Strong [O\textsc{iii}] $\lambda$5007 Emission-line Compact Galaxies in LAMOST DR9: Blueberries, Green Peas, and Purple Grapes

Siqi Liu$^{1,2}$, A-Li Luo$^{1,2,3}$, Huan Yang$^{4,5}$, Shi-Yin Shen$^{6,7}$, Jun-Xian Wang$^{4,8}$, Hao-Tong Zhang$^1$, Zhenya Zheng$^{6,7}$, Yi-Han Song$^{1}$, Xiao Kong$^1$, Jian-Ling Wang$^1$ and Jian-Jun Chen$^{1,2}$

$^1$CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, People’s Republic of China; lal@nao.cas.cn
$^2$University of Chinese Academy of Sciences, Beijing 100049, People’s Republic of China
$^3$College of Computer and Information Management & Institute for Astronomical Science, Dezhou University, Dezhou 253023, People’s Republic of China
$^4$CAS Key Laboratory for Researches in Galaxies and Cosmology, University of Science and Technology of China, Hefei, Anhui 230026, People’s Republic of China
$^5$Las Campanas Observatory, Carnegie Institution of Washington, Casilla 601, La Serena, Chile
$^6$Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai, 200030, People’s Republic of China
$^7$Key Lab for Astrophysics, Shanghai, 200034, People’s Republic of China
$^8$School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, People’s Republic of China

Received 2021 November 30; revised 2021 December 31; accepted 2022 January 12; published 2022 March 4

Abstract

Green Pea and Blueberry galaxies are well known for their compact size, low mass, strong emission lines, and analogs to high-$z$ Lyα-emitting galaxies. In this study, 1547 strong [O\textsc{iii}] $\lambda$5007 emission-line compact galaxies with 1694 spectra are selected from LAMOST DR9 at the redshift range from 0.0 to 0.59. According to the redshift distribution, these samples can be separated into three groups: Blueberries, Green Peas, and Purple Grapes. Optical [Mg\textsc{ii}] $\lambda$2800 line feature, BPT diagram, multiwavelength spectral energy distribution (SED) fitting, mid-IR (MIR) color, and MIR variability are deployed to identify 23 active galactic nucleus candidates from these samples, which are excluded for the following star formation rate (SFR) discussions. We perform the multiwavelength SED fitting with GALEX UV and WISE MIR data. Color excess from the Balmer decrement shows that these strong [O\textsc{iii}] $\lambda$5007 emission-line compact galaxies are not highly reddened. The stellar mass of the galaxies is obtained by fitting LAMOST calibrated spectra with the emission lines masked. We find that the SFR is increasing with the increase of redshift, while for the sources within the same redshift bin the SFR increases with mass with a similar slope to the star-forming main sequence. These samples have a median metallicity of $12 + \log(O/H) = 8.10$. The metallicity increases with mass, and all the sources are below the mass–metallicity relation. The direct-derived $T_e$-based metallicity from the [O\textsc{iii}] $\lambda$4363 line agrees with the empirical N2-based empirical gas-phase metallicity. Moreover, these compact strong [O\textsc{iii}] $\lambda$5007 lines are mostly in a less dense environment.

Unified Astronomy Thesaurus concepts: Compact galaxies (285)

1. Introduction

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Wang et al. 1996; Su & Cui 2004), located at the Xinglong Observatory in Hebei Province northeast of Beijing, has an effective aperture of 4 m. Comprising 4000 fibers, the LAMOST facility can collect multiple spectra at the same time. The LAMOST extragalactic survey is to investigate the extragalactic objects for galaxies and QSOs (Zhao et al. 2012). Figure 1 shows our selected strong [O\textsc{iii}] $\lambda$5007 emission-line compact galaxies in Galactic coordinates, which will be discussed in Section 2.2 in detail.

Green Pea galaxies were first identified in the Galaxy Zoo project (Cardamone et al. 2009) by their unique compact size and green color caused by strong [O\textsc{iii}] $\lambda$5007 emission lines. Cardamone et al. (2009) found 251 Green Pea galaxies from Sloan Digital Sky Survey (SDSS) DR7 by photometric color criteria, where 80 of the sources have spectroscopic data. Following studies have enlarged the sample size for the Green Pea galaxies. Izotov et al. (2011) contain 805 star-forming luminous compact galaxies (LCGs) with high H$\beta$ luminosities in the redshift range $z = 0.02 - 0.63$ from SDSS DR7 spectroscopic data. Jiang et al. (2019) contain 800+ Green Pea galaxies from the spectroscopic database of SDSS DR13 and use these samples to do direct $T_e$-based metallicity calibration.

Blueberry galaxies at $z < 0.05$ are the lowest-mass young starburst galaxies (Yang et al. 2017) where the [O\textsc{iii}] $\lambda$5007 emission line is located within the g band and the H$\alpha$ emission line is located within the r band. Purple Grape galaxies (Izotov et al. 2011; Brüggen et al. 2020) are galaxies where either the [O\textsc{iii}] $\lambda$5007 emission line is located in the i band at $z > 0.36$ and the UV continuum is redshifted to the g band or the [O\textsc{iii}] $\lambda$5007 emission line is located in the g band and the H$\alpha$ emission line is located in the i band ($0.112 < z < 0.36$).

In this work, we compile a large catalog consisting of strong [O\textsc{iii}] $\lambda$5007 emission-line compact galaxies with 1694 spectra from LAMOST DR9 at the redshift range from 0.0 to 0.722. Among the sources, 219 have SDSS spectra detections from SDSS DR16 (Blanton et al. 2017; Ahumada et al. 2020). In conjunction with multiband photometry, we can systematically learn about the properties for these high star formation rate (SFR), compact galaxies spanning a wide range of redshift coverage.

In Section 2, we describe sample selection criteria, how we do the flux calibration of the LAMOST spectra, and the multiwavelength data set we use for the following discussions. In Section 3, we describe the multiwavelength spectral energy...
distribution (SED) fitting result and how we identify the active galactic nucleus (AGN) candidates from the samples. In Section 4, we discuss the physical properties for these sources, including the stellar mass of the galaxies from spectral fitting, the SFR, the gas-phase metallicity, and the environment. In Section 5, we summarize the results addressed in this work.

In this work, we use Wilkinson Microwave Anisotropy Probe (WMAP) 9 cosmology (Hinshaw et al. 2013), AB magnitudes (Oke & Gunn 1983), a Chabrier initial mass function (IMF; Chabrier 2003), and Bruzual & Charlot (2003) stellar population synthesis models.

2. Sample and Data

2.1. Sample

There are mainly two ranges in which we select the strong [O III] $\lambda$5007 samples from the LAMOST DR9 of the extragalactic survey based on the color selection: from both dedicated and nondedicated “Green Pea” targets of the input catalog.

The dedicated “Green Pea” targets in the input catalog originate from the PI project of the LAMOST extragalactic survey add-on program covering a large area of the North Galactic Cap and a strip in the South Galactic Cap. By now, 2309 spectra have been observed in LAMOST DR9. The initial color criteria determined from SDSS DR12 ugriz photometry are as follows. For the sources in $z < 0.12$, where the [O III] $\lambda$5007 is located within the $g$ band, the criteria are twofold: loose color criteria with $u - g \leq 0.3$, $r - g \leq 0.1$, and $i - g \leq -0.7$ but requiring type = 3 to be a galaxy, or no constraint on morphology but requiring strict color criteria with $u - g \leq 0.5$, $r - g \leq 0.5$. For the sources in $0.12 < z < 0.4$, where the [O III] $\lambda$5007 is located within the $r$ band, the criteria are displayed in the following equation (the same as Equations (1)–(5) in Cardamone et al. 2009):

$$
\begin{align*}
  u - r & \leq 2.5, \\
  r - i & \leq -0.2, \\
  r - z & \leq 0.5, \\
  g - r & \geq r - i + 0.5, \\
  u - r & \leq 2.5(r - z). \\
\end{align*}
$$

For the nondedicated “Green Pea” targets in the input catalog (by “nondedicated” we mean that the initial purpose for observing this target was not intended for observing Green Pea galaxies, Blueberry galaxies, or other compact emission-line galaxies), we select an extra 23 Blueberry galaxies (25 spectra), 19 Green Pea galaxies (19 spectra), and 2 Purple Grape galaxies (3 spectra) using the same color criteria as in Equation (1), as displayed in Figure 2.

Next, we remove the sources with a large radius in $r$ band whose petroRad_r > 5". We only keep the sources where the mode keyword is equal to 1 from the PhotoObjAll in SDSS DR13, which means the primary object. Furthermore, we visually inspect the SDSS images and only keep the isolated sources. For the last step, we visually inspect each spectrum to ensure that the strong [O III] $\lambda$5007 emission lines lie within the spectra and only keep the sources where the flux of the [O III] $\lambda$5007 line is above $3 \times 10^{17}$ erg s$^{-1}$ cm$^{-2}$.

Finally, a total of 1547 strong [O III] $\lambda$5007 emission-line compact galaxies with 1694 spectra are selected. The LAMOST pipeline has classified these sources into four types: 1642 GALAXY, 3 STAR, 35 QSO, and 14 Unknown.

Figure 3 shows the spectral redshift distribution of these samples. The majority of our sources are unresolved or barely resolved, where the Petrosian radius of the source is close to the FWHM of the point-spread function (PSF) as displayed in Figure 4. We display our sources in the $u - r$ vs. $Mr$ color–magnitude diagram in Figure 5. As we expected, the majority of the strong [O III] $\lambda$5007 galaxies are located in the blue
cloud region of the color–magnitude diagram under the division line defined in Baldry et al. (2004).
All the sources within the redshift range $0.59 < z < 0.73$ are AGN candidates, whose selection criteria will be discussed in Section 3.

2.2. LAMOST Spectra Flux Calibration

The 16 LAMOST spectrographs are designed to have a theoretical resolution of $R = 1800$, covering the wavelength range of 3600–9000 Å. The blue portion spectrograph covers the wavelength range of 3690–5900 Å, and the red portion spectrograph spans the wavelength coverage of 5700–9100 Å, with 200 Å coverage overlap (Luo et al. 2012).
The response curves of the 16 spectrographs have been removed from LAMOST spectra, but the spectra is not physically calibrated (Du et al. 2016). We recalibrate the LAMOST spectra according to the SDSS $gri$ photometry following Wang et al. (2018) as described below: LAMOST spectra are convolved with the SDSS $gri$ filters (Fukugita et al. 1996), to obtain the synthetic magnitude for these three bands, and then compared with the SDSS photometric magnitude. A zeroth-order or first-order polynomial is used to fit the magnitude difference array and apply this correction to the LAMOST spectra. An example of the flux calibration result is shown in Figure 6.

For some sources in which the LAMOST data reduction pipeline did not calculate their redshift, we need to determine the redshift by fitting the emission lines.

2.3. UV and MIR Photometry

We cross-match the sources with the Galaxy Evolution Explorer (GALEX) mission (Martin et al. 2005) and the revised GALEX catalog of UV sources (GUVCat_AIS) from GR6+7 (which contains 82,992,086 sources; Bianchi et al. 2017) using a radius of 3″ and have obtained 1266 cross-matched results.
The Wide-field Infrared Survey Explorer (WISE: Wright et al. 2010, 2019) mission has conducted a mid-IR (MIR) survey with 3.4, 4.6, 12, and 22 μm bandpasses. We cross-match our sources in 10″ separation with the ALLWISE source catalog, which contains photometry and astrometry of over 747 million objects. Although this separation is large, we have checked by eye every source to ensure that there is no contamination source within the search radius. We perform the following criteria in signal-to-noise ratio (S/N) and $\chi^2$ space to remove the fake objects from the cross-matched results, as Koenig & Leisawitz (2014, Section 3.1.1) have performed. There are 138 sources after cross-matching with WISE.
We (Capek 2019) cross-matched with Spitzer Space Telescope (Werner et al. 2004) data with Spitzer Enhanced Imaging Products using 3″-radius separation and obtained 32 sources that have detections in the Infrared Array Camera (Fazio et al. 2004) and 37 that have detections in the Multiband Imaging Photometer for Spitzer (MIPS) (Rieke et al. 2004). We use
galaxies as the majority of the sources are located in the upper left classi
diagram in Figure 9. We do not consider that all the sources
are AGNs. We plot the distribution for these galaxies in the BPT
Kauffmann et al.
types according to the de

Spitzer and WISE photometry for multiband SED fitting as
described in the next subsection.

3. Multiwaveband SED Fitting and AGN Candidates

3.1. CIGALE SED Fitting

We fit the sources with multiwavelength photometry with CIGALE (Burgarella et al. 2005; Noll et al. 2009; Boquien et al. 2019) trying to use the full photometric data for the 138 sources. As described in the above section, however, GALEX (97 sources) and Spitzer (5 sources) data are available for some of them, in addition to SDSS and WISE photometry for the 138 sources. For the configuration of the SED creation modules, we use the delayed \( \tau \) star formation history, BC03 (Bruzual & Charlot 2003), Chabrier IMF (Chabrier 2003), nebular emission lines, the dust attenuated modified starburst model, the dust emission model from Casey (2012), and the Fritz et al. (2006) AGN model. An example of the CIGALE SED fitting result is shown in Figure 7.

3.2. AGN Candidates

We identify AGNs from our samples using five aspects: (1) the existence of \([\text{Mg} \, \text{II}] \lambda 2800\) emission lines in the optical spectra (as addressed in Section 2.1), (2) the BPT diagram, (3) AGN fraction determination from CIGALE multiwavelength SED fitting, (4) MIR color, and (5) MIR variability.

For the optical spectra, we spectroscopically confirm 19 AGNs with 21 spectra from LAMOST DR9 with the existence of \([\text{Mg} \, \text{II}] \lambda 2800\) emission lines. We display the stacked spectra with three-epoch \([\text{Mg} \, \text{II}] \lambda 2800\) detections in Figure 8 and name the source by its LAMOST designation J162736.10+56225.8, where we have three spectra from LAMOST.

Using the BPT diagram (Baldwin et al. 1981; Veilleux & Osterbrock 1987), we classify the galaxies into different spectral types according to the definitions in Kewley et al. (2001), Kauffmann et al. (2003), and Kewley et al. (2006). There are 1423 galaxies (1547 spectra) that are classified as star-forming galaxies, 16 galaxies (16 spectra) as composite, and 71 galaxies (73 spectra) as AGNs. We plot the distribution for these galaxies in the BPT diagram in Figure 9. We do not consider that all the sources classified as AGNs through the BPT diagram are real AGN sources, as the majority of the sources are located in the upper left region of the diagram, where the high ionization is caused by the high \([\text{O} \, \text{III}]\) lines excited from the star-forming region, not the nuclear region. On the other hand, there might be AGN contaminants located in the star-forming region, especially in the low-metallicity regime as addressed in Harish et al. (2021).

We have confirmed the existence of AGNs from the CIGALE multiwaveband photometry fitting result, where we use the Fritz et al. (2006) model. Among the 138 sources that have multiwavelength photometry, we have 2 sources where the AGN fraction is over 20%. An example of the SED fitting result is shown in Figure 7. For the MIR color selection, we follow the criteria in Equation (1) of Jarrett et al. (2011) and have obtained 23 AGN candidates under these criteria. Multiepoch exposures in the WISE catalog can be obtained from the AllWISE Multiepoch Photometry Table and NEO-WISE-R Single-Exposure Source (L1b) Source Table. Some filters are applied to select out the reliable detections in the good-quality frames, explained in detail in the Explanatory Supplements.\(^{11}\) The light curves of the selected sources are displayed in Figures 10 and 11. There are 23 sources with the

\(^{11}\) https://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec3_2.html and https://wise2.ipac.caltech.edu/docs/release/wise/expsup/sec2_3.html. Here we select the sources with quality_score~>0, q1_fact~>0, saa_sep~>0, moon_masked~0, cc_flag~0, w1rchi2_ep~<2, and w2rchi2_ep~<2.
in the observed frame with and exclude these 25 spectra (Figure 10. The light curve of one AGN candidate of W1 (top panel) and W2 (bottom panel), where the title marks the ObsID from LAMOST. The Pearson $r_{12}$ correlation is 0.93 with a p-value of 0.0. The single-exposure magnitudes are marked with gray scatter points, where the median value and the standard deviation of the exposures from the same epoch within 60 days are marked with black error bars.

$r_{12} > 0.6$ that meet the criteria in Harish et al. (2021) as MIR variables and cover the WISE and NEOWISE detections of at least nine epochs.

Furthermore, we have identified a strong X-ray emission source from the strong [O III] $\lambda$5007 emission-line galaxies after cross-matching with 4XMM-Newton DR11 (Webb et al. 2020; Traulsen et al. 2020). This source has the EPIC (indicates parameters combined from those from the available cameras) broadband energy at 0.2–12.0 keV of $1.033 \times 10^{43}$ erg s$^{-1}$, which is higher than the luminosity of the AGNs in nearby galaxies addressed in Maitra et al. (2019). The source with $L_{0.5-4keV} = 9.27 \times 10^{42}$ erg s$^{-1}$ is higher than the AGN identification criteria for X-ray source in Section 4.4 of Xue et al. (2011) that a source with an intrinsic X-ray luminosity of $L_{0.5-8keV} = 3 \times 10^{42}$ erg s$^{-1}$ is identified as a luminous X-ray AGN. Located at $z = 0.6022$, this source has the EP 9 broadband energy covering the energy range of 0.5–4.5 keV in the observed frame with $L_{0.8-7.2keV} = 9.26 \times 10^{42}$ erg s$^{-1}$, which is a lower limit for the $L_{0.5-8keV}$. The SFR from the far-UV (FUV) flux is $33.27 M_\odot$ yr$^{-1}$ and $L_{0.5-8keV}/SFR > 41.44$. Comparing this source with the Brorby et al. (2016) $L_{0.5-8keV}/SFR$ level of 39.85 with a scatter of 0.25 dex, which is already higher than the Mineo et al. (2012) relation of 39.59, our source is much higher.

We summarize the sources that have been identified as AGNs with over four of the methods mentioned above in Table 1. A total of 12 galaxies have been identified with three methods (8 identified by [Mg II] $\lambda$2008 emission lines, MIR color, and MIR variability, and 4 galaxies identified with BPT diagram, MIR color, and MIR variability). For the galaxies that are identified with one or two methods, there are 11 spectra (9 galaxies) that have been identified with the existence of [Mg II] $\lambda$2800 lines. We consider all the sources that are identified as AGNs with more than three methods or have been identified with the existence of [Mg II] $\lambda$2800 lines as AGN candidates and exclude these 25 spectra (23 galaxies) from the SFR discussions in the following section.

4. Physical Properties

4.1. Color Excess from the Balmer Decrement

To ensure the quality of the flux ratio measurement, we select the sources at $0.173 \leq z \leq 0.385$, where both the H$\alpha$ and the H$\beta$ emission lines are located within the red portion of the spectrograph. To ensure the validity of the detection, we apply an S/N cut (median value of the flux over the flux error of the H$\beta$ region) over 3 to select the sources. A total of 62 spectra satisfy this criterion, and their flux ratios are displayed in Figure 12. We measure the color excess of the galaxies from the flux ratio of H$\alpha$ to H$\beta$ assuming the case B recombination with the intrinsic line ratio of 2.86. The color excess from the flux ratio is calculated with

$$E(B - V)_{gas} = \frac{\log_{10}(f_{H\alpha}/f_{H\beta})/2.86}{0.4 \times (k(H\beta) - k(H\alpha))},$$

where $k(H\alpha) = 3.33$ and $k(H\beta) = 4.6$ as in Jiang et al. (2019). We assume that the nebular gas in these galaxies emits at $T = 10^4$ K and $n_e = 10^4$ cm$^{-3}$. We show the distribution of the color excess in Figure 13. As is discussed in Section 4.1 in Cardamone et al. (2009), some of the sources have a flux ratio below 2.86; we attribute this to the uncertainties in flux calibration and low extinction and manually set their color excess to be 0.

4.2. Mass and SFR

The main goal for this subsection is to investigate the mass–metallicity relation and compare the SFR derived from the H$\alpha$ emission line and the FUV. We use 262 spectra (252 galaxies) with valid Starlight spectral fitting results for deriving the mass–metallicity relation, among which 4 are AGN candidates, 67 spectra (64 galaxies) are Blueberry galaxies, 123 spectra (116 galaxies) are Purple Grape galaxies (0.05 $\leq z < 0.112$), and 68 spectra (68 galaxies) are Green Pea galaxies. We use 24 samples with positive color excess from the previous subsection and FUV detection from GALEX to derive the extinction-corrected SFR.

To derive the physical properties for these sources (Cid Fernandes et al. 2005; Mateus et al. 2006), we fit the recalibrated LAMOST spectra with the emission lines masked. The fitting procedure is based on a base set of 45 spectra of three metallicities and 15 ages, with the stellar population model based on the BC03 high-resolution spectra and the Calzetti et al. (2000) extinction curve. The initial value of the velocity shift is set to be $v_0 = 0.0$ km s$^{-1}$, and the velocity dispersion is of $v_{disp} = 150.0$ km s$^{-1}$. We only consider the Starlight fitting of the LAMOST spectra a valid fitting result,
where \( \chi^2_{NLamost} < 0.2 \), adev < 30, and S/N (6000 Å < \( \lambda_{\text{rest-frame}} < 6500 \) Å) > 2.0, and have obtained 262 sources. Figure 14 displays an example of the Starlight SED fitting result.

Figure 15 demonstrates the relation of the stellar mass of the galaxies derived from Starlight with SFR (measured from the \( H_\alpha \) luminosity without extinction correction; Dopita & Ryder 1994; Kennicutt 1998; Panuzzo et al. 2003; Dopita 2005; Kennicutt & Evans 2012). We separate the strong [O III] \( \lambda 4363 \) emission-line compact galaxies into Blueberry galaxies, Green Pea galaxies, and Purple Grape galaxies and compare the mass/SFR relation for these three sets of samples with the star-forming main sequence (SFMS). Within each redshift bin, we separate the sources into several mass bins and calculate the median value and the standard deviation of the mass and metallicity for the sources within this mass and redshift bin. The main-sequence SFR – \( M_\star \) relation from Speagle et al. (2014) is as follows:

\[
\log \text{SFR}(M_\star, t) = (0.84 \pm 0.02 - 0.026 \pm 0.003 \times t) \times \log M_\star - (6.51 \pm 0.24 - 0.11 \pm 0.03 \times t),
\]

where \( t \) is the age of the universe in Gyr. Similar to the discussions in Section 4.4 in Cardamone et al. (2009), our strong [O III] \( \lambda 5007 \) samples have much higher sSFR at this redshift range, where typically for the galaxies at \( z \sim 0.2 \) the sSFR is around \( 10^{-9} \) yr\(^{-1} \) (Brinchmann et al. 2004; Bauer et al. 2005).

With the \( H_\alpha \) emission-line measurement result and the FUV flux, we intend to use this sample to check whether the SFRs derived from these two indicators agree. The calculation of the SFR from the recombination emission lines (24 spectra that do not have \( H_\alpha \) spectral coverage are dropped out) and FUV flux follow the relation defined in Hao et al. (2011), Murphy et al. (2011), and Kennicutt & Evans (2012), where the corresponding parameters are in Table 2:

\[
\log M_\star(M_\bigodot, \text{yr}^{-1}) = \log L_\bigodot - \log C_\bigodot.
\]

As discussed in Section 4.2, we convert from \( E(B-V) \) to \( A_V \) using \( A_V = R_V \times E(B-V) \) for the emission lines for the 62 high \( H/\beta \) S/N galaxies, under the assumption of the Calzetti et al. (2000) extinction curve and \( R_V = 3.1 \). Within these 62 samples, 36 samples are with positive color excess. A total of 24 out of the 36 samples have GALEX FUV detections. Figure 16 illustrates the comparison of SFR without extinction correction (left panel) and after extinction correction (right panel) with these 24 samples.

### 4.3. Gas-phase Metallicity

In this subsection, we first show the distribution of metallicity measured from the N2-based method with 1337 spectra, where the [N II] \( \lambda 6585 \) emission line is covered within the spectral range. Then, we demonstrate the mass–metallicity relation of 252 sources that have Starlight-derived mass and the N2-based metallicity. Furthermore, we discuss the gas-phase metallicity calculated from the empirical N2-based method and the direct \( T_e \)-based method with 21 galaxies whose specific flux of the [O III] \( \lambda 4363 \) emission lines is above \( 8.85 \times 10^{38} \) erg\(^{-1} \) s\(^{-1} \) Å, where the detection of [O III] \( \lambda 4364 \) is rare (Gao et al. 2017).

We use the N2-based method (van Zee et al. 1998; Pettini & Pagel 2004; Marino et al. 2013) to empirically calculate the gas-phase metallicity, because the two lines used in this method are close and thus less prone to the flux calibration error. The definition for this index is

\[
N_2 = \log_{10} ([\text{N II}] \lambda 6585/\text{H}\alpha),
\]

and this index can be converted to gas-phase metallicity with the following relation:

\[
12 + \log(O/H) = 8.90 + 0.57 \times N_2.
\]

Figure 17 displays the distribution of the gas-phase metallicity derived from the N2-based method. The gas-phase metallicity is higher than that of the pure Blueberry galaxies (Yang et al. 2017), similar to that of Cardamone et al. (2009).

We show the distribution of the mass–metallicity relation in Figure 18. We separate 248 galaxies (258 spectra), which consist of 64 Blueberry galaxies (67 spectra), 68 Green Pea galaxies (68 spectra), and 116 Purple Grape galaxies (123 spectra), into seven mass bins and calculate the median value and the standard deviation of the mass and metallicity for the sources within this mass bin.

We demonstrate that the metallicity increases with the mass by comparing the points with the smallest mass bin and the largest mass bin excluded, where there are fewer data points in these two mass bins. All of these strong [O III] \( \lambda 4363 \) emission-line compact galaxies are below the Tremonti et al. (2004) mass–metallicity relation.

Besides the empirical method to calculate the gas-phase metallicity, there is also the direct \( T_e \)-based method. However, the auroral [O III] \( \lambda 4363 \) line is not easy to detect. In this work, 21 galaxies whose specific flux of the [O III] \( \lambda 4363 \) emission lines are above \( 8.85 \times 10^{38} \) erg\(^{-1} \) s\(^{-1} \) Å are used to test the direct \( T_e \)-based method to calculate the electron temperature and the metallicity referring to Izotov et al. (2006, Section 3.1) and Jiang et al. (2019). The purpose of this discussion is to show that we have detected the auroral [O III] \( \lambda 4363 \) emission line, which increases the samples of detections in LAMOST (Gao et al. 2017), and to demonstrate that the current N2-based empirical metallicity relation and direct \( T_e \)-based metallicity relation in Jiang et al. (2019) are in agreement within the 1σ level.

This approach assumes two electron temperatures for \( O^+ \) and \( O^{++} \) in a two-zone photoionization model. We calculate the \( O^{++} \) electron temperature with the following equation (as

**Table 1**

| LAMOST Designation | [Mg II] \( \lambda 2008 \) | BPT | CIGALE AGN Fraction | MIR Color Criteria | MIR Variability |
|--------------------|---------------------|-----|----------------------|-------------------|-----------------|
| J234141.49+140028.1 | 1 time SF 0.20 | in | \( N_{ep} = 17, \sigma_1 > 0, \sigma_2 > 0, r_{12} = 0.64 \) |
| J021459.09-014459.2 | 1 time SF 0.21 | in | \( N_{ep} = 17, \sigma_1 > 0, \sigma_2 > 0, r_{12} = 0.61 \) |
in Izotov et al. (2006), Equations (1) and (2):

\[ t = \frac{1.432}{\log[(\lambda 4959 + \lambda 5007)/\lambda 4363]} - \log C_T, \]

where \( t = 10^{-4}T_e([O III]) \) and

\[ C_T = (8.44 - 1.09t + 0.5t^2 - 0.08t^3) \frac{1 + 0.0004x}{1 + 0.044x}, \]

where \( x = 10^{-4}N_e^{-0.5} \). We estimate the electron temperature of \([O II]\) for the low-metallicity situation in Izotov et al. (2006, Section 3.1, Equation (14)) as follows:

\[ T_e([O III]) = -0.577 + T_e([O III]) \times (3.065 - 0.498T_e([O III])). \]

This is also the method that Jiang et al. (2019) use for calculation, and they state that their measurement of the oxygen abundance depends little on the relation between \( T_2 \) and \( T_3 \). Similarly, we use the term \( T_2 = 10^{-4}T_e([O II]) \) and \( T_3 = 10^{-4}T_e([O III]) \) for clarity. Our spectra do not cover the \([S II]\) \( \lambda 6717 \) or \([S II]\) \( \lambda 6731 \) emission lines. The derived \( N_e \) is always smaller than \( 10^3 \) cm\(^{-3}\), and as in Jiang et al. (2019), \( N_e = 10, 100, \) or \( 10^3 \) cm\(^{-3}\); the results do not vary much. We use \( x = 0 \) for our calculations.

Table 2

| Band | \( L_x \) Units | \( \log C_x \) |
|------|----------------|----------------|
| FUV  | ergs s\(^{-1}\) (\( \nu L_\nu \)) | 43.35 |
| H\(_\alpha\) | ergs s\(^{-1}\) | 41.27 |

With Equations (3) and (5) in Izotov et al. (2006), we calculate the ionic abundances as follows:

\[ 12 + \log \frac{O^+}{H^+} = \log \frac{\lambda 43727}{H_\beta} + 5.961 + \frac{1.766}{T_1} - 0.40\log T_2 - 0.034T_2 + \log(1 + 1.35x), \]

and

\[ 12 + \log \frac{O^{2+}}{H^+} = \log \frac{\lambda 4959 + \lambda 5007}{H_\beta} + 6.200 + \frac{1.251}{T_1} - 0.55\log T_1 - 0.014T_3, \]
Because the majority of the ions of oxygen are \( \text{O}^+ \) and \( \text{O}_2^+ \), we use \( \frac{\text{O}}{\text{H}} = \frac{\text{O}^+}{\text{H}} + \frac{\text{O}_2^+}{\text{H}} \) to determine oxygen abundance.

Figure 19 demonstrates the comparison of the gas-phase metallicity estimated from the N2-based method and the direct \( T_e \)-based method. We calculate the error bars in Figure 19 by the Monte Carlo method. For the error bars in the N2-based metallicity, we generate 100 realizations for each source within the wavelength range of 6500–6640 Å, with the flux error from the rescaled LAMOST spectra, and measure the \([\text{N}]_\text{II} \) λ6585 line and the \( \text{H}_\alpha \) line flux for each realization. We calculate the N2-based metallicity for each realization and use the standard deviation from 100 realizations as the error bar for that source.

For the error bars in the direct \( T_e \)-based method, similarly, we generate 100 realizations for each source, but in a tighter wavelength range 4320–4383 Å, and measure the \([\text{O}]_\text{III} \) λ4363 line only for each realization. We note that 15 out of 21 sources are located on the black line within the 1σ region. This verifies that the current N2-based empirical metallicity relation and the direct \( T_e \)-based metallicity relation in Jiang et al. (2019) are in agreement within the 1σ level.

Because the majority of the ions of oxygen are \( \text{O}^+ \) and \( \text{O}_2^+ \), we use \( \frac{\text{O}}{\text{H}} = \frac{\text{O}^+}{\text{H}} + \frac{\text{O}_2^+}{\text{H}} \) to determine oxygen abundance.

Figure 19 demonstrates the comparison of the gas-phase metallicity estimated from the N2-based method and the direct \( T_e \)-based method. We calculate the error bars in Figure 19 by the Monte Carlo method. For the error bars in the N2-based metallicity, we generate 100 realizations for each source within the wavelength range of 6500–6640 Å, with the flux error from the rescaled LAMOST spectra, and measure the \([\text{N}]_\text{II} \) λ6585 line and the \( \text{H}_\alpha \) line flux for each realization. We calculate the N2-based metallicity for each realization and use the standard deviation from 100 realizations as the error bar for that source.

For the error bars in the direct \( T_e \)-based method, similarly, we generate 100 realizations for each source, but in a tighter wavelength range 4320–4383 Å, and measure the \([\text{O}]_\text{III} \) λ4363 line only for each realization. We note that 15 out of 21 sources are located on the black line within the 1σ region. This verifies that the current N2-based empirical metallicity relation and the direct \( T_e \)-based metallicity relation in Jiang et al. (2019) are in agreement within the 1σ level.

4.4. Environment

We identify a parent sample of comparison star-forming galaxies from the "emissionLinesPort" catalog (Thomas et al. 2013), whose spectroscopic classification is "Galaxy," whose BPT classification is "star-forming," and which covers the same redshift range as our Blueberry galaxies (\( z \leq 0.05 \)), Purple Grape galaxies (0.05 < \( z < 0.112 \) and 0.36 < \( z < 0.59 \)), and Green Peas (0.112 ≤ \( z < 0.36 \)). We ensure that the parent comparison sample covers the same mag-z space as our selected samples in the \( r \) band as displayed in the following plot. The Neighbors table from SDSS16 provides the angular separation to the nearby objects within 0.5. We only consider the primary detections as the valid neighbors. We calculate the projected linear separation between our source and all the valid neighbors by multiplying the angular separation from the Neighbors table by the luminosity distance at the redshift of our
strong [O III] λ5007 emission-line compact dwarf galaxies. We consider the smallest linear separation as the distance to the nearest neighbor. We display the distance for our samples to the nearest neighbor in different redshift bins in Figure 20.

We also perform the Kolmogorov–Smirnov test for these four sets of comparisons, and the p-value between the distance to the nearest neighbor of the parent sample and our selected sample in four redshift bins is all smaller than 0.05. Therefore, it is apparent that the strong [O III] λ5007 samples are farther away from the nearest neighbor compared with typical star-forming galaxies.

5. Result

We have selected 1547 unique strong [O III] λ5007 emission-line compact galaxies with 1694 spectra from LAMOST DR9. A total of 1342 galaxies are observed spectroscopically for the first time. Our samples enlarge the current samples of LCGs, Green Pea galaxies, and Blueberry galaxies. The following LAMOST extragalactic survey will keep providing a large amount of spectra for these targets and provide duplicate detections for the same source, helpful for observing the line variability.

We have confirmed 23 AGN candidates (25 spectra) from these strong [O III] λ5007 emission-line compact galaxies by LAMOST optical spectra, BPT diagram, CIGALE SED fitting, MIR color, and MIR variability.

Our samples show that strong [O III] λ5007 emission-line galaxies have higher SFR compared with the main-sequence SFR. The SFR from the FUV luminosity with the Hα luminosity agrees with the coefficients after extinction correction from Kennicutt & Evans (2012). From the mass–SFR plot, we show that with the increase of redshift, the SFR is increasing. For the sources within the same redshift bin, the SFR increases with mass with a similar slope to the SFMS.

For the gas-phase metallicity for these samples, these samples have a median metallicity of $12 + \log(O/H)$ of 8.10. In the mass–metallicity plot, all the sources are below the Tremonti et al. (2004) mass–metallicity relation. The metallicity increases with mass. A total of 21 galaxies are found with [O III] λ4363 emission lines. The direct $T_e$-based metallicity measurement result is in agreement with the N2-based metallicity result.

Lastly, Strong [O III] λ5007 emission-line galaxies are in a less dense environment.

This work is supported by National Science Foundation of China (No. U1931209) and National Key R&D Program of China (No. 2019YFA0405502). Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. S.L. thanks the useful discussion with Sophia Y. Dai, Qing Liu, Er-Lin Qiao, and He-Yang Liu. J.X.W. acknowledges support from National Science Foundation of China (No. 1890693 & 12033006). This research uses Astropy (Astropy Collaboration et al. 2013, 2018), TOPCAT (Taylor 2005), Scipy (Virtanen et al. 2020), Numpy (Harris et al. 2020), and Matplotlib (Hunter 2007).

ORCID iDs

Siqi Liu @ https://orcid.org/0000-0002-5345-4175

A-Li Luo @ https://orcid.org/0000-0001-7865-2648

Huan Yang @ https://orcid.org/0000-0003-2260-7420

Shi-Yin Shen @ https://orcid.org/0000-0002-3073-5871

Jun-Xian Wang @ https://orcid.org/0000-0002-4419-6434

Hao-Tong Zhang @ https://orcid.org/0000-0002-6617-5300

Zhenya Zheng @ https://orcid.org/0000-0002-9634-2923

Yi-Han Song @ https://orcid.org/0000-0001-7255-5003

Xiao Kong @ https://orcid.org/0000-0001-8011-8401

Jian-Ling Wang @ https://orcid.org/0000-0001-8274-3980

References

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3

Allende Prieto, C., Lambert, D. L., & Asplund, M. 2001, ApJL, 556, L63

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Bauer, A. E., Drory, N., Hill, G. J., & Feulner, G. 2005, ApJL, 621, L89

Bianchi, L., Shiao, B., & Thilker, D. 2017, ApJS, 230, 24

---

12 https://doi.org/10.12149/101085

13 http://www.astropy.org
