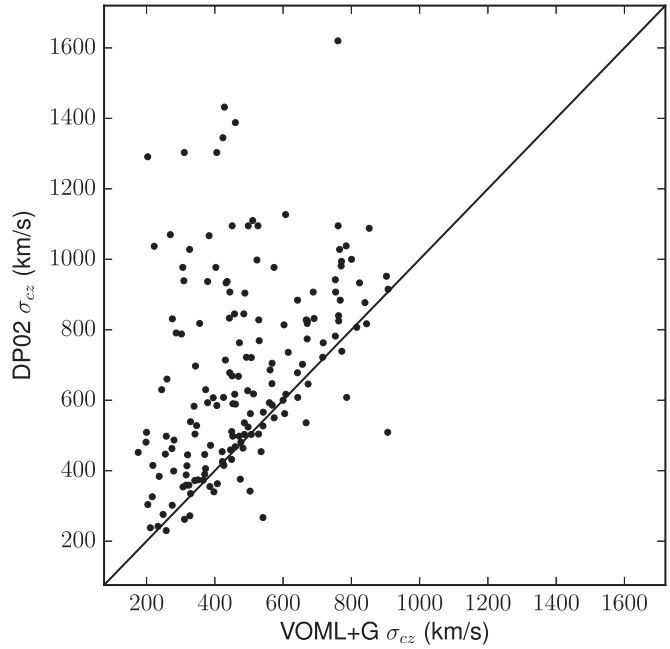


**Figure 13.** Scatter plot of velocity dispersion for clusters in De Propris et al. (2002) and corresponding matches in the VoML+G catalog cluster ($N_g \geqslant 10$). About 2/3 of the DP02 velocity determinations are considerably larger (in the mean by 93%) than the values of the VoML+G correspondences.

0.8 ($p$-value $< 0.0001$) and an rms scatter of 21 around the identity relation, which is unsurprising because of the difference between cluster centroids. It should be noted that the multiplicities determined by either ourselves or DP02 from the 2dFGRS for the Abell-APM-EDCC clusters are generally much smaller than the counts (richnesses) determined by Abell.

A comparison between the VoML+G and DP02 velocity dispersions, displayed in Figure 13, shows that the DP02 values are generally larger (in the mean by 93%) and with dispersion of 328 km s$^{-1}$. Our estimation of the velocity dispersion for a majority of the clusters is done within $R_{200}$, close to the virial radius, whereas the DP02 cluster velocity dispersions were estimated within $1R_{\rm A}$. It is thus likely that DP02 included galaxies outside the virial radius for some of their cluster identifications, which consequently could result in overestimated velocity dispersions. This expectation is actually seen to occur in Figure 13; the DP02 velocity dispersion estimates for more than one-half of the clusters exceed the values determined through the VoML+G identification.

This confirms that our velocity dispersions are good, and not overestimates. DP02 used the same 1000 km s$^{-1}$ velocity gap as we do, but because they applied it over the Abell radius, they had a larger probability of including interlopers at large radii. Our focus on the cluster core ($R_{200}$) is the reason why we are not affected by such interlopers when estimating the velocity dispersion.

## 6. Properties of the Catalog Clusters

We now proceed to study the properties of the galaxy systems contained in the Cluster Catalog. The properties of the "known" and "new" systems are displayed separately to determine whether the latter have differentiating features.

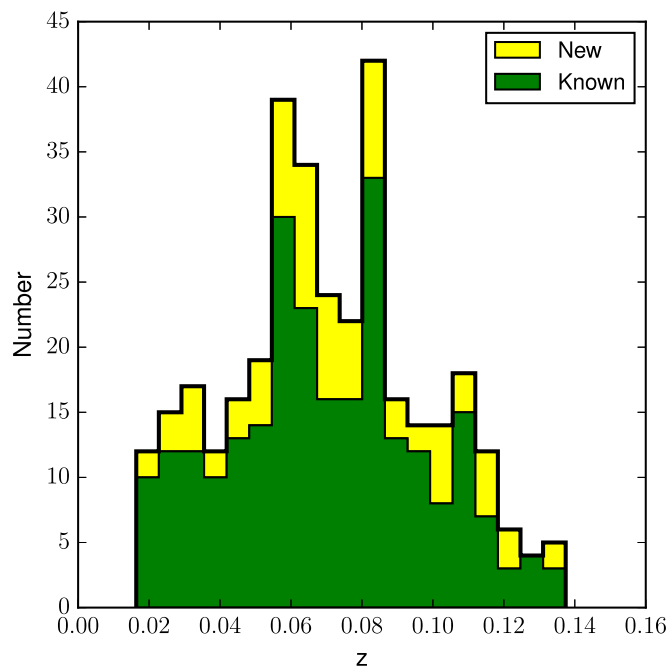


**Figure 14.** Histogram showing the distribution (black thick line) in redshift for clusters in the VoML+G catalog ($N_g \geqslant 10$). The counts are separated into "known" (in NED) and "new" (not in NED) clusters, corresponding to the filled areas in green and yellow, respectively. It is seen that the new clusters span the whole redshift range.

### 6.1. Redshift Distribution

Figure 14 shows the histogram of the redshifts of the clusters, with 0.01 bins, decomposed into the contributions from "known" (green) and "new" (yellow) detections. The "new" clusters contribute over the whole $z$-range, so they are not associated with a particular redshift range. The histogram has several peaks at particular redshifts, such as $z \sim 0.06$,

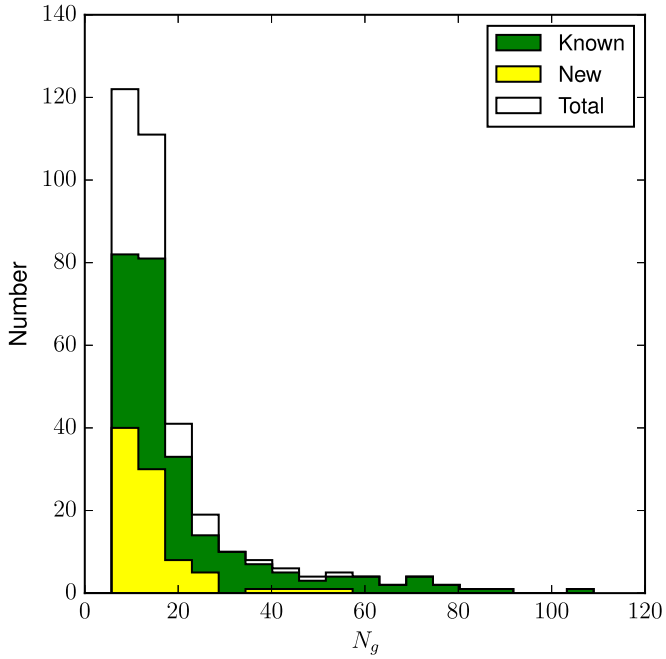


**Figure 15.** Histogram (thin black line) showing the multiplicity distribution of the clusters in the VoML+G catalog ($N_g \geqslant 10$). Overlapping histograms for the "known" and "new" clusters are shown in green and yellow, respectively. Most of the "new" clusters have between 10 and 15 member galaxies.

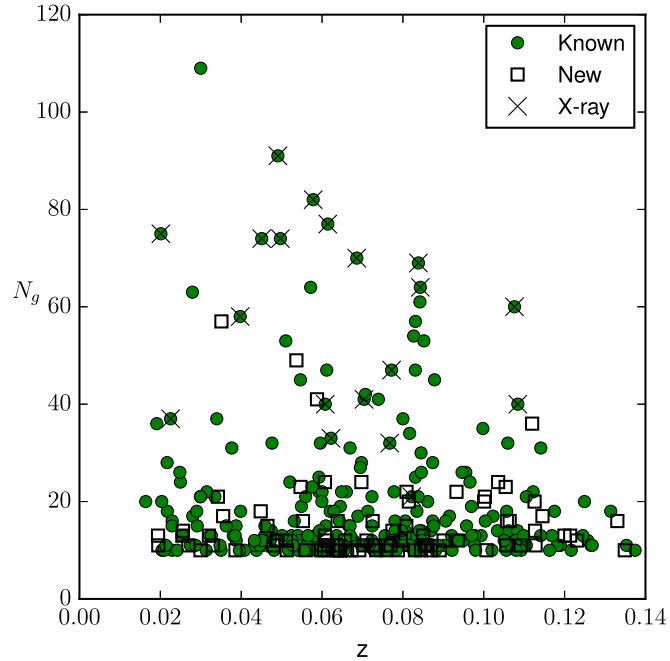


**Figure 16.** Scatter plot in the $N_g$–$z$ plane shown for the VoML+G catalog ($N_g \geqslant 10$) clusters. The multiplicities span a wide range (1–100), with both the "known" and "new" clusters represented to $N_g \sim 60$; only some "known" clusters are seen beyond and reaching $N_g \sim 100$. Nearly 4/5 of the "new" clusters have between 10 and 20 galaxies.

$z \sim 0.08$, and $z \sim 0.11$, attributable to the combined effect of NGP and SGP LSS features. A marked drop in the number of clusters is seen at $z \sim 0.11$.

### 6.2. Spatial Distribution of Clusters

Figure 4, already presented in Section 4.1.1, displays cone diagrams with the projected spatial distribution of VoML+G clusters as a function of R.A. and redshift, for the clusters in the north and south sections of the 2dF. Although these plots include the full VoML+G cluster sample, we are focused here only on the distribution of the Catalog Clusters, i.e., with $N_g \geqslant 10$. Figure 4 shows that the clusters themselves are generally clustered, a large fraction of them located in denser regions corresponding to superclusters. It can be noted that the number and importance of superclusters are somewhat different in the NGP and SGP sections, suggesting that cluster studies over the whole 2dFGRS would reduce the effect of cosmic variance.

### 6.3. Multiplicity Distribution

Figure 15 shows the multiplicity distribution for the Catalog clusters and also the distributions for separate subsets corresponding to the "known" and "new" detections. The distribution for "new" detections is unlike the one for "known" detections because of the dominance by low-multiplicity clusters, with mean and median multiplicities of 19 and 13, respectively. Figure 16 shows a scatter plot of the "known" and "new" detections in the $N_g - z$ plane, where the crosses mark the clusters with a REFLEX counterpart. It is seen, ignoring the four detections with $N_g \gtrsim 40$, that the range of multiplicity over redshift of the "new" detections is approximately constant and that some systems with $N_g \gtrsim 20$ are present up to $z \sim 0.11$. A redshift bias is observed in the maximum detectable multiplicity, which is unsurprising for a magnitude-limited galaxy survey. It can also be noted that nearly one-half of the clusters with more than 50 galaxies have REFLEX counterparts.

### 6.4. Velocity Dispersion Distribution

Figure 17 shows the distribution of the rest-frame velocity dispersion (corrected to the cluster rest frame) for the Catalog clusters, as well as the distributions for subsets corresponding to the "known" and "new" detections. It is seen that these subsets have similar velocity dispersion distributions, and thus the "new" clusters are not biased in velocity dispersion with respect to the Catalog clusters, their mean velocity dispersion being 375 km s$^{-1}$. Figure 18 shows a scatter plot of the catalog clusters in the $\sigma_{cz}$–$z$ plane displayed with the same symbols as in a previous graph. It is seen that the range over which the velocity dispersions are determined is nearly unbiased with respect to redshift, up to $z \sim 0.11$, the bulk range being 50–1000 km s$^{-1}$. This uniform sensitivity is important because it shows that the algorithm allows probing approximately the same cluster mass range up to $z \sim 0.11$. Figure 18 also shows that all the clusters with REFLEX correspondence occur for $\sigma_{cz} \gtrsim 400$ km s$^{-1}$ and that some LSS features are reflected in the distribution of points over the 0.06–0.08 redshift interval. The two outlier clusters with $\sigma_{cz} \sim 1300$ km s$^{-1}$ seen in both Figures 17 and 18 are "new" and possibly correspond to systems with important substructure.

## 7. Cluster's Galaxy Contents via Late-type Fractions

In the previous section we found that although the Catalog clusters without NED correspondences ("new") are preferentially low-multiplicity systems, their velocity dispersions span over the same range as the "known" clusters. This suggests that the "new" clusters are on average as massive as the "known"

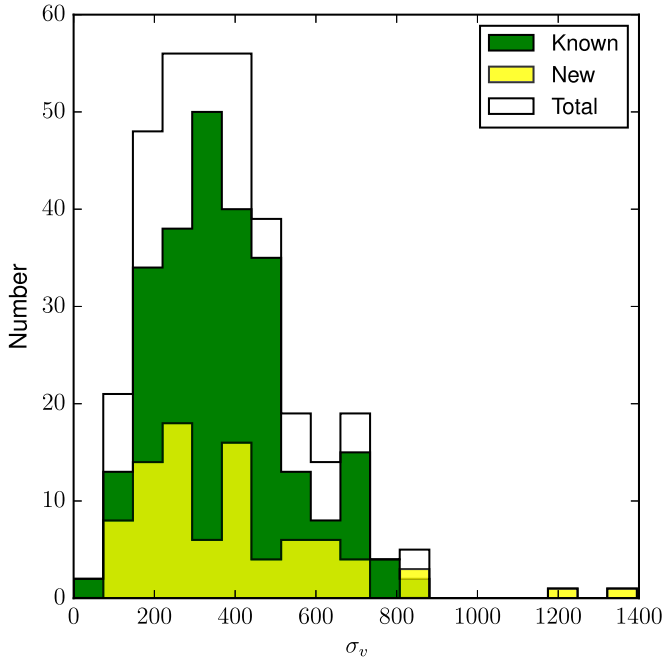


**Figure 17.** Histograms of velocity dispersion in the rest frame (thin black line) for the VoML+G catalog ($N_g \geqslant 10$) clusters. Overlapping histograms for the "known" and "new" clusters are shown in green and yellow, respectively; these subsets are seen to basically span the same redshift range. There are two outlier clusters with $\sigma_{cz} > 1000$ km s$^{-1}$, both "new" ones, which possibly correspond to systems with important substructure.

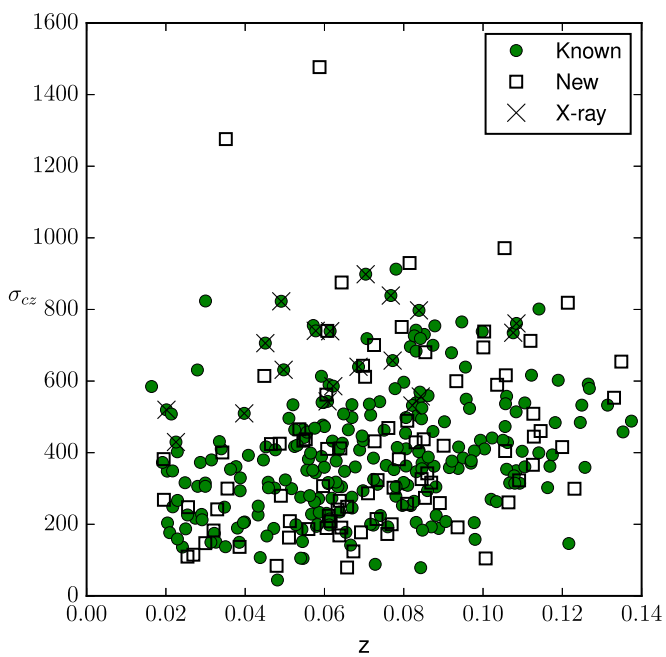


**Figure 18.** Display of the VoML+G catalog ($N_g \geqslant 10$) clusters in the velocity dispersion–redshift plane. The "known" clusters are represented by green circles, with superposed crosses indicating a REFLEX correspondence, and the "new" clusters are represented by open squares. Disregarding the outliers clusters with $\sigma_{cz} > 1000$ km s$^{-1}$, it is seen that the $\sigma_{cz}$ range of the VoML+G clusters is remarkably uniform over the (0.04–0.10) redshift range.

ones but contain fewer galaxies, i.e., that they are high mass-to-light ratio ($M/L$) systems.

The cluster galaxy content is another property of interest for the study of the Catalog clusters. A procedure is defined to compute this property that needs to be applied on a particular cluster subset in order to avoid biases connected with the magnitude-limited character of the 2dFGRS, with respect to the properties under investigation. Madgwick (2003) investigated the correlation of 2dFGRS galaxy morphologies with the spectroscopic parameter $\eta$ (see Section 2). He showed that, adopting an $\eta$ threshold, the galaxies could be separated with a high degree of certainty into two bins, early (elliptical, S0) type and late (spirals, irregulars) type, and in particular that a ~75% success rate in recognizing early-type galaxies can be achieved with the $\eta < -1.4$ condition and a ~70% rate for late-type galaxies with $\eta \geqslant -1.4$.

For the characterization of the galaxy population of the VoML+G clusters we use an $\eta = -1.4$ cut. Every cluster galaxy is classified into either early or late type, and then for each catalog cluster we compute the fraction of late-type galaxies. The *raw* late-type fraction, $f_L^o$, is the direct quotient between the number of late-type galaxies and the total number of galaxies in the cluster. Given that the probability of correctly assigning a galaxy to one type or the other is approximately the same, we expect that the late-type fraction computed for a cluster will be more robust than the assignation of galaxy type to individual 2dF galaxies because the errors will partially compensate owing to the binarity of the classification.

In this section we provide a summary of previous estimates of late-type, or spiral, fractions in clusters and compare methods in terms of classification procedure (e.g., spectral classification vs. visual morphology), the adopted magnitude limit, and the radii within which the fractions were computed. Second, we examine the images of galaxy members belonging to a couple of the "new" clusters to illustrate how successful the application of the $\eta$ parameter is in separating the galaxies into two bins. Third, we compute the distribution of raw late-type fractions for the Catalog clusters and search for correlations between late-type fraction and cluster properties.

Finally, we investigate whether the magnitude-limited character of the 2dF survey biases the calculation of the *raw* late-type fractions. This is done by calculating *proper* late-type fractions, $f_L$ counting only member galaxies with luminosities above a fixed limit. Although we find that the differences between *raw* and *proper* late-type fractions are not large, for our analysis we use only *proper* late-type fractions and adopt an approximately complete and volume-limited cluster subset (*the statistical subset*).

### 7.1. Previous Work on Galaxy Content

It is interesting to review the major studies that have determined the galaxy content in large cluster samples at low $z$. We have identified two such studies, listed in Table 4, that are relevant for the discussion of our results on the fraction of late-type galaxies for the clusters in the *statistical sample*. De Propris et al. (2004) studied how the fraction of blue galaxies ($f_b$) varies as a function of cluster properties and in the field at the same redshift in the local universe using a sample of 60 clusters belonging to the DP02 study (see Section 3) and that have at least 40 members. The authors stress that these clusters to some extent share the biases of the Abell, APM, and EDCC catalogs from which they were originally drawn; they provide an approximately complete and volume-limited ensemble of nearby clusters, spanning a large range of properties (such as richness and velocity dispersion) determined from the same 2dFGRS data set. The VoML+G algorithm has been shown to be effective in identifying galaxy systems consistent with

**Table 4**
Previous Studies of Galaxy Content in Low-$z$ Clusters

| Study | Cluster Number | Redshift | Classification | Magnitude Limit | Aperture | Population Fraction |
|---|---|---|---|---|---|---|
| De Propris et al. (2004) | 60 | 0.02–0.11 | Color–Magnitude | 19.45 ($b_J$) | $R_{200}$ | 0.13 (Blue fraction) |
| Fasano et al. (2012) | 77 | 0.04–0.07 | Morphology | 18 ($V$) | $R_{200}/2$ | 0.23 (Spiral fraction) |

Abell-like clusters; however, the Catalog clusters contains many other bound galaxy systems that may eventually have contrasting properties. De Propris et al. (2004) performed the separation between blue/red galaxies through the color–magnitude diagram and computed the fractions using the parameters listed in Table 4. Among their results, with blue fractions evaluated within $R_{200}$ and $R_{200}/2$, the most relevant ones to our study are that there is no dependence on redshift of the blue fraction over the $z < 0.11$ range, that the scatter found in the blue fraction is suggested to be a real property of the sample, and that although $\overline{f_b} = 0.13$, there are a few clusters that have $f_b > 0.4$ but none attaining $f_b > 0.6$.

Fasano et al. (2012) studied the morphology of galaxies in the WIde-field Nearby Galaxy-clusters Survey (WINGS; Fasano et al. 2006), consisting of wide-field, photometric data, in the optical bands ($B$ and $V$) of several hundred thousand galaxies in the fields of 77 nearby *ROSAT* brightest clusters ($0.04 \leqslant z \leqslant 0.07$), selected from three X-ray flux-limited samples. These clusters also share the biases of the Abell catalog, like the DP02 sample, but in addition are constrained to the upper range of cluster X-ray luminosity. Fasano et al. (2012) generated automated estimates of morphological types for a total number of 39,923 galaxies through the application of the tool MORPHOT to the $V$-band WINGS imaging and determined morphological fractions as a function of the clustercentric distance. Through their study defined by the parameters listed in Table 4, and adopting conventional "broad" morphological galaxy classes, they found that at an $R_{200}/2$ clustercentric distance ellipticals, S0s, and spiral galaxies constitute ~33%, 44%, and 23%, respectively, of the whole galaxy population in the WINGS clusters. Although the fraction of spirals used by Fasano et al. (2012) is evidently not identical to the fraction of blue galaxies used by De Propris et al. (2004), their respective mean values are close. As mentioned before, the Cluster Catalog has been shown to be effective to find X-ray-bright galaxy systems consistent with the kind of clusters used by Fasano et al. (2012) in their study.

### 7.2. Raw Late-type Fraction Distribution

Before proceeding with the calculation of the cluster's late-type fractions, we present the optical images for galaxies belonging to two clusters from the Catalog (Table 2) to illustrate the application of the $\eta = -1.4$ cut established by Madgwick (2003) for the separation of their member galaxies into early and late types. Postage stamp images of 1 arcmin on a side, extracted from SDSS DR9 downloaded images, are shown for each of the member galaxies, where the individual $\eta$ value and classification as either early (E) or late (L) type are marked therein. Fifty-one galaxies belonging to the VoML+G cluster No. 135, at $\bar{z} = 0.035$ and with $f_L^o \sim 0.70$, are displayed in Figure 19, and 47 galaxies belonging to VoML+G cluster No. 129, at $\bar{z} = 0.061$ and with $f_L^o \sim 0.60$, are displayed in Figure 20. A perusal examination of the morphologies of the galaxies of VoML+G No. 135 indicates a 100% success rate for the spectroscopic classification of late types and a corresponding 80% for the early types. Similar examination of the images for the VoML+G No. 129 galaxies indicates a 100% success rate for the spectroscopic classification of late types and a corresponding 100% for the early types. Although Madgwick (2003) proved that $\eta$ is effective in assigning a 2dF galaxy to one of these two bins, its successful application for the calculation of the late-type fractions in a couple of VoML+G clusters reassures us that the implementation of the procedure is robust.

Figure 21 displays the catalog clusters in the $f_L^o$–redshift plane and codes information on both whether they are "known"/"new" clusters and whether they are contained in the REFLEX catalog. This diagram shows that the "known" and "new" clusters span over the same $f_L^o$ range (0–0.9); it is for $f_L^o \lesssim 0.8$ that the representative points intermix well over the total redshift interval, and in addition this reveals that all the REFLEX clusters have $f_L^o < 0.5$.

It should be stressed that some regions of the $f_L^o$–redshift plane may be affected by the magnitude-limited character of the 2dFGRS. The latter does not prevent cluster detections except for redshifts larger than ~0.12, where the galaxy density starts to drop drastically. However, when regions of increasing redshift are considered, the relative number density of late-type galaxies compared to early-type galaxies is reduced, and conversely, it increases toward smaller redshifts, thus affecting the *raw* late-type fraction values. In fact, the two regions devoid of clusters in Figure 21, indicated by gray areas, can be explained by selection effects. First, the fact that no clusters with more than 80% of early-type galaxies are detected within $z \sim 0.04$ is a consequence of the fact that we are including more low-luminosity galaxies, which are naturally more likely to be late-type galaxies. Second, the virtual absence of $z \gtrsim 0.09$ clusters with more than 60% of late-type galaxies is related to higher luminosity limits imposed at higher redshifts by the survey brightness limit.

The redshift dependence of the relative number density of early-type (late-type) galaxies can be seen in Figure 22, which shows a scatter plot of the absolute magnitude with redshift corresponding to the cluster member galaxies. The early-type galaxies are represented by red dots and the late-type galaxies by blue dots. The black solid curves correspond to the relations between absolute magnitude and redshift for both the lower and upper 2dF magnitude limits. For all the clusters, the cluster's galaxy data points indeed reach the curve minimum luminosity at the corresponding redshift. This figure reveals that galaxies with the lowest luminosities are only included in clusters at the lowest redshifts, thus favoring larger fractions of late-type galaxies in comparison to the fractions for clusters at higher redshifts. Also, it can be seen that for redshifts larger than ~0.09 there is a nearly complete dominance of early-type galaxies. These effects are due to selection biases that must largely be avoided before addressing the analysis of the distribution of late-type fractions in clusters. This will be achieved by selecting a subset of the Catalog clusters.

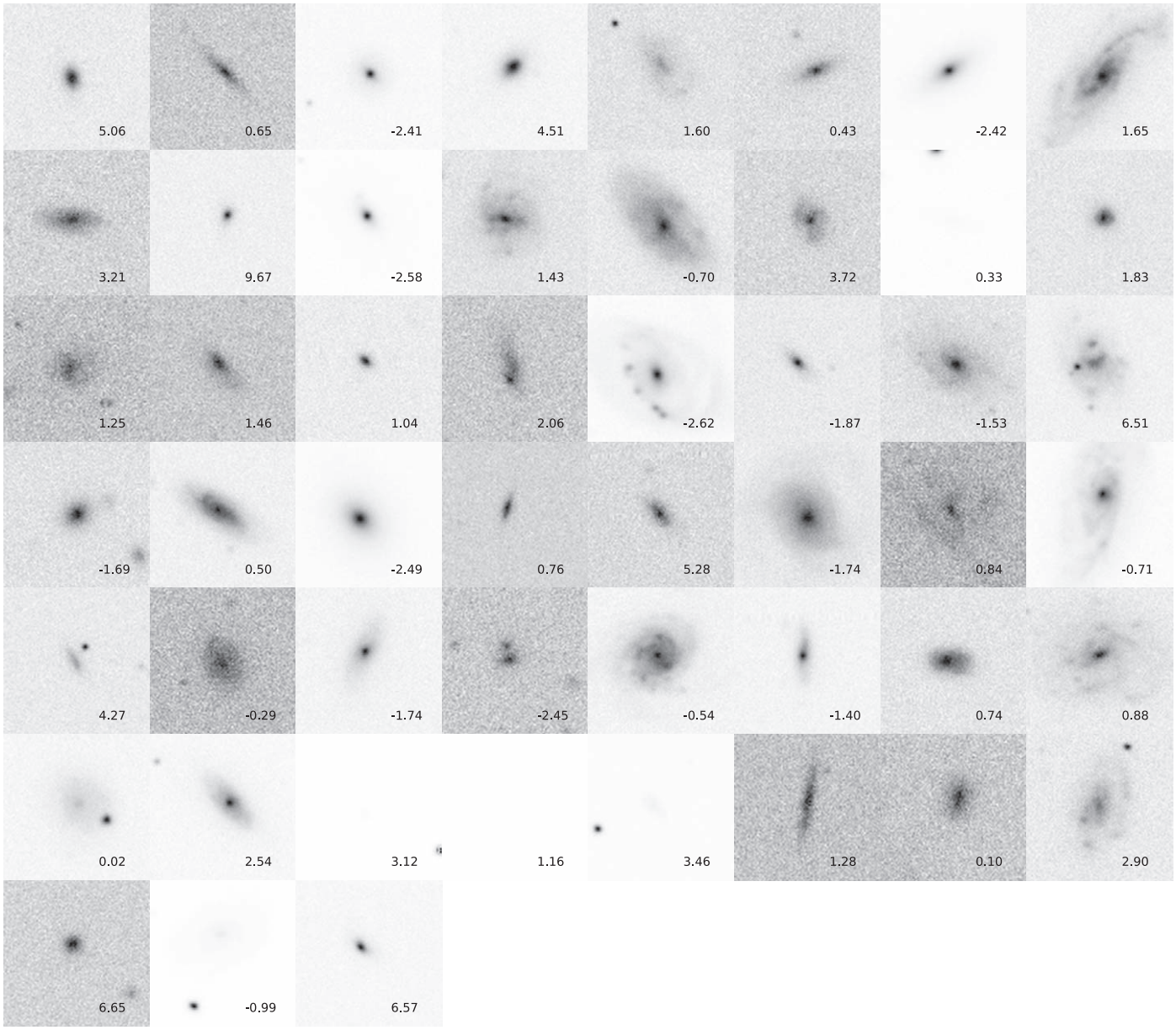


**Figure 19.** 1 arcmin postage stamp images from SDSS DR9 $g$-band images showing member galaxies belonging to the cluster VoML+G No. 135 (at $z = 0.035$). The number at the bottom of each stamp is the galaxy $\eta$ value, and the predicted galaxy type (early or late) is in its upper right corner. A visual inspection allows the verification that the galaxy morphologies in this case are fully consistent with galaxy types predicted from the corresponding 2dFGRS spectroscopic indexes and separation cut at $\eta = -1.4$ (see the text).

### *7.3. The Proper Late-type Fraction Calculation*

In principle, by considering only cluster members with luminosities above a fixed limit, we may get a better estimation of a cluster late-type fraction. We adopt a minimum galaxy luminosity of $M_{b_j} = -18.5$: galaxies at $z > 0.09$ that are less luminous than this limit are not included in the 2dFGRS (see Figure 22). The condition $z_{\rm cluster} \lesssim 0.09$ would then in principle suffice; however, we adopted also a low redshift limit of $z = 0.04$ because below this $z$-value a large proportion of the cluster member galaxies do not satisfy the $M_{b_j} = -18.5$ condition (see Figure 22). These redshift limits define a complete volume-limited sample, i.e., the *statistical subset*. Hereafter we use only the *proper* late-type fractions, $f_L$, computed from galaxy members with $M_{b_j} \leqslant -18.5$ for the clusters belonging to the *statistical subset*. Nevertheless, we compute both the *raw* and *proper* late-type fractions for all the Catalog clusters and list them in Table 2, together with $N_g'$, defined to be the remaining number of galaxies in a cluster after the absolute magnitude cut has been applied.

We analyze the effect of the magnitude cut in the estimation of the late-type fractions by computing the differences between *raw* and *proper* fractions. The resulting distribution of differences is shown in Figure 23. The distribution is strongly peaked at zero, with median equal to zero. The mean late-type fraction difference is 0.03, and the standard deviation is 0.12, the latter being compatible with the expected error of the fraction as a statistical estimator. Notwithstanding, we adopt the *statistical subset* to study the late-type fraction distribution because it reduces the luminosity bias.

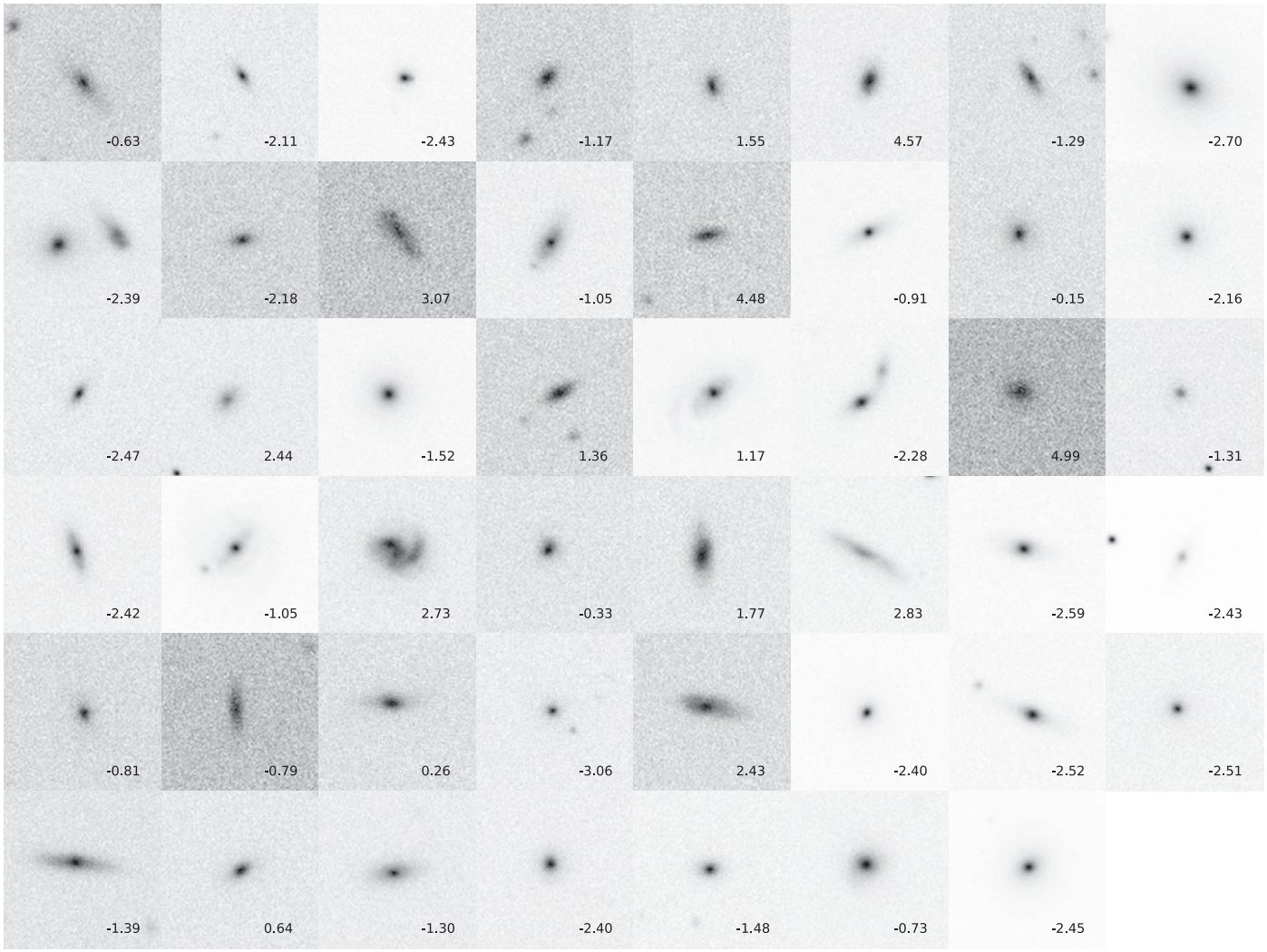


**Figure 20.** 1 arcmin postage stamp images from SDSS DR9 $g$-band images showing member galaxies belonging to the cluster VoML+G No. 129 (at $z = 0.061$). The number at the bottom of each stamp is the galaxy $\eta$ value, and the predicted galaxy type is in its upper right corner. A visual inspection allows the verification that the galaxy morphologies are in this case largely consistent with galaxy types predicted from the corresponding 2dFGRS spectroscopic indexes and separation cut at $\eta = -1.4$ (see the text).

Out of the 207 clusters in the *statistical subset*, 74% (154) are represented in NED as either galaxy clusters or groups. This percentage coincides with the corresponding one for the Cluster Catalog.

## 8. The Galaxy Content of Clusters in the Statistical Sample

### *8.1. Late-type Fraction Distribution*

Figure 24 shows the distribution of late-type fraction for the *statistical subset*. The curves displayed are the histogram (normalized) and the kernel density estimate of the data. The distribution strongly decays toward the extreme values of the late-type fraction, an expected behavior as they represent limits of the galaxy content in clusters. The distribution of late-type fraction hints at a bimodality, with peaks close to 0.3 and 0.6. The kernel density estimate shows this as a flattened peak induced by the bandwidth selected (Silverman's rule was used). In order to test the bimodality, we used a Gaussian mixture model (GMM; see Feigelson & Babu 2012; Ivezić et al. 2014), a model widely used in the search for underlying populations.

We perform fits to the data using a single-Gaussian component model and a double-Gaussian component model. We exclude in the fit clusters with $f_L$ equal to zero or one, because the finite range of $f_L$ produces an excess of these objects that cannot be accurately modeled using Gaussians, and also they are more likely to be due to incomplete sampling. We use AIC (Akaike Information Criterion; Akaike 1973; Feigelson & Babu 2012) to choose the best model, where the model with a lower AIC value is closer to the true distribution. The AIC for the one-component model is $-92$, whereas the one for the two-component model is $-95$, and therefore the two-component model is the most likely. The parameters of the two-component fit are given in Table 5, and Figure 24 displays the fitted Gaussian components. We also apply a likelihood ratio test for the null hypothesis of a single-Gaussian that gives a $p$-value of 0.002, and therefore the null hypothesis is rejected with a confidence level of $3\sigma$. We conclude that the data corresponding to the *statistical subset* support the notion of two cluster populations in the late-type fraction distribution. One with a peak fraction at $\sim$0.25, with few clusters with $f_L \geqslant 0.4$ but none attaining $f_L \geqslant 0.6$, is akin to the results found by De Propris et al. (2004) for the fraction of blue galaxies in Abell-like clusters. A second one implies a population with a

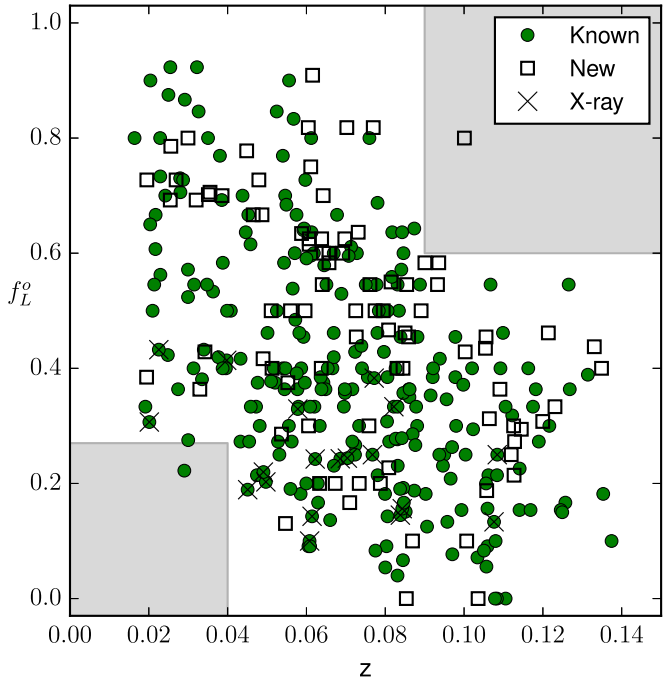


**Figure 21.** Scatter plot representing the VoML+G catalog clusters ($N_g \geqslant 10$) in the raw late-type fraction ($f_L^o$)–redshift plane. The symbol codes are the same as in the previous figures. The known clusters span the $\sim 0 - 0.7 f_L^o$ range and intermix well with the new ones, while the new clusters span the whole $f_L^o$ range (0–1). There are two regions devoid of clusters, indicated in gray, which are understood to arise from selection effects (see the text).

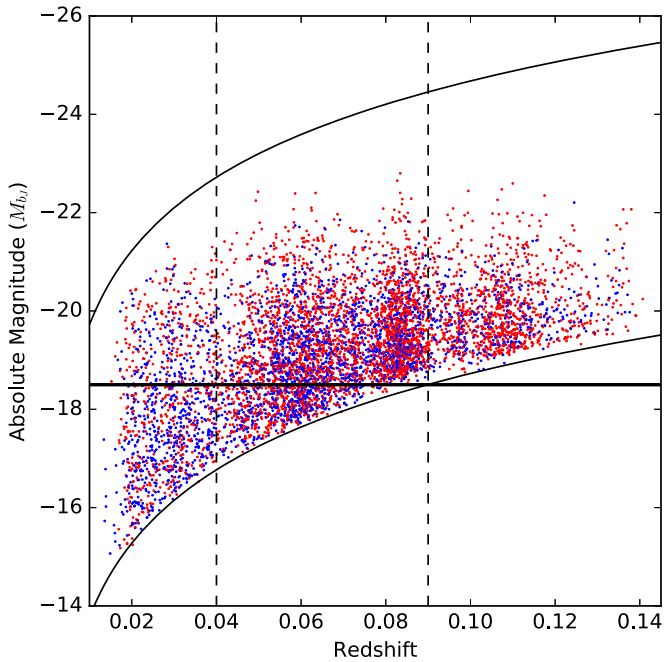


**Figure 22.** Scatter plot in the absolute magnitude–redshift plane representing all the member galaxies that belong to clusters in the VoML+G catalog ($N_g \geqslant 10$). Early-type galaxies are represented by red dots, and the late-type galaxies are represented by blue dots. The region between the curves is where galaxies may lie because of the flux limits of the 2dFGRS, and the thick horizontal line marks the luminosity limit $M_{bJ} = -18.5$. Only galaxies within 0.04–0.09 (dashed vertical lines) and luminosities above the limit are used in the analysis of the cluster's late-type fractions.

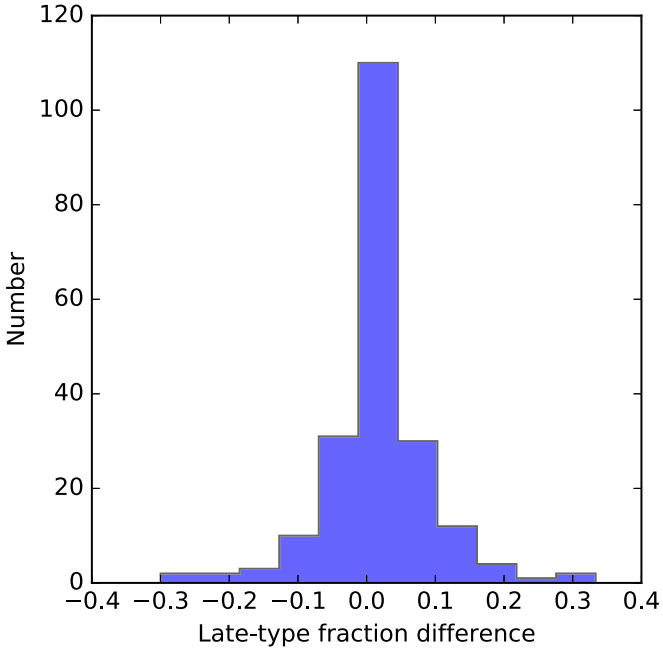


**Figure 23.** Histogram of the differences between raw and proper (see the text) late-type fractions for clusters in the VoML+G catalog ($N_g \geqslant 10$). A cluster proper fraction is computed considering only galaxy members with $M_{bJ} \leqslant -18.5$.

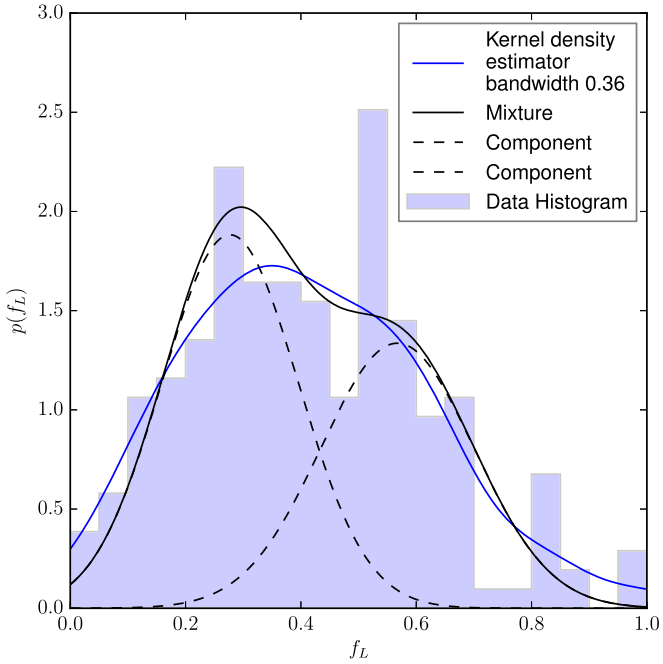


**Figure 24.** Probability density of the late-type fraction for the VoML+G statistical subset of clusters ($N_g \geqslant 10$ within $z = 0.04$–009). We estimate probability density using a normalized histogram (blue filled steps) and a kernel density estimator (blue line). The observed distribution hints at the presence of two modes. The fitted Gaussian mixture model (see the text) is shown using a black solid line (full mixture). Each Gaussian component is shown using dashed lines.

peak fraction at $\sim$0.6, a result that is not reminiscent of previous reports.

There is a possibility that clusters with lower multiplicity and therefore with corresponding larger errors in the estimated late-type fraction could affect the statistics of our previous analysis.

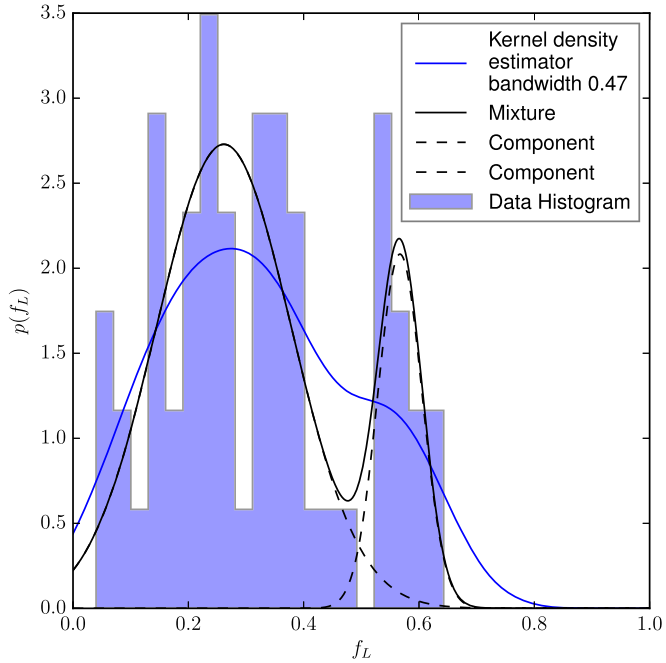


**Figure 25.** Probability density of the late-type fraction for the statistical subset of clusters, but now including only clusters with 20 or more members. We estimate probability density using a normalized histogram (blue filled steps) and a kernel density estimator (blue line). The observed distribution hints at the presence of two modes more clearly than the full statistical subset. The fitted Gaussian mixture model (see the text) is shown using a black solid line (full mixture). Each Gaussian component is shown using dashed lines.

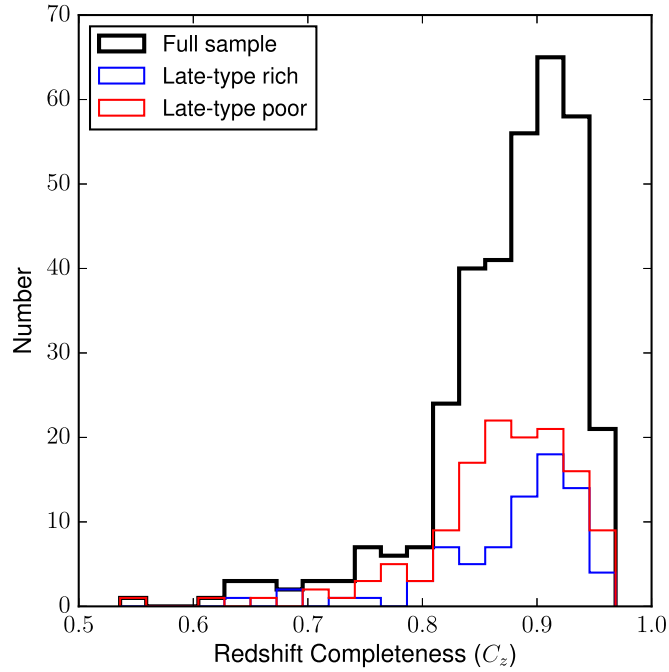


**Figure 26.** Histograms showing the distributions of redshift completeness for the (i) VoML+G catalog clusters ($N_g \geqslant 10$), shown by a thick black line; (ii) late-type-poor clusters in the statistical subset ($N_g \geqslant 10$ within $z = 0.04$–009), displayed in red; and (iii) late-type-rich clusters in the statistical subset, displayed in blue. Most of the clusters have associated $C_z$ values larger than 80%.

**Table 5**
Parameters of the Gaussian Mixture Model with Two Components

| Parameter | Estimate Full Statistical Sample | Estimate $N_g \geqslant 20$ |
|---|---|---|
| Component 1 mean | 0.28 | 0.26 |
| Component 2 mean | 0.57 | 0.57 |
| Component 1 variance | 0.014 | 0.014 |
| Component 2 variance | 0.018 | 0.0015 |
| Component 1 weight | 0.56 | 0.8 |
| Component 2 weight | 0.44 | 0.2 |

Consequently, we repeat the Gaussian mixture model fitting, but using only clusters with 20 or more members, i.e., retaining the more accurate late-type fractions only. The result of the repeated analysis is shown in Figure 25 and Table 5. Suppressing the lower-multiplicity clusters, the bimodality is clearer, and the fit of two Gaussians is also better, with an AIC of $-48$ in comparison with an AIC of $-41$ for a single-Gaussian component. The likelihood test in this case gives a $p$-value of $4 \times 10^{-4}$ (more than $3\sigma$). This corroborates and strengthens the previous analysis.

The usual approach to perform classifications when a mixture is present is to choose the limit between populations where the probability of each component is equal, which happens at $f_L = 0.43$ for the *statistical subset* and at $f_L = 0.5$ in the sample with $N_g \geqslant 20$. Because we want to avoid any potential uncertainties in the late-type fraction estimations, we adopt the limit given by the richer $N_g \geqslant 20$ sample as the limit between classes. This corresponds to a more conservative limit for the *statistical subset* and agrees with a naive separation at $f_L = 0.5$. Therefore, we separate the *statistical subset* into late-type-poor ($f_L < 0.5$) and late-type-rich ($f_L \geqslant 0.5$) clusters. Out of the 207 clusters in the subset, 131 fall in the late-type-poor class (63%) and 76 into the late-type-rich class (37%).

Before we proceed to compare the multiplicities, velocity dispersions, luminosities of brightest cluster galaxies, and spatial distribution corresponding to the late-type-poor and late-type-rich cluster classes, we check that the classification of a cluster to either of these classes is unrelated to the 2dFGRS redshift completeness ($C_z$) in the area of the sky on which it lies. The resulting median redshift completeness for both cluster classes was 0.85. Figure 26 shows the histograms of the redshift completeness, separately for late-type-poor (in blue) and late-type-rich (in red) clusters, where it can be seen that the curves are very similar and both have a peak at $C_z \sim 0.8$. A K-S test cannot refuse the null hypothesis that the two distributions are statistically identical ($p$-value $= 0.63$). The histogram in black shown in Figure 26 corresponds to the whole cluster sample and confirms its consistency with the histograms by class. We conclude that the slight variations in redshift completeness as a function of position in the sky do not favor the detection of either late-type-poor or late-type-rich clusters.

### 8.2. Multiplicity Distributions

The distribution of multiplicity for late-type-rich and late-type-poor clusters is shown in Figure 27. The mean of late-type-rich clusters is $15.7 \pm 0.9$, whereas the mean of the late-type-poor clusters is $22 \pm 1$ (38% larger). The figure shows that the late-type-poor distribution has a slightly heavier tail and is consistent with the higher multiplicities observed. We perform a K-S test for the null hypothesis of the same parent distribution, and we obtained a K-S statistic of 0.1879, which corresponds to a

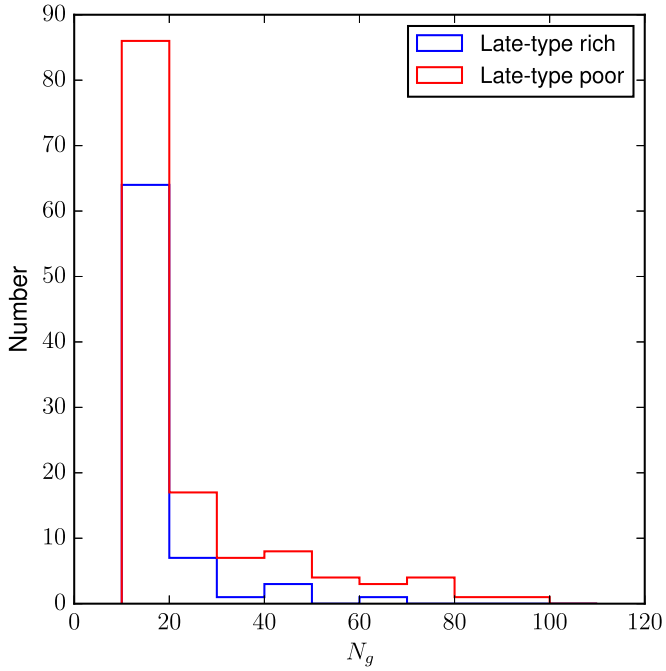


**Figure 27.** Histograms showing the distributions of multiplicity for the VoML +G statistical subset of clusters ($N_g \geqslant 10$ within $z = 0.04$–009) separated into late-type-poor (red) and late-type-rich (blue). The late-type-rich distribution has both smaller variance and average multiplicity in comparison with the late-type-poor distribution, which presents a tail toward larger multiplicities.

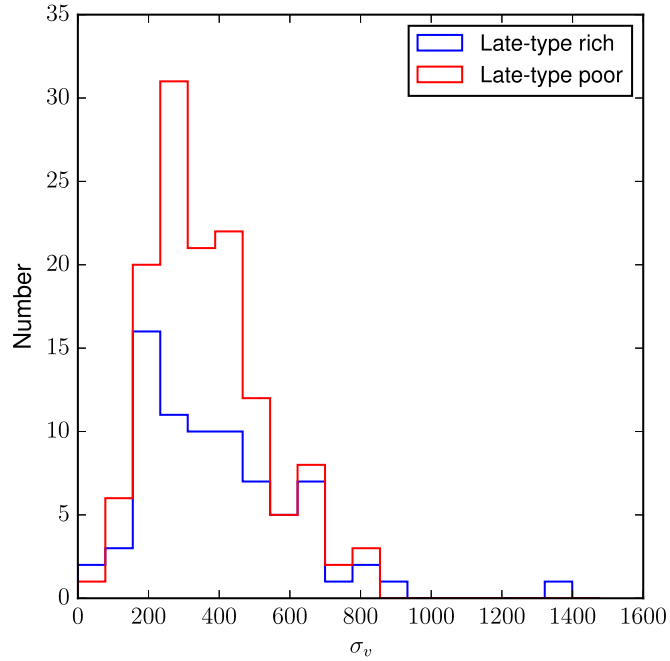


**Figure 28.** Histograms showing the distribution of (rest-frame) velocity dispersion for the VoML+G statistical subset ($N_g \geqslant 10$ within $z = 0.04$–009) separated into late-type-poor (red) and late-type-rich (blue) clusters. The late-type-poor distribution is leptokurtic (more peaked than normal), whereas the late-type-rich distribution is platykurtic (broader than normal).

$p$-value of 0.0084. Therefore, we cannot reject the hypothesis of the same parent distribution at the required $3\sigma$ significance. Nevertheless, the $p$-value is smaller than 5%, a $2\sigma$ significance level; therefore, the difference is mildly significant, and it hints at a possible difference in multiplicity between the classes, but the effect is small and cannot be confirmed with the current sample. Alternatively, the Mann–Whitney $U$ statistic for $N_g$ is 9452, with a $p$-value of 0.0017. The significance level for a one-sided test is one-half that for a two-sided test, with a $p$-value of $<0.0015$ for the $3\sigma$ level and a $p$-value of $<0.025$ for the $2\sigma$ level.

Therefore, we cannot reject the hypothesis of a same distribution at $3\sigma$, but we do reject it at the $2\sigma$ level. In summary, there is mildly significant evidence that late-type-poor clusters have larger multiplicities on average.

### 8.3. Velocity Dispersion Distributions

The distribution of $\sigma_{cz}$ for late-type-rich and late-type-poor clusters is shown in Figure 28. Figure 28 shows the presence of a late-type rich cluster with a very high velocity dispersion, over 1400 km s$^{-1}$. A visual inspection of the sky distribution and redshift distribution of its member galaxies suggests that it is not dynamically relaxed. Therefore, this very high velocity dispersion plausibly is an overestimation due to a far from equilibrium state. Consequently, we consider these two cluster as outliers, and they are removed from the following statistical estimators and analyses. Both distributions have different kurtosis, with late-type-poor clusters having an excess kurtosis of 0.08, i.e., leptokurtic (larger kurtosis than normal, therefore more peaked), whereas late-type-rich clusters have $-0.07$, i.e., platykurtic (negative excess kurtosis, flatter than a normal distribution). This indicates that the late-type-poor cluster velocity dispersion is more concentrated than for late-type-rich clusters. A possible explanation of this behavior is a bias to higher velocity dispersion due to the role of nonrelaxed dynamics in late-type-rich clusters. However, we have not explored this possibility, as it is beyond the scope of this paper. The mean velocity dispersion of late-type-poor clusters is higher ($378 \pm 12$ km s$^{-1}$) than for late-type-rich clusters ($341 \pm 16$ km s$^{-1}$). We obtain a K-S statistic of 0.2070, which corresponds to a $p$-value of 0.0028, which is smaller than our adopted significance level of 0.3%; thus, the hypothesis of *same distribution* is rejected at a $3\sigma$ significance level. Alternatively, a Mann–Whitney test with the hypothesis that the late-type-rich distribution is smaller than the late-type-poor distribution has a $U = 9717$ and a $p$-value of 0.0095; thus, the hypothesis *that the peaks of the distribution* are statistically equal is rejected at a $2\sigma$ significance level. In summary, we find mild evidence that the velocity dispersion distributions are different for late-type-rich and late-type-poor clusters.

Directly related to the velocity dispersion is the virial mass, which is computed using the Biviano et al. (2006) formula (see Pereira et al. 2017). Figure 29 shows the histogram of the logarithm of the mass (for better visualization). The difference between late-type-rich and late-type-poor is more evident in term of mass, but this is an effect of the cubic dependency on velocity dispersion of the virial mass. Again, the late-type-rich distribution is platykurtic, whereas the late-type-poor distribution is leptokurtic. Doing the statistics on the logarithm of the virial mass (in units of $h^{-1}\,M_{\rm Sun}$), we find that the mean mass for late-type-rich clusters is $12.66 \pm 0.07$ and for late-type-poor clusters is $12.91 \pm 0.04$. We perform the same statistical analysis as before, but because the masses are computed in a deterministic fashion, both variables are perfectly correlated, and the results are the same as for velocity dispersion, i.e., we find statistically significant evidence for a difference in the mass distribution in general, but only mildly significant for the

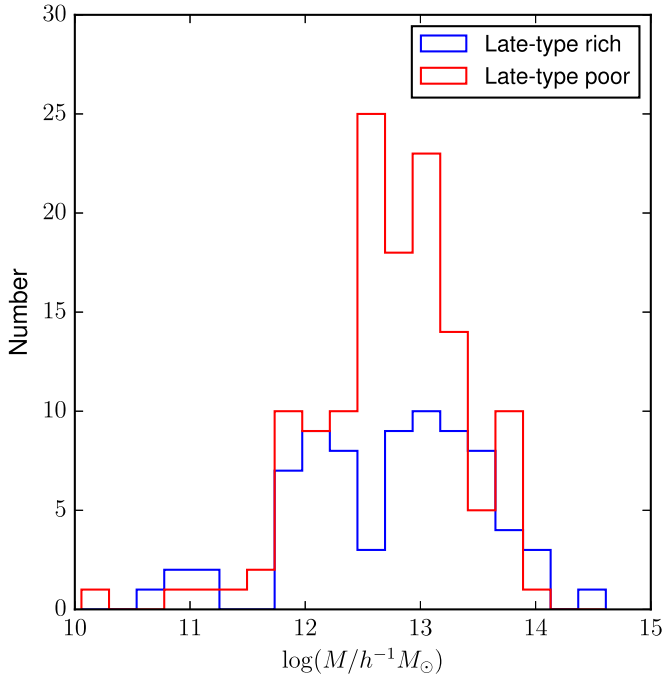


**Figure 29.** Histograms showing the distributions of the logarithm of virial masses for the VoML+G statistical subset of clusters ($N_g \geqslant 10$ within $z =$ 0.04–009) separated into late-type-poor (in red) and late-type-rich (in blue). Both distributions span the same mass range.

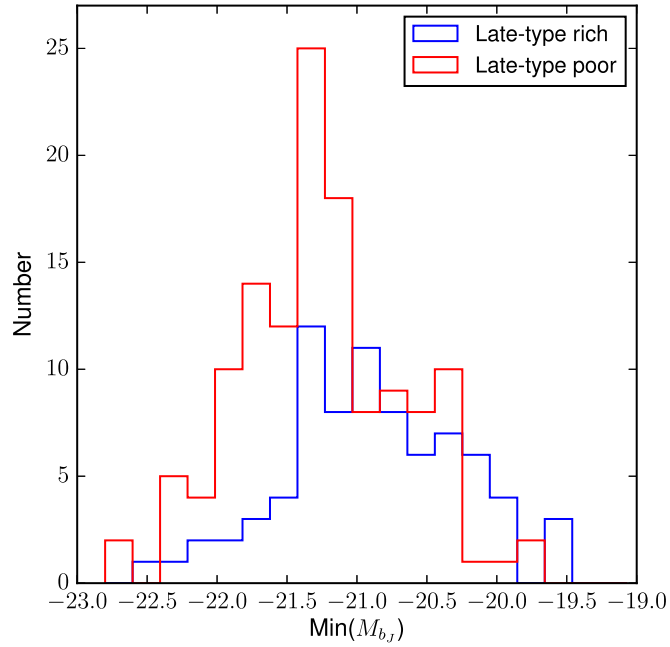


**Figure 30.** Histograms showing the distribution of minimum absolute magnitude $M_{bJ}$ of member galaxies for the VoML+G statistical subset of clusters ($N_g \geqslant 10$ within $z = 0.04$–009) separated into late-type-poor (red) and late-type-rich (blue) clusters. A K-S test rejects the hypothesis of same distribution with a significance of $\sim 3.3\sigma$.

peak (mean location) of the distribution. This difference must be studied in detail because there are unaccounted-for biases in this analysis, as we mentioned before in the case of velocity dispersion. The main bias in this analysis is the overestimation of the velocity dispersion in nonrelaxed clusters, which implies overestimation of the virial mass. This issue is beyond the scope of this paper.

### 8.4. Range of the Cluster's Brightest Galaxy

Figure 30 shows the histogram of the brightest galaxies belonging to clusters in the statistical sample ($N_g \geqslant 10$ within $z =$ 0.04–009), separately for late-type-poor and late-type-rich clusters. The median value of minimum absolute magnitudes ($M_{bJ}$) is $-21.27$ for the former and $-20.96$ for the latter; the resulting difference of 0.31 mag is very significant (the standard errors of the average values being 0.05 and 0.07 mag). The quantitative comparison between the histograms using a K-S test gives a $p$-value of 0.04% about $3.3\sigma$, allowing us to conclude with a high degree of confidence that these distributions are different.

### 8.5. Statistics on Detection Case and NED Matches

Paper I established that VoML+G finds physically valid galaxy systems without distinction by detection case. Nevertheless, for the sake of clarity and with a risk of distracting the attention of the reader from the physically relevant aspects, we now investigate how the cluster detection cases ($a$ or $b$) are related to the cluster's late-type fractions. Out of the 207 clusters of the *statistical subset*, 126 are case $a$ and 81 are case $b$. In Section 3 it was commented that as a consequence of the algorithm the case $a$ clusters are expected to be predominantly virialized clusters while the case $b$ clusters would be unvirialized and possibly with substructure.

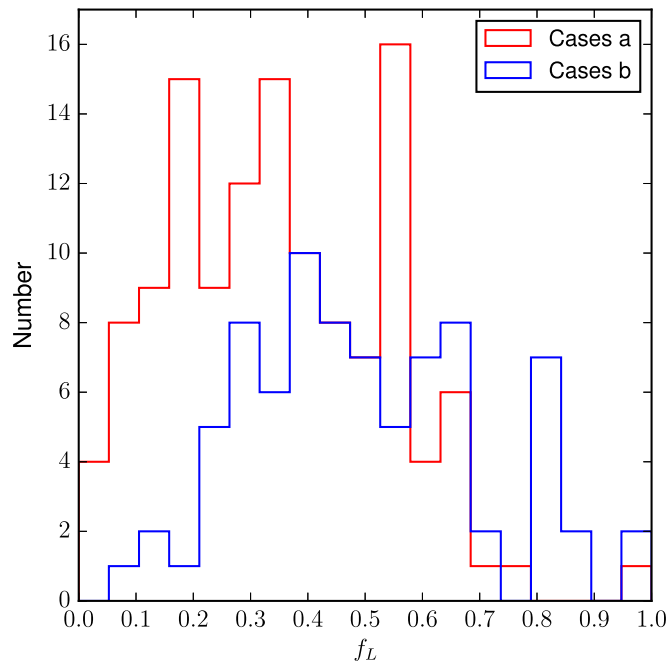


**Figure 31.** Histograms showing the distribution of late-type fraction for the case $a$ (blue) and case $b$ clusters (red). Most of the case $a$ detections are late-type-poor clusters (see the text). Case $b$ detections have late-type fraction over a wide range, resulting in similar numbers of late-type-rich and late-type-poor clusters.

Of the 76 late-type-rich clusters in the *statistical subset*, 36 are case $a$ and 40 are case $b$; thus, at least one-half of them (47%) would be virialized. Of the 131 late-type-poor clusters, 90 are case $a$ and 41 are case $b$, so a much larger fraction of them (69%) would be virialized. The distribution of case $a$ and $b$ clusters with late-type fraction is shown in Figure 31. In Figure 31 it can be seen (i) that nearly all the clusters with

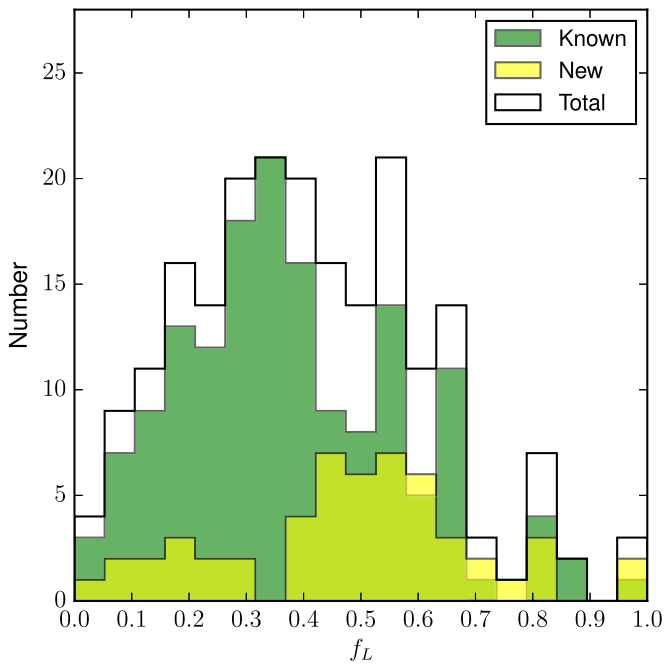


**Figure 32.** Histogram showing (thin black line) the distribution of late-type fraction for the statistical subset ($N_g \geqslant 10$ within $z = 0.04$–009). Overlapping histograms corresponding to the "known" and "new" clusters are displayed in green and yellow, respectively. Most of the "known" clusters correspond to late-type-poor clusters, whereas the "new" cluster detections include correspondences with late-type-rich and late-type-poor clusters.

$f_L \lesssim 0.25$ are case *a* detections and thus are consistent with Abell regular clusters (virialized systems with galaxy content dominated by early types); (ii) clusters with late-type fractions over the wide range 0.25–0.70 are roughly equally contributed by case *a and b* detection cases (except by two peaks in the case *a* histogram); and (iii) all the clusters (except one) with $f_L > 0.80$ are case *b* detections and thus predominantly unvirialized. Although for $f_L \lesssim 0.25$ there is a clear correlation between late-type fraction and case detection, overall the case *b* detection cases have nearly the same number of correspondences with late-type-poor and with late-type-rich clusters, indicating that complex dynamics and geometries are present in both cluster classes.

The distribution of late-type fraction for the NED matches is also valuable to understand the characteristics of the clusters that are not included in this database. The distribution is shown in Figure 32, and it is clear that the "known" clusters (in NED) are dominated by late-type-poor clusters whereas "new" clusters are evenly populated by late-type-rich and late-type-poor clusters. This suggests that the high fraction of late-type galaxies in a cluster was a likely cause for nondetection in previous surveys.

### 8.6. LSS Environment by Cluster Class

The spatial distribution of clusters belonging to the *statistical subset* is shown in Figures 33 and 34 for NGP and SGP 2dFGRS fields, respectively. The figures show the wedge diagram of late-type-rich (blue squares) and late-type-poor clusters (red squares) for the redshift interval $z = 0.04$–0.09. Qualitatively, the cluster distributions suggest that the late-type-poor clusters tend to occupy different regions on large scales in comparison to the late-type-rich clusters. We perform a preliminary analysis of the spatial clustering of late-type-rich and late-type-poor clusters through counts-in-cells (Peebles 1980) and estimate the average two-point correlation function within the volume of spherical cells for the cluster sample. We compute the counts by filling the volume uniformly with a grid of spherical cells of various sizes. Using the counts of clusters, we compute the statistic

$$\bar{\xi} = \frac{\mathrm{var}(N) - \langle N \rangle}{\langle N \rangle^2}, \tag{1}$$

where $N$ is the count variable, $\langle N \rangle$ is the mean count, and $\bar{\xi}$ corresponds to the average correlation function of the cluster distribution, or equivalently, the variance of the cluster overdensity field. The error is estimated using bootstrap resampling of the counts, which provides a better estimate of the error, although it is not entirely accurate, as the volume used might still have some cosmic variance (i.e., deviation of the fair sample due to the finite volume of the survey), and the selection effects from the galaxy survey and cluster finder have not been properly assessed. This statistic is computed for different radii and for late-type-poor, late-type-rich, and the full sample. The result is shown in Figure 35. The result is that late-type-rich clusters have higher mean correlation function than the full sample and the late-type-poor clusters, and therefore they are more clustered. At large scales late-type-poor and late-type-rich clusters tend to have similar correlations, which converge to zero for radii larger than ~110 Mpc, a situation that corresponds to the expectations, as the matter distribution is close to homogeneous at such a scale.

### 8.7. Sky Images of Late-type-rich Clusters

In this section we show two sky images corresponding to fields that contain a late-type-rich cluster. Given the unusual characteristics of the galaxy content of these clusters, it is evident that they are not morphologically similar to Abell-like clusters and that in a 2D projection they are inconspicuous. The chosen VoML+G clusters are the same selected to exemplify the application of the Madgwick (2003) criterion to separate 2dF galaxies into either of the early- or late-type bins (see Section 7.2).

The first example, VoML+G No. 135, is a late-type-rich cluster with substructure, found within the NGP section of the 2dFGRS, with 57 member galaxies, a redshift of 0.035, a line-of-sight velocity dispersion of 1276 km s$^{-1}$, and a late-type fraction of 0.68. Figure 36 shows an image from the SDSS DR9 of approximately 30 arcmin in size, centered on the calculated centroid of the cluster. The member galaxies are marked according to a code: early-type galaxies are indicated by red circles and late-type galaxies by blue squares. The absolute *B*-magnitude of the brightest member galaxy detected is −21.3. The individual member galaxies are shown in Figure 19. $R_{200}$ for this cluster is of the order of $1R_A$. The velocity distribution is flat, and the peaks suggest the presence of two clumps; it is not well represented by a single Gaussian with $\sigma_{cz} = 1276$ km s$^{-1}$. Although a case *a* detection, the cluster is not virialized.

The second example, VoML+G No. 129, is a virialized late-type-rich cluster, found within the NGP section of the 2dF, with 47 member galaxies, a redshift of 0.061, a line-of-sight velocity dispersion of 590 km s$^{-1}$, and a late-type fraction of 0.55. Figure 37 shows an image from the SDSS DR9 of approximately 20 arcmin in size, centered on the calculated

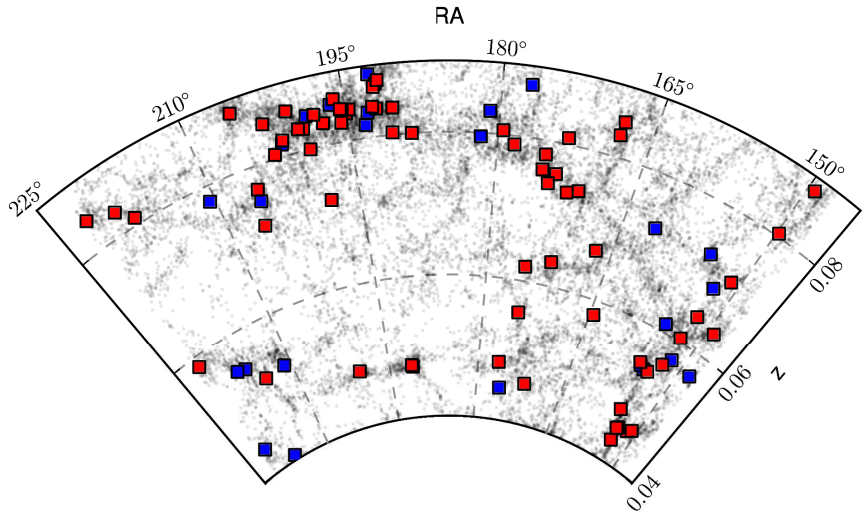


**Figure 33.** Wedge diagram showing the projected spatial distribution of the VoML+G statistical subset clusters ($N_g \geqslant 10$ within $z = 0.04$–009) in the 2dF NGP section. Filled red and blue squares represent late-type-poor and late-type-rich clusters, respectively. The gray points correspond to galaxies of the database in the displayed redshift interval.

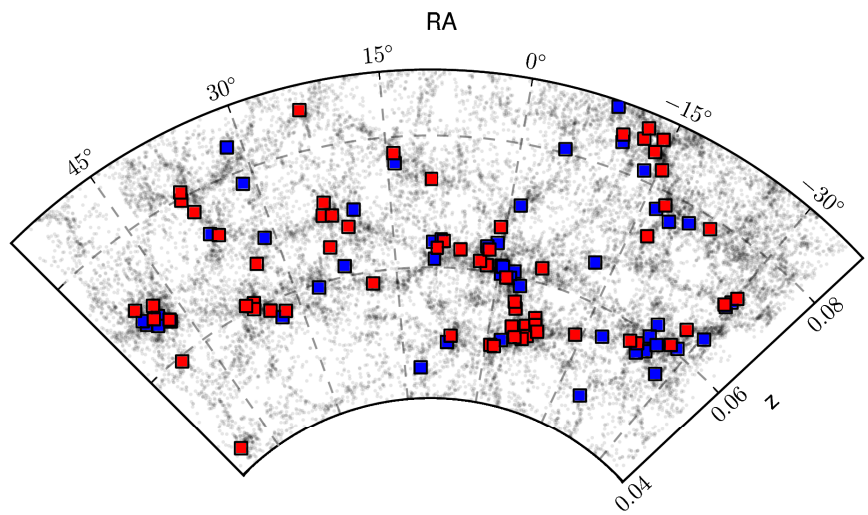


**Figure 34.** Wedge diagram showing the projected spatial distribution of the VoML+G statistical subset clusters ($N_g \geqslant 10$ within $z = 0.04$–009) in the 2dF SGP section. Filled red and blue squares are late-type-poor and late-type-rich clusters, respectively. The gray points correspond to galaxies of the database in the displayed redshift interval.

centroid of the cluster. Galaxies are coded as in Figure 36. The absolute $B$-magnitude of the brightest member galaxy detected is $-21.3$. The individual member galaxies are shown in Figure 20. The velocity distribution of cluster member galaxies is very well fit by a Gaussian, and there is a good kinematical mix between early and late types. Note that as a case *a* detection this cluster is expected to be virialized.

## 9. Discussion

The application of the VoML+G cluster finder (Paper I Pereira et al. 2017) to the real 2dFGRS identified 1901 clusters with at least two galaxies (the full sample), 341 of them with 10 or more galaxy members (the cluster catalog), which is interestingly close to the corresponding numbers obtained from its application to the mock 2dFGRS (1614 and 358, respectively). The 2dFGRS has been used to search for galaxy groups and for the study of the Abell-APM-EDCC clusters in the area of concern, thus providing the opportunity to compare them with the VoML+G detections. Because the VoML+G algorithm was designed to detect groups and clusters of galaxies, imposing minimal conditions, our expectation was that it would recognize known systems and, at the same time, that it could eventually expand the currently accepted parameter space. One of our aims is to study systems for which we can calculate masses (Paper I showed that this is possible when the number of galaxy members is at least 10), and in particular to identify a redshift range over which the algorithm has a flat selection function so as to perform a census of the $N_g \geqslant 10$ systems therein.

The general comparison of the *full catalog* with the 2PIGG catalog of galaxy groups showed that ∼90% of the VoML+G clusters (full sample) have correspondence in the 2PIGG catalog, albeit with characterizations that may differ considerably. The cross-check of the 2PIGG and VoML+G detection corresponding to a few clusters with $N_g > 45$ (VoML+G) showed that these algorithms, applied over the same data, identify nearly coincident structures only for the denser and

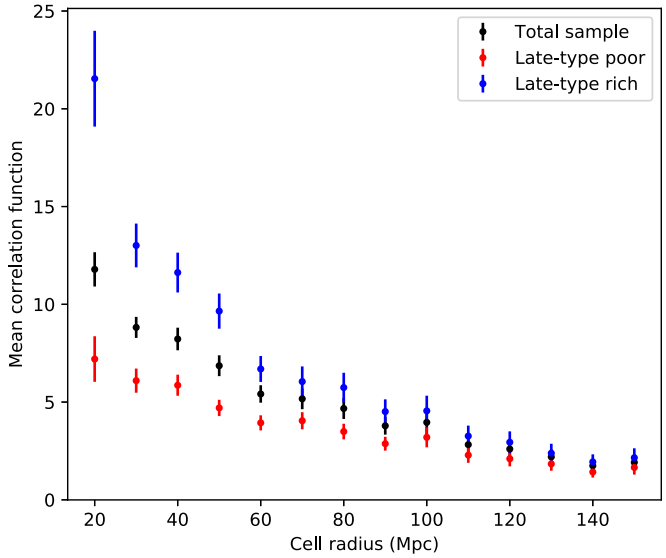


**Figure 35.** Estimated mean correlation function vs. cell radius displayed with black dots for the full statistical subset of clusters ($N_g \geqslant 10$ within $z = 0.04$–009), with red dots for the subset of late-type-poor clusters (see the text), and with blue dots for the late-type-rich cluster (see the text) subset. The error bars are computed using bootstrap resampling, which, although more accurate than Poissonian, still might be affected by the selection effects in the cluster samples and by cosmic variance. The late-type-rich clusters have a larger mean correlation function at all scales (more clustering); differences between samples become insignificant for radii larger than 110 Mpc.

isolated overdensities. The other examples highlighted a difficulty of the 2PIGG algorithm, and of FoF methods in general, in the sense that when the overdensities are not quite isolated 2PIGG merges in with the cluster core the connecting galaxy filaments.

This paper is specially focused on the clusters whose masses can be homogeneously estimated, that is, with $N_g \geqslant 10$ (the 341 catalog clusters). Before we proceeded to the identification of the *statistical subset*, we determined that ∼75% of the catalog clusters had correspondences with either clusters or groups in NED and investigated whether the "new" (not in NED) clusters had peculiar properties. We found that "new" clusters are preferentially low-multiplicity systems, not different from some of the "known" ones, but we suspected that other properties, to be determined, were at play to make them more difficult to detect by previous searches.

One of the intended capabilities of the VoML+G algorithm was the automatic rediscovery of Abell-like clusters. A test done using the 166 catalog clusters with correspondences in the DP02 study of Abell-APM-EDCC showed that our centroid, mean redshift agreed closely with the DP02 values; however, agreement between multiplicity determinations was achieved only when compared with *expanded versions* of the VoML+G clusters (from $R_{200}$ to $R_A$). Also, important discrepancies were found between velocity dispersion determinations with the DP02 values being generally larger than ours. Although DP02 used the same 1000 km s$^{-1}$ velocity gap as us, the explanation of the discrepancy resides in that they applied it over the Abell radius, resulting in a larger probability of including interlopers at large radii. Our focus on the cluster core ($R_{200}$) is the reason why we are not affected by such interlopers when estimating the velocity dispersion. It confirms that our velocity dispersions are good, and not overestimates. Further confirmation of the capacities of VoML+G to identify rich regular clusters was provided through a comparison with the REFLEX catalog, which showed that 82% of the REFLEX clusters in the area of concern have correspondences in the cluster catalog.

The next step after establishing the relation of VoML+G clusters with some of the main catalogs of clusters and groups of galaxies was to conduct an independent self-consistent study of the $N_g \geqslant 10$ VoML+G clusters contained in a volume selected to minimize the effect of detection biases. Using the subset of the 207 catalog clusters within the 0.04–0.09 redshift interval (the *statistical subset*), once restricted only to galaxy members with luminosities $M_{b_j} \leqslant -18.5$, the distribution of the fractions of late-type galaxies ($f_L$) contained in the clusters revealed cases in which the majority of the galaxies were of late type. Sixty-three percent of these clusters were found to be dominated in number by early-type galaxies (i.e., $f_L < 0.5$), with the remaining 37% being dominated by late-type galaxies ($f_L \geqslant 0.5$). The statistical analysis of the late-type fraction distribution supports, with a confidence level of $3\sigma$, the existence of two populations in the late-type fraction distribution. We found mild evidence indicating that late-type-poor clusters tend to have larger multiplicities in comparison with the late-type-rich clusters and strong evidence that the brightest cluster member galaxies in the former tend to be more luminous. Also mild evidence was found favoring a difference between the velocity dispersion distributions; however, the covered range is the same for the late-type-rich and late-type-poor clusters, which implies that the masses for these two cluster classes also span the same range.

The examination through cone diagrams of the spatial distribution of the *statistical subset* revealed a tendency of the late-type-rich clusters to reside in different, more clustered large-scale environments in comparison with the distribution of the late-type-rich clusters. Through a counts-in-cells analysis we found that late-type-rich clusters have higher mean correlation function than the full subset and than the late-type-poor clusters, indicative that they are more clustered.

The late-type-poor galaxy systems may reflect and extend the results from the studies of De Propris et al. (2004) and Fasano et al. (2012) on the fraction of blue galaxies ($f_b$) and the fraction of spiral galaxies, respectively, in samples of ∼70 Abell-APM-EDCC clusters. It is important to note that the late-type fraction distribution we find for the late-type-poor component is consistent with the suggestion by De Propris et al. (2004) that the scatter they found in the blue fraction is an intrinsic feature of the sample and with their assertion that they measure $f_b > 0.4$ for some clusters in their sample but find none attaining $f_b > 0.6$. The other considerably abundant late-type-rich component (∼one-third of the total) has less of a history but is suggested to constitute a new cluster class in the nearby universe already presaged half a century ago by Oemler (1974).

A remarkable feature encountered is that late-type-rich clusters have $\sigma_v$ values that lie in the same range of the late-type-poor clusters, which is indicative that both classes cover a similar mass range. A similar mass range for the late-type-rich cluster class, together with their average lower multiplicities and luminosities, implies that they generally are systems with higher mass–luminosity ($M/L$) than the late-type-poor clusters. The existence of a wide range in $M/L$ ratios in systems of megaparsec sizes would possibly hold clues on the nature of

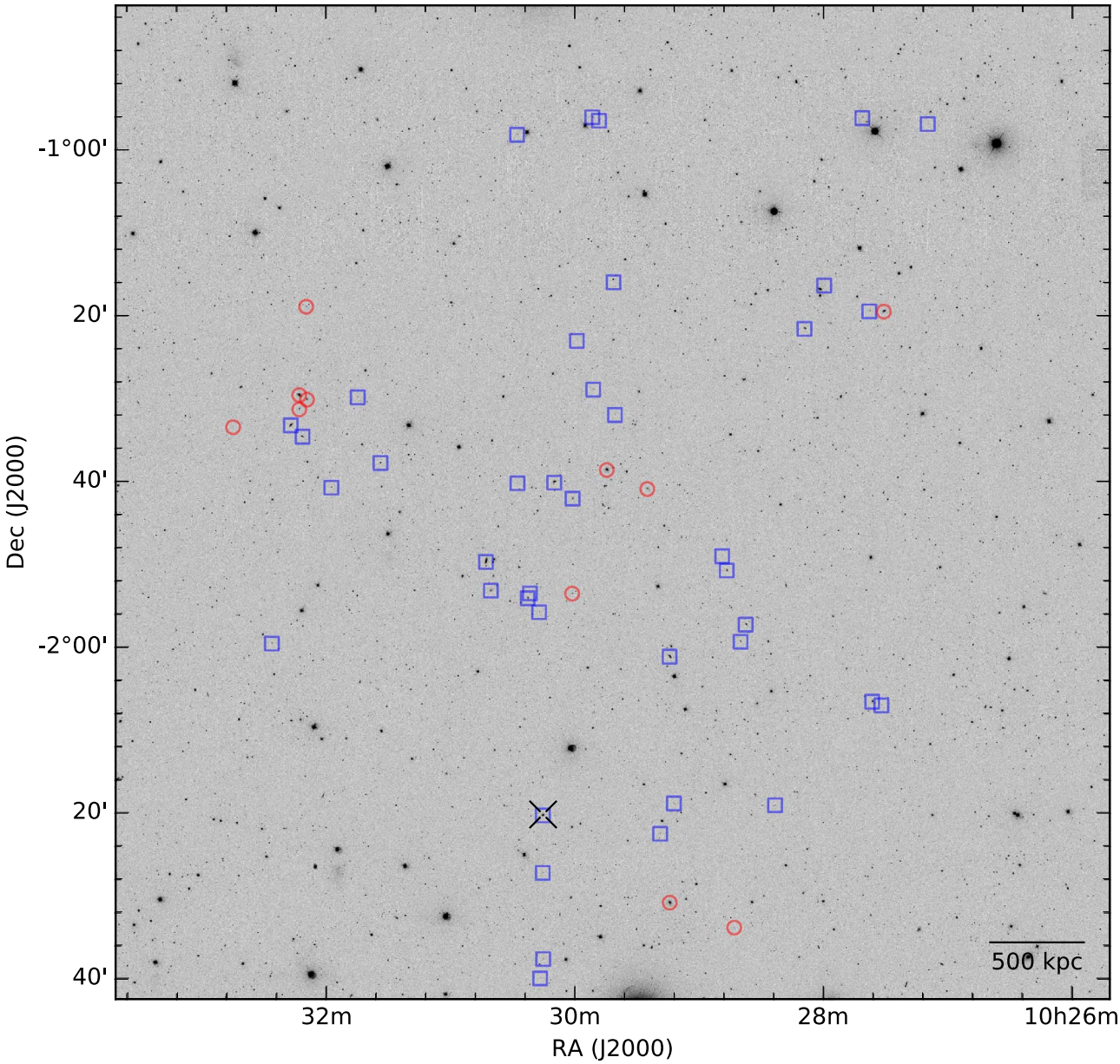


**Figure 36.** Sky view of the cluster VoML+G No. 135 (with $z = 0.035$) from an SDSS DR9 $g$-band image. The cluster detection contains 57 galaxies, the brightest one with $M_{bJ} = -21.3$ indicated by a cross-hair, represented by either blue squares or red circles that correspond to late- or early-type galaxies, respectively. The galaxy content of this cluster is dominated by late-type galaxies (68%). The cluster is not listed in NED ("new"), presumably because its less luminous, less packed galaxies, together with its galaxy contents similar to the "field," have prevented its identification in previous searches.

dark matter and the physics of galaxy formation; thus, it is of utmost importance to confirm these results with new techniques and surveys. In addition, this relatively abundant cluster class could significantly increase the contribution of clusters of galaxies to the mass budget in the low-$z$ universe.

Although a preliminary result, our analysis of the spatial clustering of late-type-poor and late-type-rich clusters suggests that the former are the less clustered. Such evidence, that late-type-poor clusters reside in different, less clustered, large-scale environments in comparison with the late-type-rich clusters, is consistent with expectations from the dark matter *halo assembly bias*. Also, the mere existence of different cluster classes implies that the common semi-empirical techniques such as Halo Occupation Distribution models (Berlind & Weinberg 2002) may need to consider galaxy morphology, star formation rate, and other aspects of galaxy evolution.

The identification of a late-type-rich cluster class further explains why the "new" clusters had escaped detection before. In addition to the role of incompleteness of catalogs and of the low multiplicity of the new clusters, there is a further element to consider, that is, that $\sim 1/2$ of them are late-type-rich clusters. The latter contain less packed, less luminous galaxies and have galaxy contents closer to the *field*, which make them hard to detect by some of the traditional methods. Another related issue is that, given the search methods employed in the past, it seems unlikely that different clustering properties of late-type-poor and late-type-rich clusters would have had a bearing on their incompleteness with respect to late-type-rich clusters.

Also, it is interesting to make (a distant) parallel with the ultrafaint galaxies. Simon & Geha (2007) developed a procedure to identify extragalactic bound star systems (very

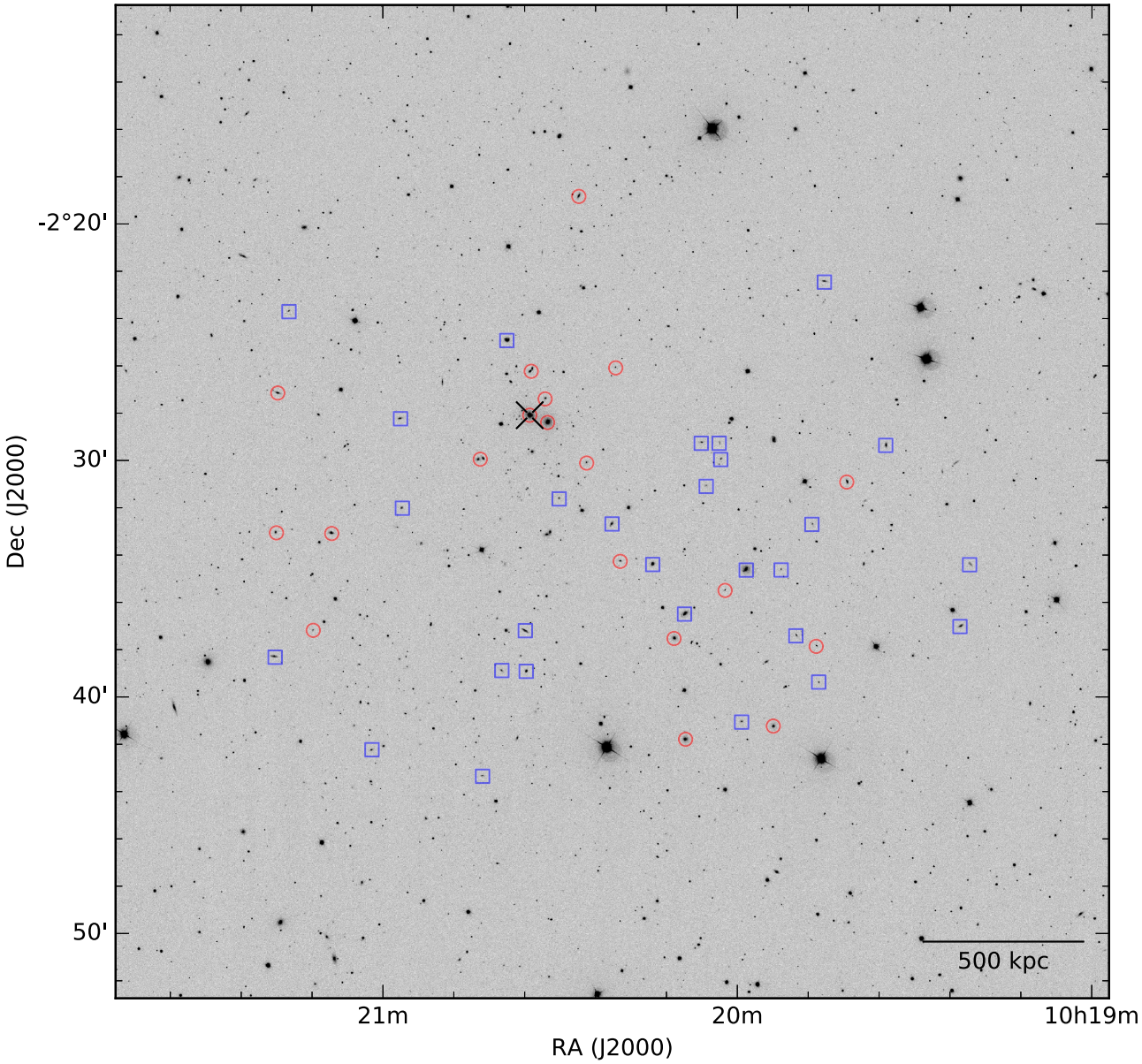


**Figure 37.** Sky view of cluster VoML+G No. 129 (with $z = 0.061$) from an SDSS DR9 $g$-band image. Symbols have the same meaning as in Figure 36. The cluster detection contains 47 galaxies, the brightest one with $M_{bJ} = -21.3$. The galaxy content of this cluster is slightly dominated by late-type galaxies (55%). This cluster has not been detected in previous searches (not in NED, "new"), presumably because of its less luminous, less packed galaxy members and its galaxy contents similar to the "field."

few dim stars) whose velocity dispersions suggest high masses (virialization assumed), concluding that they are previously unnoticed galaxies with the highest dark matter densities (highest *M/L*). Also, the claim by van Dokkum et al. (2018) of the discovery of a galaxy lacking dark matter is relevant to the possibility of extreme values of *M/L*.

The reader might ask why the 2PIGG group search and posterior work did not reveal the new cluster population that we have identified. One possible reason is that the 2PIGG algorithm was optimized to test the consistency between observed galaxy groups and the predictions for dark matter halos. Another one is that the prevalent view then was that the galaxy content of low velocity dispersion systems was dominated by late-type galaxies while the high velocity dispersions systems were typified by a dominant early-type population. And lastly, a more specific reason at play is that the 2PIGG algorithm included more tenuously associated galaxies in clusters than the VoML+G, an effect confirmed in our graphical comparisons between the respective cluster detections based on the same galaxy data set.

### 9.1. Interpretation

The dark matter *halo assembly bias* (Gao et al. 2005; Croton et al. 2007) provides a theoretical context in which to interpret the existence in the local universe of two cluster classes for which, even though their respective mass distributions run over the same range, their galaxy contents are found to be dissimilar. The dark matter *halo assembly bias* implies that the particular appearance of a cluster today depends on how galaxy properties were established during halo assembly and thus expected to be correlated with the cluster large-scale environment. Miyatake et al. (2016) have claimed a detection of dark

matter *halo assembly bias* based on galaxy clusters in a $z$-range of 0.1–0.33; in their study they defined two classes of clusters through a structural parameter (packed and loose). Our finding is reminiscent of the one by Miyatake et al. (2016), to the extent that the two cluster classes identified in the present study (late-type-rich and late-type-poor) are confirmed to be embedded in different large-scale environments.

The recent observational result by Ricciardelli et al. (2017) on the existence of correlations between a particular large-scale environment, cosmic voids, and morphological segregation of galaxies may be complementary for the interpretation of our findings.

### 9.2. Future Work

Future detailed studies of the morphology of member galaxies for a few representative VoML+G clusters through deep imaging over extended areas could be used to check our late-type fraction estimates and to characterize their large-scale environments. Of considerable importance would be the confirmation of the masses of the late-type-rich clusters through gravitational lensing studies (e.g., Cypriano et al. 2004; Saez et al. 2016), independent of the cluster's dynamical states, over extended areas. Other desirable future work includes (1) the improvement of the analysis of the selection effects in the estimation of the correlation function per cluster class, (2) the determination of limits on the X-ray emission from the late-type-rich clusters, (3) the production of versions of the VoML+G algorithm optimized to more recent redshift surveys such as the SDSS (York et al. 2000) and GAMA (Driver et al. 2011) galaxy redshift surveys, and (4) the determination of the compatibility of these cluster classes with ΛCDM cosmological numerical simulations.

## 10. Conclusions

The application of the VoML+G cluster finder (Paper I Pereira et al. 2017) to the 2dFGRS galaxy sample (Colless et al. 2001) detected 1901 clusters with two or more galaxies (the cluster *full sample*). Eighty-two percent of the Abell-APM-EDCC clusters with 10 or more galaxies as reproduced semiautomatically from the 2dFGRS by De Propris et al. (2002, DP02) have matches in the VoML+G *full sample*.

The VoML+G *full sample* contains 341 clusters with 10 or more member galaxies, which constitute the *cluster catalog* presented in Table 2. Ninety-nine percent of them have one or more correspondences in a well-studied catalog of galaxy groups (2PIGG catalog; Eke et al. 2004) detected automatically in the same 2dFGRS database. Approximately 25% of the catalog clusters are "new" in the sense that they have not been reported previously in the literature either as groups or as clusters (NED).

From the comparison performed for six sky regions containing VoML+G clusters with high multiplicity ($N_g > 45$), we found that the 2PIGG algorithm generates nearly the same galaxy configuration when the cluster has a dense central region and is surrounded by relatively few galaxies. When the overdensities are not quite isolated, the 2PIGG algorithm merges in connecting filaments with the cluster core, an effect that highlights a difficulty of 2PIGG, and of FoF methods in general. We conclude that the focus of the VoML+G algorithm on the cluster core ($R_{200}$) is the main reason why our detections largely avoid this difficulty.

Nearly 50% of the clusters (166) in the VoML+G cluster catalog have correspondences with the Abell-APM-EDCC clusters in the list by DP02. The mean-$z$ matching has an rms of 0.001 (3% relative rms). The comparison of multiplicities of the *expanded* versions of the VoML+G clusters, from $R_{200}$ to $1R_A$, with the DP02 determinations reveals a linear relation with an rms scatter of 21 km s$^{-1}$. The determined VoML+G radial velocity dispersions ($\sigma_{cz}$) are considerably smaller than the DP02 values by an average percentage difference of 93%. Although DP02 used the same 1000 km s$^{-1}$ velocity gap as us, the explanation of the discrepancy resides in that they applied it over the Abell radius, resulting in a larger probability of including interlopers at large radii. We conclude that our focus on the cluster core ($R_{200}$) is the reason why we are not affected by such interlopers when estimating the velocity dispersion.

Using the *statistical subset* consisting of 207 clusters, we found that (i) 63% of the systems are dominated by early-type galaxies (i.e., the late-type-poor clusters, $f_L < 0.5$) with a corresponding mean multiplicity and logarithmic virial mass (in units of $M_\odot$) of 22 ± 1 and 12.91 ± 0.04, respectively; and (ii) 37% of the systems are dominated by late-type galaxies (i.e., the late-type-rich clusters, $f_L \geqslant 0.5$) with a corresponding mean multiplicity and logarithmic virial mass (in units of $M_\odot$) of 15.7 ± 0.9 and 12.66 ± 0.07, respectively. The late-type-rich clusters have less conspicuous brightest member galaxies (by about 1 mag) and fewer and less spatially concentrated galaxies. The statistical analysis of the data corresponding to the *statistical subset* supports, with a $3\sigma$ confidence level, the notion of two cluster populations in the late-type fraction distribution, a conclusion that is reinforced when only clusters with more than 20 galaxies are retained for the analysis. We found mild evidence indicating that late-type-poor clusters tend to have larger multiplicities in comparison with the late-type-rich clusters and strong evidence that their brightest cluster galaxies tend to be more luminous. The velocity dispersion distributions for late-type-rich and late-type-poor clusters span the same range of values, while the statistical tests provide mild evidence that these distributions are different.

We used a counts-in-cells analysis (Peebles 1980) to estimate the average two-point correlation function within the volume of spherical cells for the *statistical subset* and separately by cluster class. We found that late-type-rich clusters have a higher mean correlation function than the full *statistical subset* and than the late-type-poor clusters, thus tentatively conclude that they are more clustered.

The identification of a late-type-rich cluster class plausibly provides the remaining argument as to why the "new" clusters had escaped detection before. In addition to the role of incompleteness of catalogs and of the low multiplicity of the "new" clusters, the fact that $\sim 1/2$ of them are late-type-rich clusters may be the clue to this riddle, the reason being that the latter contain less packed, less luminous galaxies and have galaxy contents closer to the *field*, which make them hard to detect by the traditional methods.

The presence of an abundant ($\sim 1/3$ of the total), previously unnoticed population of massive clusters in the local universe can be further tested through deep imaging of a subset of the late-type-rich clusters. Such observations over extended areas could be used to determine late-type fractions, to characterize their LSS environments, and to deduce mass distributions by gravitational lensing studies of these areas. Furthermore, of considerable interest would be studies to establish limits on the

X-ray emission from the late-type-rich clusters and to search for similar cluster populations in the SDSS and/or GAMA galaxy redshift surveys through the VoML+G algorithm.

The dark matter *halo assembly bias* (Gao et al. 2005; Croton et al. 2007) is a possible theoretical context in which to interpret these new findings.

We foremost acknowledge the 2dF scientific initiative, made possible by the UK and Australian funding agencies, the staff of the Anglo-Australian Observatory, and the 2dFGRS team for carrying out this superb project and for the setup of a public database. This research has made use of NASA's Astrophysics Data System, as well as the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. We acknowledge the use of the SDSS public database. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

S.P., L.E.C., and N.H.-K. received partial support, and C.P.H. was funded, through CONICYT Anillo project ACT-1122. L.E.C. and G.M. acknowledge partial support from CONICYT project Basal AFB-170002. The initial development of the algorithm was supported by Fondecyt grant Nos. 1040499 and 7040066 (L.E.C., R.G.C. and D.P.). It is also a pleasure to thank Eduardo Marcelo Campusano, Hernán Pulgar, Harold Francke, and Carlos Hervías for help provided in the initial stages of this work. We acknowledge encouraging discussions with Paolo Coppi and Joaquín Prieto. Last but not least, we acknowledge the comments by the anonymous referee that greatly contributed to the improvement of the paper.

## References

Abell, G. O. 1962, in IAU Symp. 15, Problems of Extra-galactic Research, ed. G. C. McVittie (New York: Macmillan), 213
Abell, G. O., Corwin, H. G., Jr., & Olowin, R. P. 1989, ApJS, 70, 1
Akaike, H. 1973, in Proc. 2nd Int. Symp. Information Theory, ed. B. N. Petrov & F. Csaki (Budapest: Akademiai Kiado), 267
Allard, D., & Fraley, C. 1997, J. Am. Stat. Assoc., 92, 1485
Bahcall, N. A. 2000, in Allen's Astrophysical Quantities, ed. A. N. Cox (4th ed.; New York: AIP Press), 613
Baugh, C. M., Croton, D. J., Gaztañaga, E., et al. 2004, MNRAS, 351, L44
Benson, B. A., de Haan, T., Dudley, J. P., et al. 2013, ApJ, 763, 147
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Biviano, A., Murante, G., Borgani, S., et al. 2006, A&A, 456, 23
Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, A&A, 425, 367
Boselli, A., & Gavazzi, G. 2006, PASP, 118, 517
Clerc, N., Pierre, M., Pacaud, F., & Sadibekova, T. 2012, MNRAS, 423, 3545
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Colless, M., Peterson, B. A., Jackson, C., et al. 2003, arXiv:astro-ph/0306581
Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303
Cucciati, O., Marinoni, C., Iovino, A., et al. 2010, A&A, 520, A42
Cypriano, E. S., Sodré, L., Jr., Kneib, J.-P., & Campusano, L. E. 2004, ApJ, 613, 95
Dalton, G. B., Maddox, S. J., Sutherland, W. J., & Efstathiou, G. 1997, MNRAS, 289, 263
De Propris, R., Colless, M., Peacock, J. A., et al. 2004, MNRAS, 351, 125
De Propris, R., Couch, W. J., Colless, M., et al. 2002, MNRAS, 329, 87, (DP02)
Dressler, A. 1980, ApJ, 236, 351
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
Eke, V. R., Baugh, C. M., Cole, S., et al. 2004, MNRAS, 348, 866
Fasano, G., Marmo, C., Varela, J., et al. 2006, A&A, 445, 805
Fasano, G., Vanzella, E., Dressler, A., et al. 2012, MNRAS, 420, 926
Feigelson, E. D., & Babu, G. J. 2012, Modern Statistical Methods for Astronomy (Cambridge: Cambridge Univ. Press)
Flaugher, B. 2005, IJMPA, 20, 3121
Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66
Haines, C. P., Pereira, M. J., Smith, G. P., et al. 2015, ApJ, 806, 101
Henry, J. P., Evrard, A. E., Hoekstra, H., Babul, A., & Mahdavi, A. 2009, ApJ, 691, 1307
Ivezić, Ž., Connelly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy (Princeton, NJ: Princeton Univ. Press)
Jing, Y. P., Mo, H. J., & Börner, G. 1998, ApJ, 494, 1
Lewis, I., Balogh, M., De Propris, R., et al. 2002, MNRAS, 334, 673
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Lumsden, S. L., Nichol, R. C., Collins, C. A., & Guzzo, L. 1992, MNRAS, 258, 1
Madgwick, D. S. 2003, MNRAS, 338, 197
Madgwick, D. S., Lahav, O., Baldry, I. K., et al. 2002, MNRAS, 333, 133
Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, MNRAS, 406, 1759
Merchán, M., & Zandivarez, A. 2002, MNRAS, 335, 216
Miyatake, H., More, S., Takada, M., et al. 2016, PhRvL, 116, 041301
More, S., Kravtsov, A. V., Dalal, N., & Gottlöber, S. 2011, ApJS, 195, 4
Nussinov, S., & Zhang, Y. 2018, arXiv:1807.00846
Oemler, A., Jr. 1974, ApJ, 194, 1
Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144
Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Pereira, S., Campusano, L. E., Hitschfeld-Kahler, N., et al. 2017, ApJ, 838, 109, (Paper I)
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Proctor, R. N., de Oliveira, C. M., Dupke, R., et al. 2011, MNRAS, 418, 2054
Ricciardelli, E., Cava, A., Varela, J., & Tamone, A. 2017, ApJL, 846, L4
Robotham, A., Phillipps, S., & de Propris, R. 2010, MNRAS, 403, 1812
Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, ApJ, 708, 645
Saez, C., Campusano, L. E., Cypriano, E. S., Sodré, L., & Kneib, J.-P. 2016, MNRAS, 460, 4453
Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20
Seljak, U. 2000, MNRAS, 318, 203
Simon, J. D., & Geha, M. 2007, ApJ, 670, 313
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Natur, 435, 629
Tago, E., Einasto, J., Saar, E., et al. 2006, AN, 327, 365
van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Natur, 555, 629
Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, ApJ, 692, 1060
Wen, Z. L., Han, J. L., & Yang, F. 2018, MNRAS, 475, 343
Whitbourn, J. R., & Shanks, T. 2014, MNRAS, 437, 2146
Yang, X., Mo, H. J., van den Bosch, F. C., & Jing, Y. P. 2005, MNRAS, 356, 1293
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791
Zheng, Z., Zehavi, I., Eisenstein, D. J., Weinberg, D. H., & Jing, Y. P. 2009, ApJ, 707, 554