



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

Galaxy Population Identification With a Phylogenetic Approach

Por: MONSERRAT LORETO MARTINEZ MARIN

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la
Univerisdad de Concepción para optar al grado académico de Magíster en
Astronomía.

Profesor Guía: Dr. Ricardo Demarco López

**Comisión: Dr. Ricardo Demarco López, Dr. Guillermo Cabrera
Vives & Dr. Neil Nagar**

Enero 2020

Concepción, Chile



© 2019, Monserrat Loreto Martínez Marín

Ninguna parte de esta tesis puede reproducirse o transmitirse bajo ninguna forma o por ningún medio o procedimiento, sin permiso por escrito del autor.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento

Acknowledgement

I would first like to thank my thesis advisor Ricardo Demarco, for his advice, encouragement and trust placed in me to do this job. As well as my co-advisor Guillermo Cabrera for his extraordinary support and always steered me in the right direction whenever I needed it. I would also like to thank the experts who were involved in this work, Dr. Pierluigi Cerulo, Dr. Rodrigo Herrera, Dr. Nathan Leigh and Dr. Inger Jørgensen. Without their passionate participation, valuable comments and advises, this thesis could not have been successfully conducted.

I wish to acknowledge the support and great love of my family, my mother, Loreto; For his unconditional love even when I'm the biggest scare of his life, my father, Osvaldo; For instilling curiosity in science from an early age. My brother and sister, Nelson and Micaela; For their care and love, my friends Mariela, Catalina, Paulo, Carolina, Daniela and Ignacio; For being my second family. And finally my significant other, Paula; Because without her love none of this would be possible.

Summary

Galaxy evolution is one of the main open questions in modern astrophysics. Theories of evolution arise from the observed morphological differences of galaxies. We propose a new approach interpreting the results without previously biasing the sample by a specific characteristic of galaxies like mass, morphology, sSFR or metallicity. For this we make a phylogenetic analysis at different environments like Coma cluster to 475 galaxies with redshifts $0,015 < z < 0,033$ and 438 galaxies of the coeval field with redshifts $0,035 < z < 0,054$. The phylogenetic study retrieve possible evolutionary paths for galaxies based on the information of stellar population of galaxies by a set of 30 absorption indices from the Sloan Digital Sky Survey (SDSS). The branches of the tree are interpreted as galaxy populations and the length between the nodes (NodeLength) as the internal chemical variation of the galaxy population. The NodeLength works as chemical index, thus connecting galaxy stellar populations with other structural and local environmental properties of galaxies. We found a contrast of how galaxies are chemically related at different environments, the Coma cluster is a complex systems with multiple populations of galaxies, while the field is a more homogeneous population, presenting one main galaxy population. The Coma cluster present three main galaxy populations and several minor structures in the Red Sequence of the Color-Magnitude Diagram. The main populations we find in Coma are 3: (i) One mostly of S0 morphology, which internal hierarchical structure shows correlations with stellar age, metallicity, and sSFR, N_aD index. This tells us that galaxies of this population has suffered a recent period of star formation;(ii) another one with mostly elliptical morphology, do not show internal correlations and has old quenched galaxies with high metallicity and stellar mass; and (iii) one that has the youngest galaxies with spiral and lenticular morphology, which belong to the blue cloud of the Cluster. This population do not show internal correlations, but their galaxies with S0 morphology, show a chemical correlation with $Fe4383$ not present in other galaxy populations with the same morphology. On the other hand the main population present in the field, has an internal hierarchical structure that show a weak correlation with sSFR, but not dependence with any specific absorption index.

Keywords – Galaxy evolution, Galaxy Clusters, Abundances, Spectral Indices

Abstract

We propose a phylogenetic approach as a novel and robust tool capable of detecting galaxy populations (GPs) based on their chemical composition. This method clusters galaxies into the hierarchical structure of a phylogenetic tree that connects galaxies through features we call "nodes". The branches of the tree are interpreted as different GPs and the length between nodes (NodeLength) as the internal chemical variation along a branch. The NodeLength works as a chemical index, thus relating galaxy stellar populations with other structural and local environmental properties of galaxies. We apply the phylogenetic approach using 30 abundance indices from the Sloan Digital Sky Survey to 475 galaxies in the Coma cluster ($0,015 < z < 0,033$) and 438 galaxies in the field ($0,035 < z < 0,054$). We find that Coma is a complex system with multiple populations. It has three main GPs that can readily be identified in color-magnitude space, and several minor structures in its red sequence. On the other hand, the field is more homogeneous, presenting one main GP. This phylogenetic analysis of cluster and field galaxies, shows a contrast in terms of how chemically related galaxies are within different environments. We also find that, regardless of their morphology, galaxies can have a similar chemical composition. Therefore, this new approach to study galaxy properties and their evolution makes it possible to perform robust analyses and interpretations without introducing biases by first selecting samples based on a specific characteristic like stellar mass, morphology, sSFR or metallicity, as done with more traditional techniques.

Keywords – Galaxy evolution, Galaxy Clusters, Abundances, Spectral Indices

Table of Contents

Acknowledgement	I
Summary	II
Abstract	III
1. Introduction	1
1.1. Cladistics for astrophysics	1
1.2. Phylogenetics based on DM	1
1.3. Galaxy evolution	2
2. Theoretical Framework	5
2.1. Research Problem	5
2.2. Research Question	5
2.3. Problem Statement	5
2.4. Hypotesis	6
3. Data	7
3.1. Phylogenetic study for Globular Cluster	7
3.2. Phylogenetic study for Galaxies	7
3.2.1. Phylogenetic Study	7
3.2.2. Comparison data	7
4. Methodology	9
4.1. Phylogenetic Tree	9
5. Analysis	21
5.1. Proof of concept	21
5.1.1. Globular Cluster NGC2808	21
5.2. Phylogenetic study in Galaxies	25
5.2.1. The Coma Cluster	25
5.2.2. The Field	36
6. Discussion	47
6.1. Comparison with other methods	47
6.2. Minor Structures	49
7. Conclusion	52
References	54

List of Tables

5.1.	Galaxy distribution in the Phylogenetic tree of the Coma Cluster	25
5.2.	Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population	34
5.3.	Remaining galaxy populations for Recalculated phylogenetic trees without a specific set of line indices	36
5.4.	Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population	41
5.5.	Pearson and Spearman coefficients between NodeLength and specific abundance indices for each galaxy population in the field	46
6.1.	Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population	50
6.2.	Pearson and Spearman coefficients between NodeLength and specific abundance indices for each galaxy population in the field	50



List of Figures

4.1.	Star representation distribution of galaxies of our example, the branches in black represent the unknown length.	10
4.2.	Star representation of the distribution of our galaxy sample in our first cycle, the branches in black represent the unknown length and red branches represent Know distances.	12
4.3.	Star representation of the distribution of our galaxy sample in the second cycle. The branches in black represent the unknown length and red branches represent Know distances. In this cycle we have two nodes represented by letter U, that joins different galaxies.	15
4.4.	Star representation of the distribution of our galaxy sample in our third cycle. The branches in black represent the unknown length and red branches represent Know distances. We have a tree with three nodes represented by letter U.	17
4.5.	Star representation of the distribution of our galaxy sample in our four cycle, the branches in black represent the unknown length and red branches represent Know distances.	18
4.6.	Final Star representation of the distribution of our galaxy sample, the branches in black represent the unknown length and red branches represent Know distances. The U letters represent the nodes between galaxies. . . .	19
4.7.	Final Phylogenetic tree, each galaxy is represented by a letter at the end of the branches. Each branch is joined to an specific node U. The numbers represent the length of the branches.	19
5.1.	Consensus phylogenetic tree of the globular cluster NGC2808. The red shaded area highlight the main branch of the phylogenetic tree. The numbers correspond to the ID of each star in the globular cluster.	22
5.2.	Mg and Na anticorrelation from figure 8 C15. Different colours indicate different groups, blue (P1), green (P2), red (I1), orange (I2), and black (E)	23
5.3.	Na-Mg anticorrelation of the globular cluster NGC2808, with the five groups define by C15 differentiated by color. Blue (P1), green (P2), red (I1), orange (I2), and black (E). The magenta hollow squares represent the stars in our main branch of the phylogenetic tree.	24
5.4.	Color-Magnitude diagram of the Coma Cluster with photometric filters Johnson B and Sloan r. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate galaxy morphology where red corresponds to elliptical galaxies, blue to spiral galaxies, and yellow to S0 galaxies.	26
5.5.	Branch B1 of the phylogenetic tree of the Coma Cluster that represents our first galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the length of their separation.	27
5.6.	Branch B2 of the phylogenetic tree of the Coma Cluster that represents our second galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the chemical length of their separation.	28

- 5.7. Branch B3 of the phylogenetic tree of the Coma Cluster that represents our third galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the length of their separation. 29
- 5.8. Projected sky distribution of galaxies in the Coma Cluster belonging to different branches of the phylogenetic tree. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate the morphology of the galaxy where red corresponds to elliptical galaxies, black to spiral galaxies, and magenta to S0 galaxies. The color scale shows the number count of galaxies by degree squared in the sky for Coma cluster. White star points the galaxy NGC4839. 31
- 5.9. The histograms A, B, C and D show the distribution of galaxies from the three different branches B1, B2 and B3 with respect to stellar mass, metallicity, age and sSFR. The yellow, red and blue colors represent galaxies from branches B1, B2 and B3, respectively. 32
- 5.10. Panels A, B, C and D, show a comparison of our chemical index (length between nodes) with stellar mass, metallicity, age of the stellar population, and sSFR, respectively. Triangles, circles and squares represent B1, B2 and B3 GP. The red, yellow and blue colors represent elliptical, lenticular and spiral morphology, respectively. 33
- 5.11. Color-magnitude diagram of our Coma cluster galaxy sample with photometric filters Johnson B and Sloan r. Panels A, B, C, D, and E show a recalculated phylogenetic tree without, negative response to α/Fe enhancement, insensitive response to α/Fe enhancement, positive response to α/Fe enhancement, absorption Balmer lines, and absorption line indices [OIII], respectively. The cyan diamonds represent the galaxies that appear in the branches of each recalculated tree. 37
- 5.12. 38
- 5.13. Continue image 39
- 5.14. Comparison of absorption lines with length between nodes of our GP. Triangles, circles and squares represent the B1, B2, and B3 GP. The red, yellow and blue colors represent elliptical, lenticular and spiral morphology, respectively. Red, yellow and blue lines correspond to linear fits to each galaxy population. The Spearman correlation (SC) index is calculated for each galaxy population. 40
- 5.15. Spatial distribution of our field galaxy sample in the sky. Grey circles represent all galaxies in our sample, whereas triangles represent our main galaxy population. Red, yellow, and blue colors indicates, correspondingly, elliptical, S0 and spiral galaxies. 42
- 5.16. Galaxy population obtained in the field, plotted as a single tree. Each galaxy has its ID number from the SDSS DR7, and its image is from the SkyServer DR15. 43
- 5.17. Panels A, B, C, and D show a comparison of our chemical index (length between nodes) with stellar mass, metallicity, age of the stellar population and sSFR, respectively. Red, yellow and blue colors represent the elliptical, S0 or spiral morphology, respectively, of our galaxy population in the field. 44

- 5.18. Color-magnitude diagram of our field galaxy sample, where panels A, B, C, D, and E show a recalculated phylogenetic tree excluding absorption Balmer lines, lines insensitive to α/Fe , lines with negative response to α/Fe enhancement, lines with positive response to α/Fe enhancement, and Emission Line $[OIII]\alpha5007$, with $[OIII_1]$ and $[OIII_2]$ respectively. The cyan diamonds represent the galaxies that appear in branches of each recalculated tree. 45
- 6.1. Principal component decomposition of the 30 abundance indices for our galaxy samples of the Coma cluster and the field. Figure A present the principal components 1 and 2, and B present the principal components 1 and 4 for the Coma cluster. Figure C present the principal components 1 and 2, and D present the principal components 1 and 4 for the field. Red, blue and yellow dots, indicates the elliptical, spiral and lenticular morphology of the galaxies. Magenta and cyan diamonds indicates the centroids for the PCA decomposition 48
- 6.2. Minor structures. Image A show a Color-Magnitude diagram of the Coma Cluster with photometric filters Johnson B and Sloan r, image B show a PCA decomposition of the coma cluster and image C show a PCA decomposition of the field. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate galaxy morphology where red corresponds to elliptical galaxies, blue to spiral galaxies, and yellow to S0 galaxies. Black and grey dots indicates the structures with 1, 2 and 3 nodes and single branch respectively. 51

1. Introduction

In the last years, phyloinformatic studies have been used in astronomy to help understand different processes that could involve evolution. These studies were based either on discrete character states (DC) or distance matrices of pairwise dissimilarities (Dm). Their main objective is to study the evolutionary history and relationships among individual or groups of astronomical objects.

1.1. Cladistics for astrophysics

Astrocladistics is a Maximum parsimony method, that can be utilize as a tool to find hierarchical classification of galaxies. This classification is based on choosing specific characters to compare states of evolution. The characters can be either qualitative (e.g. morphology) or quantitative (e.g. luminosity, stellar mass, chemical composition). In order to compare this characters two states have to be defined, one have to be ancestral and the other one derived. This hierarchical classification is represented in a phylogenetic tree, called a cladogram, in which the distance between two branches reflects the level of diversification that occurred during evolution. Astrocladistics essentially attempts to represent the pattern of morphological similarity. Some examples applied to astronomy of astrocladistics are the morphological relation in dwarf galaxies of the Local Group (Fraix-Burnet et al. 2006), globular cluster classification (Fraix-Burnet et al. 2009) and stellar populations (SP) in ω Centauri (Fraix-Burnet and Davoust 2015).

1.2. Phylogenetics based on DM

Other method for inferring phylogenetic trees is based on a distance-matrix. Here we calculate some measure of dissimilarity that for our case study are absorption lines indices of each pair of galaxies, to produce a pairwise distance matrix, and then infer with the neighbour joining algorithm the phylogenetic relationships of the galaxies from that matrix. The neighbour joining algorithm constructs a tree by sequentially finding pairs of neighbours, which in our study are pairs of galaxies connected by a single interior

node. The main difference of NJ from other clustering algorithms is that it minimizes the length of all internal branches and thus the length of the entire tree. This way we can represent the branching pattern of evolution. This method was employed as an attempt to reconstruct the chemical history of stars in the solar neighbourhood (Jofré et al. 2017) and to retrieve stars from an open cluster (Blanco-Cuaresma and Fraix-Burnet 2018).

1.3. Galaxy evolution

The utility of phylogenetics studies in galaxies, came from the necessity of find a reliable method to solve galaxy evolution which remains as an open problem in modern astrophysics. Galaxies are complex systems made up of stars, gas, dust and dark matter. Galaxy evolution is the result of several complex physical processes that involve the various galaxy components in different ways through cosmic epochs. These mechanisms bring galaxies to transition from one physical state to another and to change their observed properties.

Morphology is the most immediate way to classify galaxies since it is based on a visual assessment. Following Hubble (Hubble, 1926), galaxies are classified as ellipticals, lenticulars or S0s, and spirals. A class of irregular galaxies is added to take into account objects that cannot be classified neither as elliptical, S0 or spiral. Ellipticals and S0s are commonly said early-type galaxies (ETGs), while spirals and irregulars are commonly grouped together as late-type galaxies (LTGs). (although see Scarlata et al. 2007, Huertas-Company et al. 2008, Pérez-Carrasco et al. 2019 for examples of automated, machine-learning-based morphological classification). Morphology is related to the underlying structural distribution of the stellar populations and interstellar medium. ETGs tend to be red and with no or little star-formation, while LTGs tend to be blue and star-forming (see Baldry et al. 2004, Cassata et al. 2008, Taylor et al. 2015). Up to $z \sim 6$, the star formation in galaxies is observed to be related to their stellar mass: massive galaxies are less star-forming than their low-mass counterparts (Cowie et al., 1996). The stellar mass of galaxies is finally related to their metallicity: more massive galaxies are also metal-richer (Lequeux et al. 1979; Tremonti et al. 2004; Gallazzi et al. 2005)

Regardless of the environment where galaxies reside, they show a bimodal colour distribution. Three main populations can be defined from this color-magnitude distribution:

(1) a Red Sequence of quiescent galaxies; (2) a Green Valley as an "intermediate" population; and (3) a diffuse cloud of blue star-forming galaxies. Thus, in principle, the evolution of a galaxy population can be investigated by looking at the gradual build-up of the Red Sequence as a function of redshift (Tanaka et al. 2010; Gobat et al. 2011; Muzzin et al. 2012; Lidman et al. 2004; Snyder et al. 2012).

Nevertheless, we can obtain more detailed information by studying galaxy spectra. Spectra of galaxies have information about gas and stellar populations properties that preserved record of galaxy's formation and evolution. Emission lines are used to study the physical state of the ionized gas, so we can trace black hole accretion, derive gas kinematics and star formation activity (Kauffmann et al. 2003; Schawinski et al. 2007). Absorption lines provides information of stellar populations properties and can be used to derive ages, metallicities, star formation histories and element abundances. (Davies et al. 1993; Fisher et al. 1996; Jørgensen 1999; Kuntschner 2000; Thomas et al. 2010a). The most common sets of optical absorption line indices are the Lick index system (Burstein et al. 1984; Faber et al. 1985; Worthey 1994; Worthey and Ottaviani 1997), and the [OIII] indices (González 1993). This indices group different absorption lines, therefore they measure various chemical elements, and the derivation of the abundance for an individual element abundances is non-trivial.

Studies with absorption lines compare morphology with chemical abundances and other properties like velocity dispersion (Davies et al. 1987, Burstein et al. 1988), B-band absolute magnitudes (Faber 1973; Terlevich and Melnick 1981), or effective radius (Parikh et al. 2018).

In this work we propose a new perspective to study galaxy evolution, by analyzing the grouping of galaxies according to their chemical composition through a phylogenetic approach utilizing multiple absorption indices. For this, we analyzed the stellar population chemical information of galaxies in different environments. We considered 475 galaxies in the Coma cluster of galaxies with redshifts $0,015 < z < 0,033$, and 438 field galaxies with redshifts $0,035 < z < 0,054$. Utilizing a set of 30 lines indices from the value added catalogs of the Sloan Digital Sky Survey data release 14 (SDSS) (Abolfathi et al. 2018). We generated a phylogenetic tree based on a distance matrix from the pairwise differences of these indices with the Neighbor Joining algorithm (NJ; Saitou and Nei 1987; Studier

and Keppler 1988). NJ allows us to find hierarchical relationships that can be used to uncover possible evolutionary paths of the observed galaxies. We interpret the branches of the phylogenetic tree as different galaxy populations, and the length between nodes of this branches as a gradient of chemical difference in the stellar population content of these galaxies. In this way we analyze galaxy evolution avoiding linearity, to see then how properties like morphology, stellar mass (M_*), star formation rate (SFR), specific star formation rate (sSFR; SFR/M_*) or metallicity ($[Z/H]$) relate with galaxy populations in different environments.



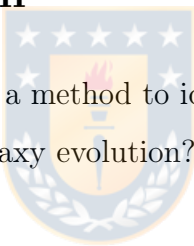
2. Theoretical Framework

2.1. Research Problem

In modern astronomy, physical relationships that link galaxy populations with galaxy evolution are not well defined yet with enough amount of detail. Since Hubble introduced his classification of galaxies based on morphology, astronomers have wondered whether this implies that galaxies evolve. Therefore, they have studied possible internal and external physical processes that could be responsible for this evolution. However, the selection criteria for galaxy populations are biased by morphology or other properties like (stellar or dynamical) mass, star formation rate, metallicity and age.

2.2. Research Question

Can we utilize a Phylogenetic tree as a method to identify galaxy populations and obtain a useful physical insight to study galaxy evolution?



2.3. Problem Statement

In the last years, phylogenetic trees have been used in astronomy to study processes that involve evolution. They are based on the chemical content of astronomical objects, and are able to disentangle them into different populations represented as the branches of the tree. The aim of this work is the utilization of this method on galaxies to identify different populations by using a set of absorption line indices. In order to have a physical insight about their chemical composition, we define an index given by the length between the so-called nodes of the branches that reflect a chemical gradient between galaxies within a population. This allows us to visualize which absorption lines are prominent, aiming at being able to better characterize and describe the overall star formation history of galaxies.

and then infer the physical process that explains the observed relationships.

2.4. Hypotesis

Phylogenetic trees can be used to identify galaxy populations and give us a physical insight about them to study galaxy evolution.



3. Data

3.1. Phylogenetic study for Globular Cluster

For the globular cluster NGC2808 we use the membership of stars from ?, obtaining 117 stars with 11 chemical abundances of $[O/Fe]_I$, $[Na/Fe]_I$, $[Mg/Fe]_I$, $[Si/Fe]_I$, $[Ca/Fe]_I$, $[Sc/Fe]_{II}$, $[Ti/Fe]_I$, $[Cr/Fe]_I$, $[Fe/H]_I$, $[Fe/H]_{II}$, $[Ni/Fe]_I$, with they corresponding errors.

3.2. Phylogenetic study for Galaxies

3.2.1. Phylogenetic Study

For the phylogentic study on the Coma cluster we use the membership of Beijersbergen et al. 2002, obtaining 475 galaxies with redshifts from 0,015 to 0,033. For the field we selected galaxies that do not belong to galaxy groups (Tempel et al. 2012) nor filaments (Tempel et al. 2014), thus obtaining a sample of 438 galaxies with redshifts from 0,035 to 0,054. For both samples we obtained the spectroscopic data from the seventh data release (DR7 Abazajian et al. 2009) of the Sloan Digital Sky Survey (SDSS; York et al. 2000). We used the spectral line strength measurement database provided by the OSSY group (Oh et al. 2011), considering the stellar absorption-line measurements for this galaxy sample.

3.2.2. Comparison data

The stellar mass is based on fits to the photometry following (Kauffmann et al. 2003), and (Salim et al. 2007) from the SDSS DR7. The *SFR* estimates are based on the technique discussed in Brinchmann et al. 2004, with a slight modification for non-star-forming galaxies where the likelihood distribution of the sSFR was constructed as a function of D4000 using the star-forming sample, also from the SDSS DR7. The ages and metallicities where calculated following Cardiel et al. 2003 with the Thomas et al. 2010b models. Finally, we utilized the visual morphological classification by TType from van Dokkum

2001 and two parameters of the morphological classification from Domínguez Sánchez et al. 2018, obtained with deep learning algorithms using convolutional neural networks (CNNs). The TType parameter separates ETGs from LTGs and gives the probability P_{S0} of being S0 versus a pure elliptical in order to have three galaxy classes: Elliptical, Spiral and S0. Elliptical galaxies were define such that have TType values from -3 to 0 and P_{S0} lower than $0,5$. S0 galaxies have TType values between -1 and 3 , and P_{S0} higher than $0,5$. Finally, Spiral galaxies have a TType value between 3 and 8 , independent of the P_{S0} value. For galaxies in the sample that do not have information on TType and P_{S0} , a visual morphological classification was assigned.




4. Methodology

4.1. Phylogenetic Tree

We perform a phylogenetic study based on a distance matrix with the NJ algorithm utilizing the open-source Scikit-bio python package. In our case study, the distance matrix elements are obtained by the summation of the euclidean distance for a given abundance index (X_k) between every pair of galaxies in our sample. Therefore, each element of our distance matrix is the total chemical distance $D_{i,j}$ from our set of abundance indices between our galaxies:

$$D_{i,j} = \sum_{k=1}^N |[X_k]_i - [X_k]_j|$$

For example if we have the following distance matrix:



$$\begin{bmatrix} & G_A & G_B & G_C & G_D & G_E & G_F \\ G_A & 0 & 5 & 4 & 7 & 6 & 8 \\ G_B & 5 & 0 & 7 & 10 & 9 & 11 \\ G_C & 4 & 7 & 0 & 7 & 6 & 8 \\ G_D & 7 & 10 & 7 & 0 & 5 & 9 \\ G_E & 6 & 9 & 6 & 5 & 0 & 8 \\ G_F & 8 & 11 & 8 & 9 & 8 & 0 \end{bmatrix}$$

The molecular clock hypothesis states that the rate of nucleotide substitution per generation is constant across lineages. If generation times were equal across lineages, samples obtained at the same calendar time would have experienced the same number of generations since their common ancestor

A clustering method would group galaxies A and C as the most similar galaxies, because their total chemical distance $D_{i,j}$ is the smallest. This would assume that the elements under study evolve in a clock-like behavior, therefore the samples obtained at the same

calendar time generation time would have experienced the same number of generations since their common ancestor, maintaining a constant rate of change per generation. Nevertheless the main characteristic of NJ is that it takes into account the development of evolution at different rates. For that, the NJ defines a new distance $M_{i,j}^*$ which subtracts the divergence (a summation over the total chemical distance $D_{i,j}$ of each possible pair of galaxies) from the original distance, and normalizes the result by the number of end nodes N :

We will solve step by step an example of six galaxies, utilizing a star representation as we can see in figure 4.7.

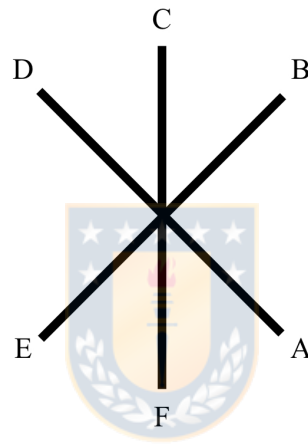


Figura 4.1: Star representation distribution of galaxies of our example, the branches in black represent the unknown length.

To know their rate corrected distances we, calculate their divergence with the following equation:

$$S_j = \frac{\sum_{i \leq j} D_{i,j}}{N - 2}$$

Now we will calculate the divergences for each galaxy.

$$\begin{aligned}
S_A &= \frac{5 + 4 + 7 + 6 + 8}{4 - 2} = 7,5 & S_B &= \frac{5+7+10+9+11}{4-2} = 10,5 \\
S_C &= \frac{4 + 7 + 7 + 6 + 8}{4 - 2} = 8 & S_D &= \frac{7+10+7+5+9}{4-2} = 9,5 \\
S_E &= \frac{6 + 9 + 6 + 5 + 8}{4 - 2} = 8,5 & S_F &= \frac{8+11+8+9+8}{4-2} = 11
\end{aligned}$$

With the divergences calculated, we will obtain the new rate corrected distances.

$$M_{i,j}^* = D_{i,j} - (S_i + S_j)$$

This rate corrected distance allow to group the the closest nodes, but at the same time the ones that are further of the rest of the galaxies.



$$M_{A,B}^* = D_{A,B} - (S_A + S_B) = 5 - 7,5 - 10,5 = -13$$

$$M_{A,D}^* = -10 \quad M_{A,E}^* = -10 \quad M_{A,F}^* = -10,5$$

$$M_{B,C}^* = -11,5 \quad M_{B,D}^* = -10 \quad M_{B,E}^* = -11$$

$$M_{B,F}^* = -10,5 \quad M_{C,D}^* = -10,5 \quad M_{C,E}^* = -10,5$$

$$M_{C,F}^* = -11 \quad M_{D,E}^* = -13 \quad M_{D,F}^* = -11,5$$

$$M_{E,F}^* = -11,5 \quad M_{A,C}^* = -11,5$$

With this new rate corrected distance, we create a new distance matrix.

$$\begin{bmatrix} & G_A & G_B & G_C & G_D & G_E & G_F \\ G_A & 0 & -13 & -11,5 & -10 & -10 & -10,5 \\ G_B & -13 & 0 & -11,5 & -10 & -11 & -10,5 \\ G_C & -11,5 & -11,5 & 0 & -10,5 & -10,5 & -11 \\ G_D & -10 & -10 & -10,5 & 0 & -13 & -11,5 \\ G_E & -10 & -11 & -10,5 & -13 & 0 & -11,5 \\ G_F & -10,5 & -10,5 & -11 & -11,5 & -11,5 & 0 \end{bmatrix}$$

Now we will create our first node U_1 with galaxies A and E for which $M_{i,j}^*$ is minimal, but we could also choose galaxies D and E. The decision of selecting one or other node will define the the zero point of our tree. We can calculate the length of the branches from each galaxy to this node.

$$S_{B,U_1} = \frac{D_1 + S_B - S_A}{N - 2} \cdot \frac{1}{2} = 1$$

$$S_{A,U_1} = \frac{D_1 + S_A - S_E}{N - 2} \cdot \frac{1}{2} = 4$$

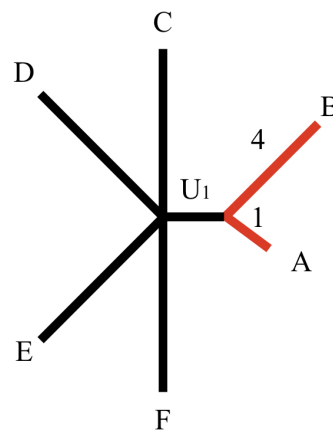


Figura 4.2: Star representation of the distribution of our galaxy sample in our first cycle, the branches in black represent the unknown length and red branches represent Know distances.

$N = 6$ Number of end-nodes.

Now we compute new distances from node U_1 to each other terminal node.

$$d_{C,U_1} = \frac{(d_{C,A} + d_{C,B} - d_{A,B})}{N - 2} = 3$$

$$d_{D,U_1} = \frac{(d_{D,A} + d_{D,B} - d_{A,B})}{N - 2} = 6$$

$$d_{E,U_1} = \frac{(d_{E,A} + d_{E,B} - d_{A,B})}{N - 2} = 5$$

$$d_{F,U_1} = \frac{(d_{F,A} + d_{F,B} - d_{A,B})}{N - 2} = 7$$

We create the new distance matrix.

$$\begin{bmatrix} & U_1 & G_C & G_D & G_E & G_F \\ U_1 & 0 & 3 & 6 & 5 & 7 \\ G_C & 3 & 0 & 7 & 6 & 8 \\ G_D & 6 & 7 & 0 & 5 & 9 \\ G_E & 5 & 6 & 5 & 0 & 8 \\ G_F & 7 & 8 & 9 & 8 & 0 \end{bmatrix}$$

With this new distance matrix, we start a new cycle to find chemical relationships between our galaxies. Therefore we compute the new divergences for the galaxies that we have left.

$N = 5$ Number of end-nodes.

$$S_U = \frac{3 + 6 + 5 + 7}{5 - 2} = 7 \quad S_C = \frac{3 + 7 + 6 + 8}{3} = 8$$

$$S_D = \frac{6 + 7 + 5 + 9}{3} = 9 \quad S_E = \frac{5 + 6 + 5 + 8}{3} = 8$$

$$S_F = \frac{7 + 8 + 9 + 8}{3} = 10,6$$

Now we compute the rate corrected distances

$$\begin{aligned}
 M_{U,C}^* &= -12 & M_{U,D}^* &= -10 & M_{U,E}^* &= -11 \\
 M_{U,F}^* &= -10,6 & M_{C,D}^* &= -10 & M_{C,E}^* &= -11 \\
 M_{C,F}^* &= -10,6 & M_{D,E}^* &= -12 & M_{D,F}^* &= -10,6 \\
 M_{F,E}^* &= -10,6
 \end{aligned}$$

With this new rate corrected distance, we create a new distance matrix.

$$\begin{bmatrix}
 & U_1 & G_C & G_D & G_E & G_F \\
 U_1 & 0 & -12 & -10 & -11 & -10,6 \\
 G_C & -12 & 0 & -10 & -11 & -10,6 \\
 G_D & -10 & -10 & 0 & -12 & -10,6 \\
 G_E & -11 & -11 & -12 & 0 & -10,6 \\
 G_F & -10,6 & -10,6 & -10,6 & -10,6 & 0
 \end{bmatrix}$$

This new rate corrected distance matrix, show us that the new node could be either galaxies D and E or Node U with galaxy C, we will choose galaxies D and E to have the node U_2 .

$$S_{D,U_2} = 3$$

$$S_{E,U_2} = 2$$

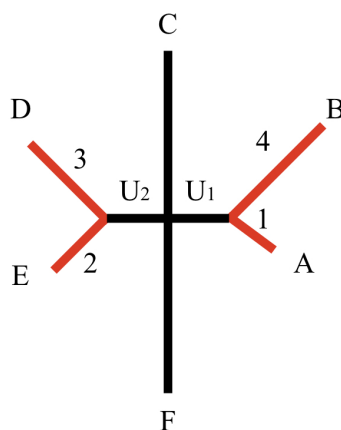


Figura 4.3: Star representation of the distribution of our galaxy sample in the second cycle. The branches in black represent the unknown length and red branches represent known distances. In this cycle we have two nodes represented by letter U, that joins different galaxies.

We calculate the new distances to each galaxy.

$N = 4$ Number of end-nodes.

$$d_{U_2, U_1} = \frac{(d_{D, U_1} + d_{D, U_2} - d_{D, E})}{N - 2} = 3$$

$$d_{C, U_1} = 3$$

$$d_{F, U_1} = 7$$

The distances between nodes like d_{U_2, U_1} will be used later as a chemical index that will show a chemical gradient useful to differentiate galaxies.

$$\begin{bmatrix} & U_1 & G_C & U_2 & G_F \\ U_1 & 0 & 3 & 3 & 7 \\ G_C & 3 & 0 & 4 & 8 \\ U_2 & 3 & 4 & 0 & 6 \\ G_F & 7 & 8 & 6 & 0 \end{bmatrix}$$

We start the third cycle of computation, we already have to nodes in our tree and we need to complete our tree calculating the new divergences for the galaxies that we have left.

$$S_{U_1} = \frac{3 + 3 + 7}{4 - 2} = 6,5 \quad S_C = \frac{3+4+8}{2} = 7,5$$

$$S_{U_2} = \frac{3 + 4 + 6}{2} = 6,5 \quad S_F = \frac{7+8+6}{2} = 10,5$$

Now we compute the rate corrected distances

$$M_{U_1,C}^* = -11 \quad M_{U_1,U_2}^* = -10 \quad M_{U_1,F}^* = -10$$

$$M_{C,U_2}^* = -10 \quad M_{C,F}^* = -10 \quad M_{U_2,F}^* = -10$$



With this new rate corrected distance, we create a new distance matrix

$$\begin{bmatrix} & U_1 & G_C & U_2 & G_F \\ U_1 & 0 & -11 & -10 & -10 \\ G_C & -11 & 0 & -10 & -10 \\ U_2 & -10 & -10 & 0 & -10 \\ G_F & -10 & -10 & -10 & 0 \end{bmatrix}$$

The new node joins galaxies C and node U_1 , and we name it U_3 .

$$S_{C,U_{1,C}} = 2$$

$$S_{U_1,U_{1,C}} = 1$$

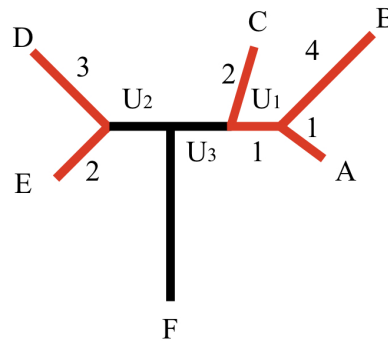


Figura 4.4: Star representation of the distribution of our galaxy sample in our third cycle. The branches in black represent the unknown length and red branches represent Know distances. We have a tree with three nodes represented by letter U.

We calculate the new distances to make our forth and final cycle to finish our tree. $N = 3$ Number of endnodes.

$$S_{U_2, U_3} = \frac{(S_{C,A} + S_{C,B} - S_{A,B})}{N - 2} = 3$$

$$S_{U_2, F} = 6$$

$$S_{U_3, F} = 6$$

$$\begin{bmatrix} & U_2 & U_3 & G_F \\ U_2 & 0 & 3 & 6 \\ U_3 & 3 & 0 & 6 \\ G_F & 6 & 6 & 0 \end{bmatrix}$$

$$S_{U_2} = \frac{2 + 6}{3 - 2} = 8 \quad S_{U_3} = 8 \quad S_F = 12$$

$$M_{U_2,F}^* = 6 \sim 8 \sim 12 = \sim 14 \quad M_{U_3,F}^* = 6 \sim 8 \sim 12 = \sim 14 \quad M_{U_2,U_3}^* = 2 \sim 8 \sim 8 = \sim 14$$

Here we could choose the combinations $U_4 = U_{2,F}$, $U_4 = U_{3,F}$ or $U_4 = U_{U_2,U_3}$ for our final node.

$$S_{U_2,U_4} = 1$$

$$S_{U_3,U_4} = 1$$

Now we finalize our tree joining F with U_4 with branch length 5.

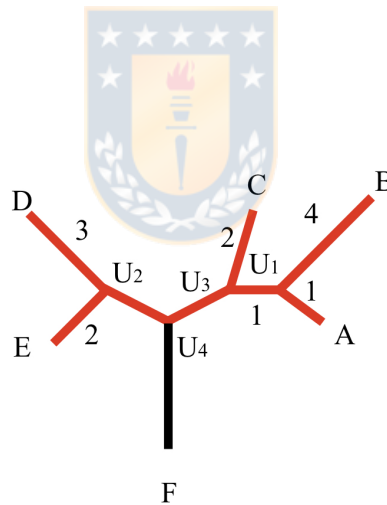


Figura 4.5: Star representation of the distribution of our galaxy sample in our four cycle, the branches in black represent the unknown length and red branches represent Know distances.

We can no longer keep on calculating distances, and we only have left the galaxy F, therefore we calculate it distance to the last node.

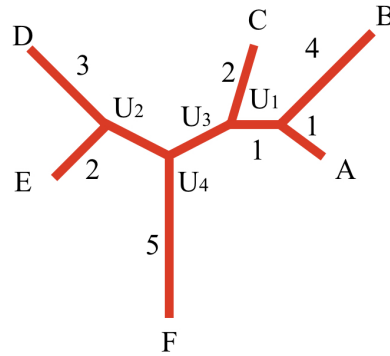


Figura 4.6: Final Star representation of the distribution of our galaxy sample, the branches in black represent the unknown length and red branches represent Know distances. The U letters represent the nodes between galaxies.

Now we have the distances for all our galaxies and we can form our final tree.

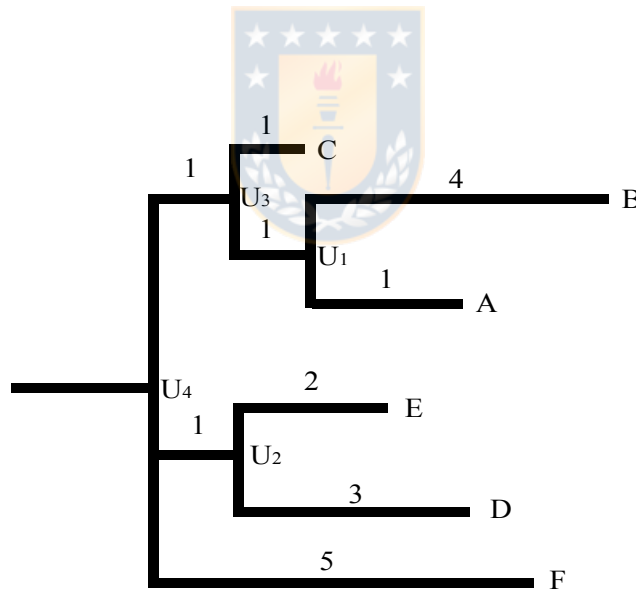


Figura 4.7: Final Phylogenetic tree, each galaxy is represented by a letter at the end of the branches. Each branch is joined to an specific node U. The numbers represent the length of the branches.

With this algorithm galaxies that have similar chemical composition, galaxies A and B, belong to the same node at the top of the tree. Their next closest neighbour in stellar content is galaxy C. Galaxies A, B and C belong to the same branch which we interpret

as the same GP. In contrast, galaxies E and D belong to another branch separated by node U_4 , therefore, a different GP. On the other hand, galaxy F does not belong to any GP. The distance between nodes U_1 and U_3 works as a chemical index that will allow us to differentiate galaxies A and B from galaxy C, even when they belong to the same population. Finally, galaxies that are in the tree canopy are the galaxies whose chemical composition are most similar, and, therefore, they have the longest distance to the root. Galaxies at the top of the tree end up being actually chemically different from galaxies at the root of the tree, even when they belong to the same population. Therefore the length between the nodes (Hereafter: NodeLength) along the tree, will help us to understand which parameters can influence the differentiation of galaxies from the top of the tree until the root.

In order to obtain a consistent result for our different galaxy samples, we will calculate 1,000 trees utilizing Montecarlo sampling from a normal distribution where the center of the distribution is the value of the absorption index and the width, the error of the measurement from SDSS DR14 (Abolfathi et al. 2018). Also, we make bootstrapping with replacement so that every time we calculate our distance matrix with Montecarlo sampling we will have a different combination of 29 or 30 absorption indices to generate our tree. Finally, we make a majority rule consensus, that considers structures in the tree that repeat at least a 50% of the time in the trees that we calculated. This procedure allows us to eliminate branches that are not statistically significant.

5. Analysis

5.1. Proof of concept

5.1.1. Globular Cluster NGC2808

As a proof of concept, we calculate a phylogenetic tree for our sample of 117 stars in the globular cluster NGC2808. We obtain one main branch from the majority consensus tree of 1000 trees appearing 730 times, as we can see in Fig 5.1.

The stars of the globular cluster were separated into five groups in Carretta 2015 (Hereafter: C15) in order to define five populations as we can see in Fig 5.2 with the classical Na-Mg anticorrelation. They order the populations by the decreasing of Mg abundance, classifying them into primordial or almost primordial abundance ratios (P1 and P2), intermediate composition (I1 and I2) and with severely extreme changes from the original composition (E). To be able to determine if two sets of data were significantly different from each other, they apply a Student's and Welch's test. They test the null hypothesis that any pair of components are extracted from a distribution having the same mean $[element/Fe]$, for $[Fe/H]$, $[O/Fe]$, $[Na/Fe]$, $[Mg/Fe]$, $[Al/Fe]$, $[Si/Fe]$, and $[Na/Mg]$. They list the number of degrees of freedom, the t-value and the two-tailed probability, finding that only E population seems to be statistically different from the intermediate and primordial populations.

We plot the Na-Mg anticorrelation highlighting the stars belonging to our main branch in Fig 5.3. We can notice that the stars of our main branch correspond to the E population defined by C15. Therefore our phylogenetic analysis can be used to detect a population of stars chemically different from the original composition of the globular cluster.

In order to study which classes of indices influence our population with different composition, we re-calculate our consensus tree without a specific chemical element. We observe that if we remove $[O/Fe]_I$, $[Na/Fe]_I$, or $[Mg/Fe]_I$, we get a tree with no structures. Therefore these three chemical elements are fundamental in the extreme population found.

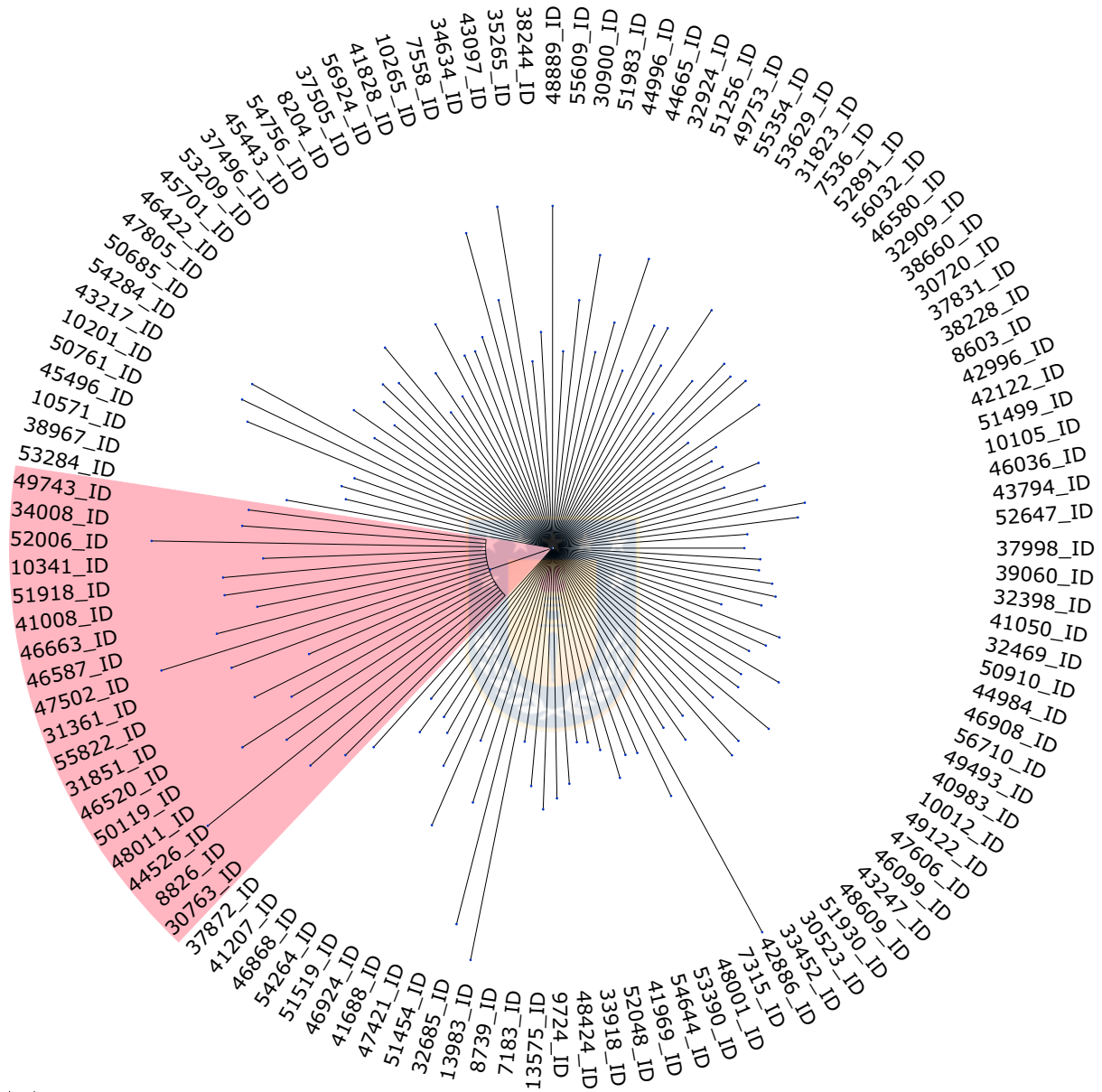


Figure 5.1: Consensus phylogenetic tree of the globular cluster NGC2808. The red shaded area highlight the main branch of the phylogenetic tree. The numbers correspond to the ID of each star in the globular cluster.

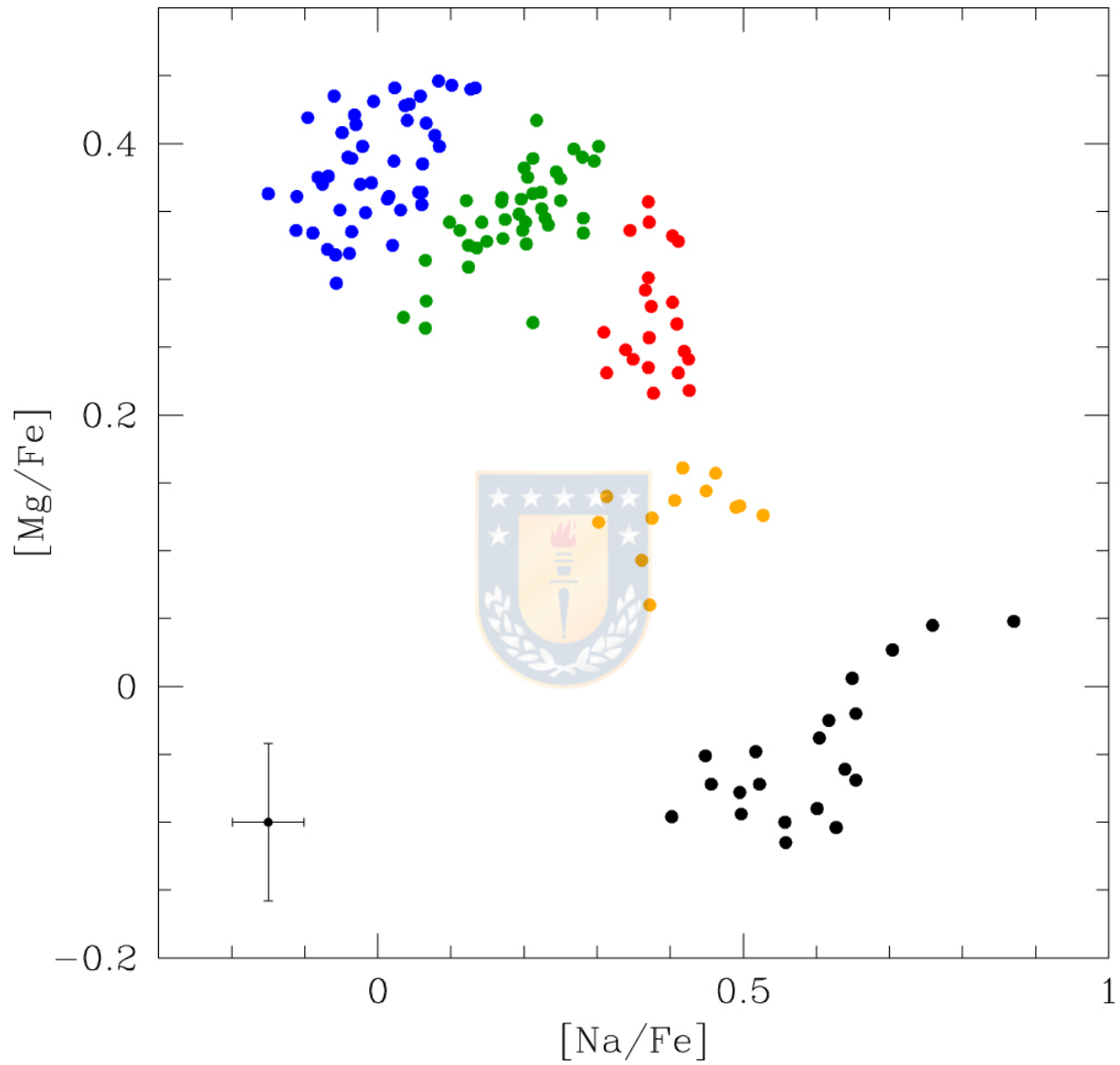


Figure 5.2: Mg and Na anticorrelation from figure 8 C15. Different colours indicate different groups, blue (P1), green (P2), red (I1), orange (I2), and black (E)

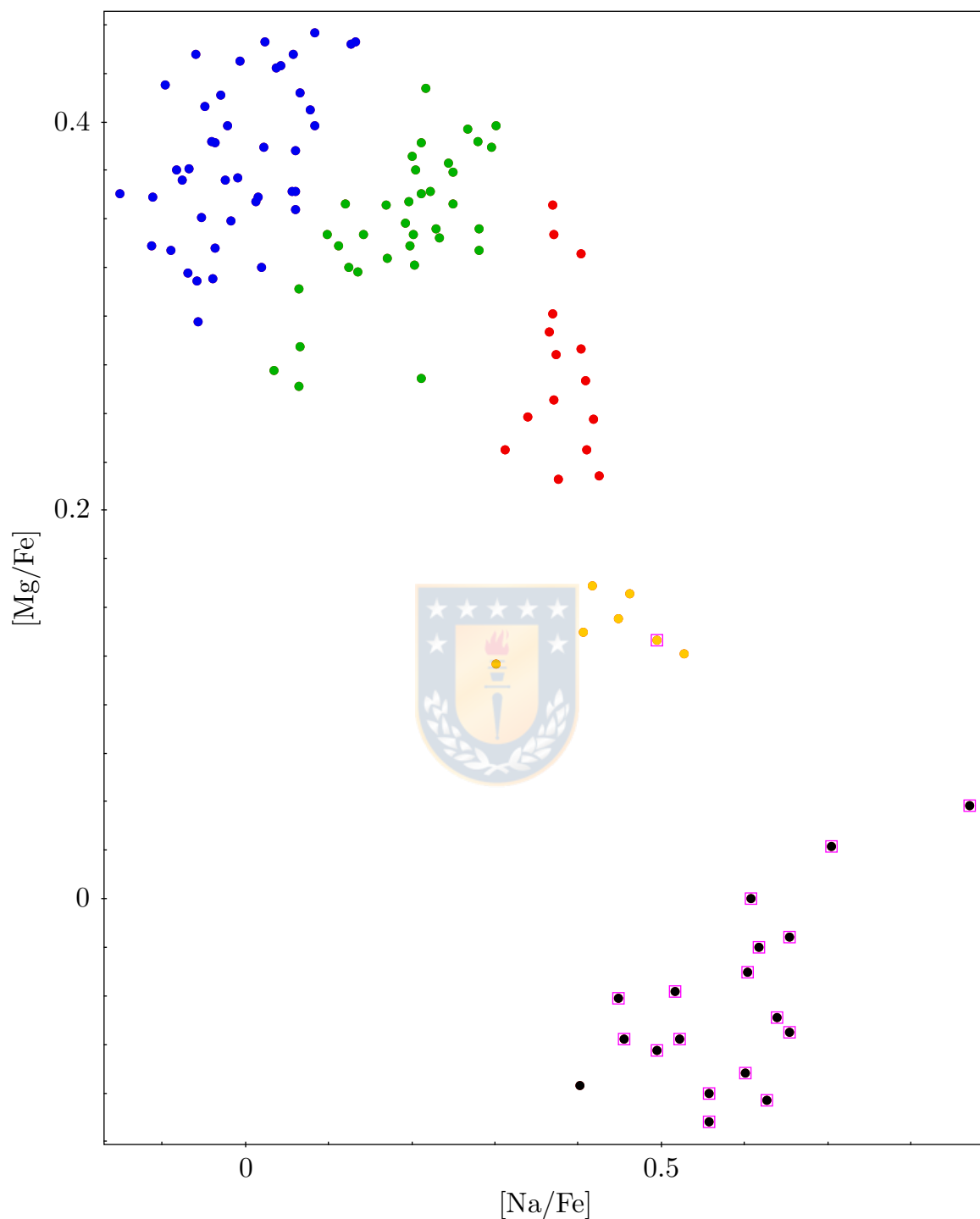


Figura 5.3: Na-Mg anticorrelation of the globular cluster NGC2808, with the five groups define by C15 differentiated by color. Blue (P1), green (P2), red (I1), orange (I2), and black (E). The magenta hollow squares represent the stars in our main branch of the phylogenetic tree.

Cuadro 5.1: Galaxy distribution in the Phylogenetic tree of the Coma Cluster

Population	No. of members
Branch 1	10
Branch 2	18
Branch 3	67
Branch with 1 node	90
Branch with 2 node	39
Branch with 3 node	18
Branch without nodes	233

5.2. Phylogenetic study in Galaxies

5.2.1. The Coma Cluster

We calculate a phylogenetic tree for our galaxy sample of the Coma cluster. Our tree presents three main branches that we arbitrarily named branch one (B1), branch two (B2) and branch three (B3). From the majority consensus rule of 1,000 trees, they appear in 683, 694 and 655 trees, respectively. Several other minor branches are found that appear at least 50 % of the time and whose complete description is shown in table 5.1.

Each of our branches represents a GP and they are differentiated by their chemical content. In what follows, we will focus in the three main GPs B1, B2 and B3. In order to prove that our GPs are different, we plot them in Color-Magnitude space, indicating also galaxy morphology.

From Figure 5.4, branches B1 and B2 belong to the bright-end of Red Sequence, and B3 belongs to the blue cloud of the cluster. This shows us that at least populations B1 and B2 are different from population B3, because they belong to different areas in the Color-Magnitude diagram. From the consensus morphology, we see that B1 has 3 elliptical galaxies and 7 S0 galaxies, B2 has 15 elliptical galaxies and 3 S0 galaxies, whereas B3 has 50 spiral galaxies and 17 S0 galaxies. We plot each population as a single tree so the morphology can be visualized. (See 5.5, 5.6 and 5.7).

If we look at how these GPs are distributed in a sky projection in the galaxy cluster, from figure 5.8 we see that that 1) galaxies in B1 population show lenticular morphology at the outskirts of the cluster and elliptical morphology in the cluster core; 2)

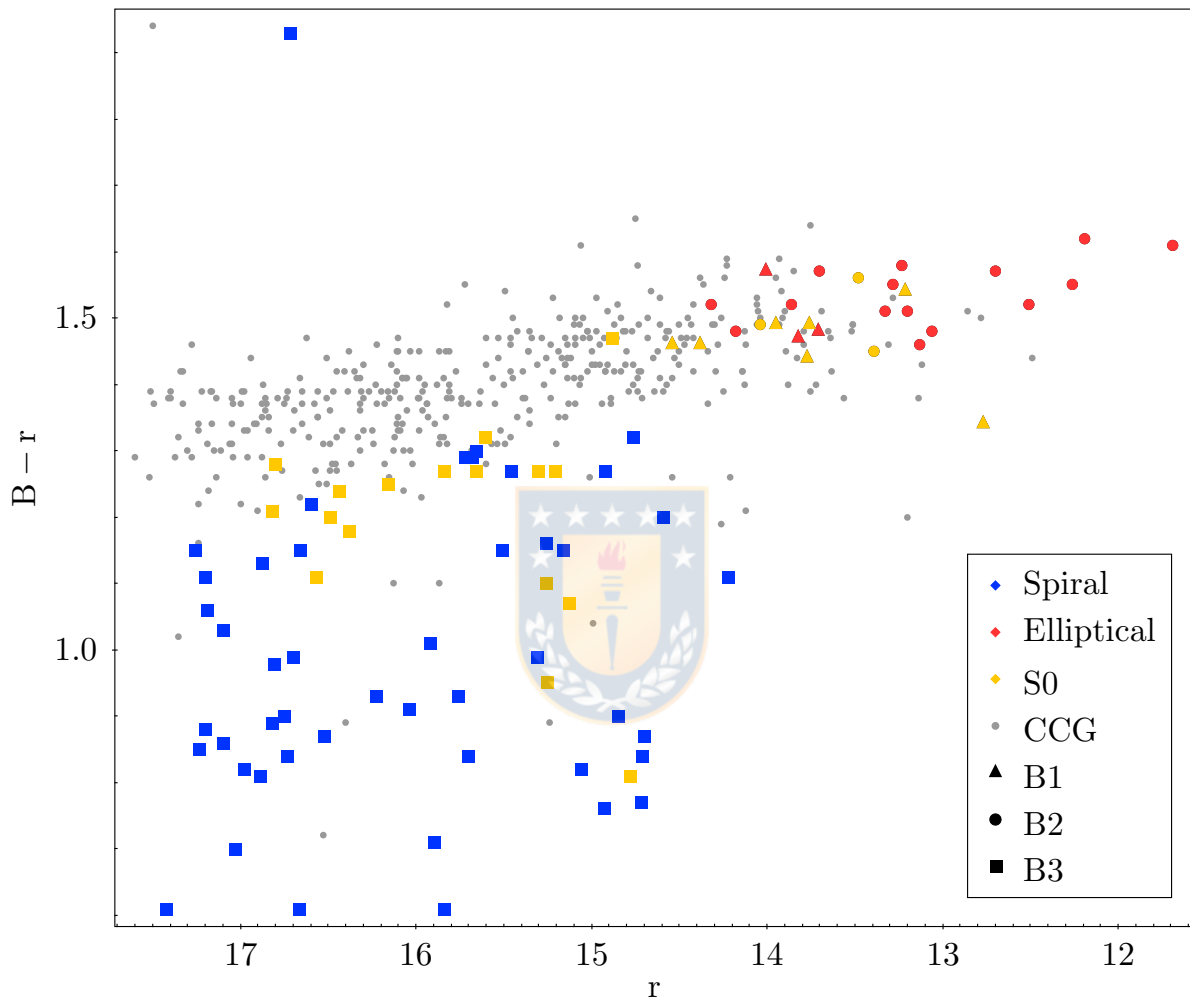


Figure 5.4: Color-Magnitude diagram of the Coma Cluster with photometric filters Johnson B and Sloan r. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate galaxy morphology where red corresponds to elliptical galaxies, blue to spiral galaxies, and yellow to S0 galaxies.

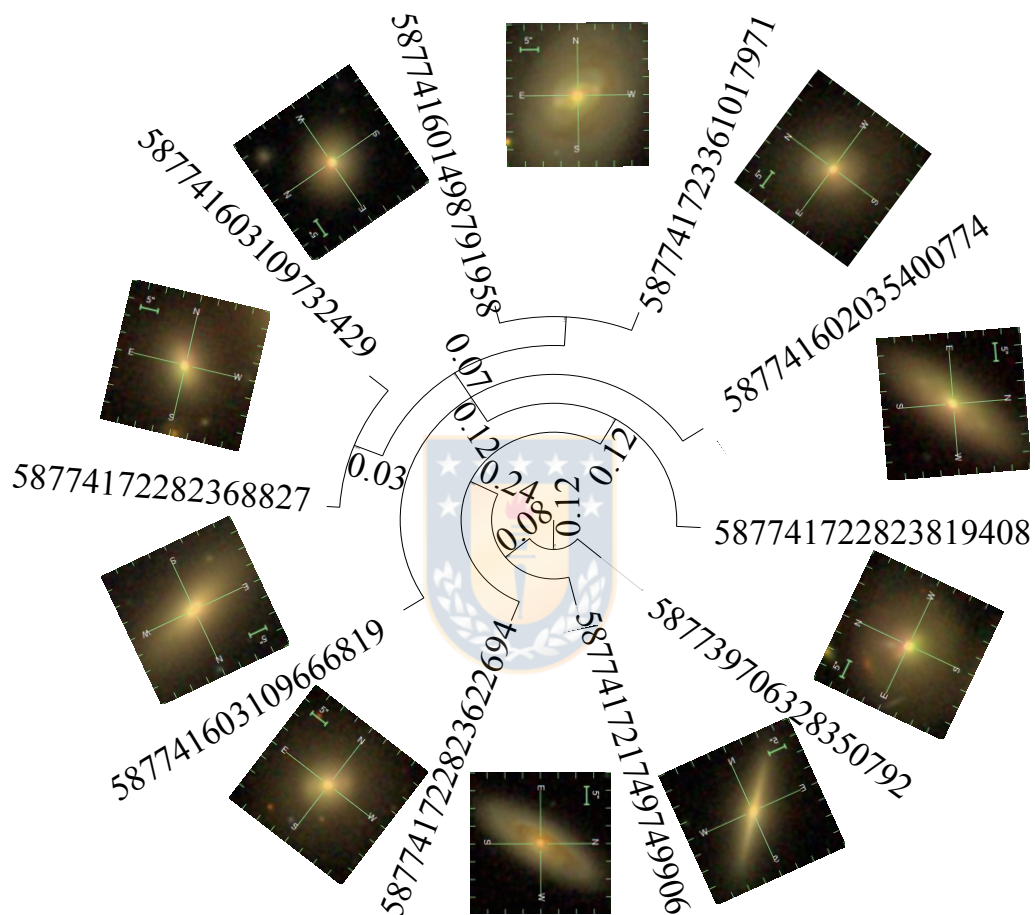


Figure 5.5: Branch B1 of the phylogenetic tree of the Coma Cluster that represents our first galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the length of their separation.

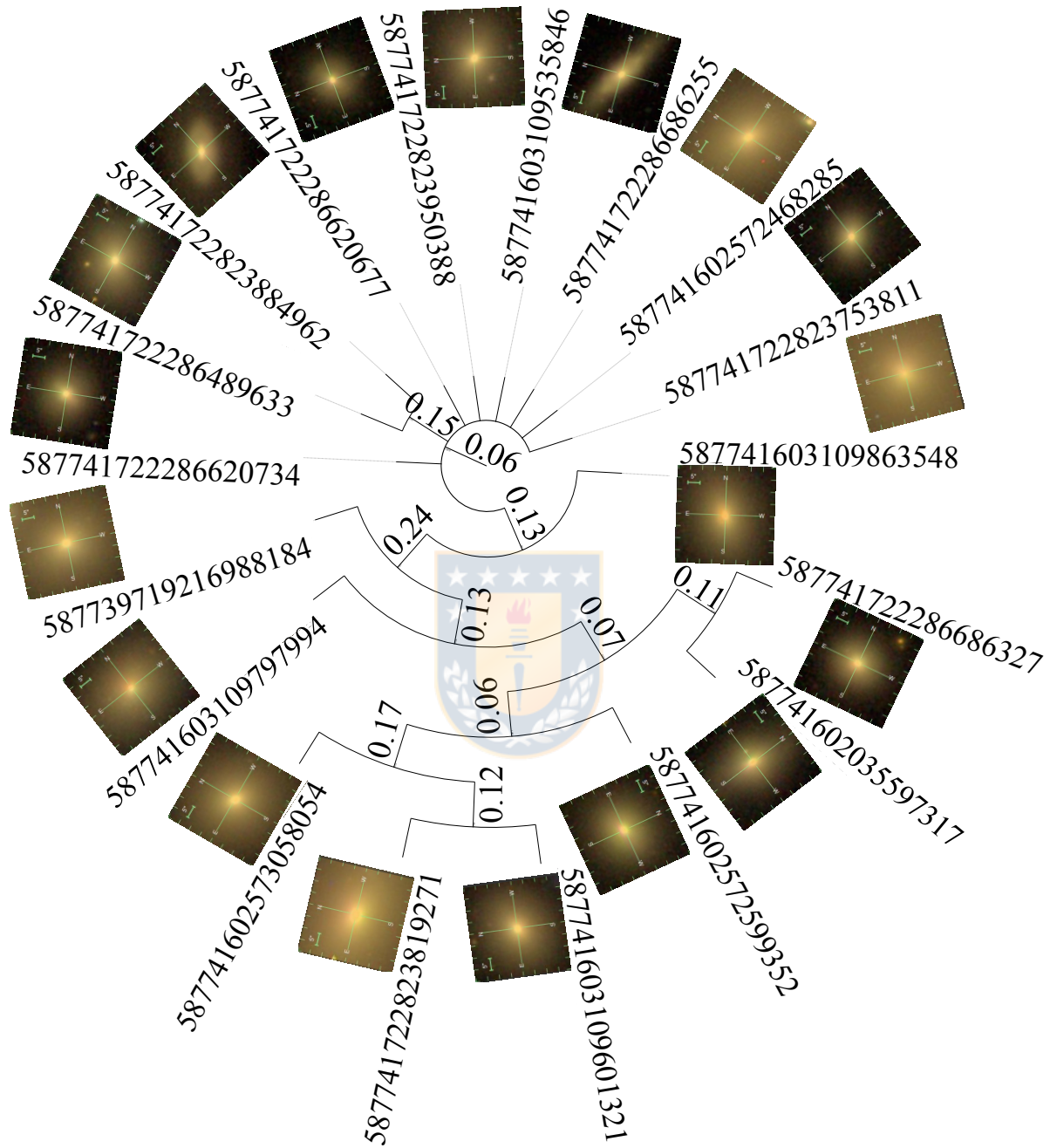


Figura 5.6: Branch B2 of the phylogenetic tree of the Coma Cluster that represents our second galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the chemical length of their separation.

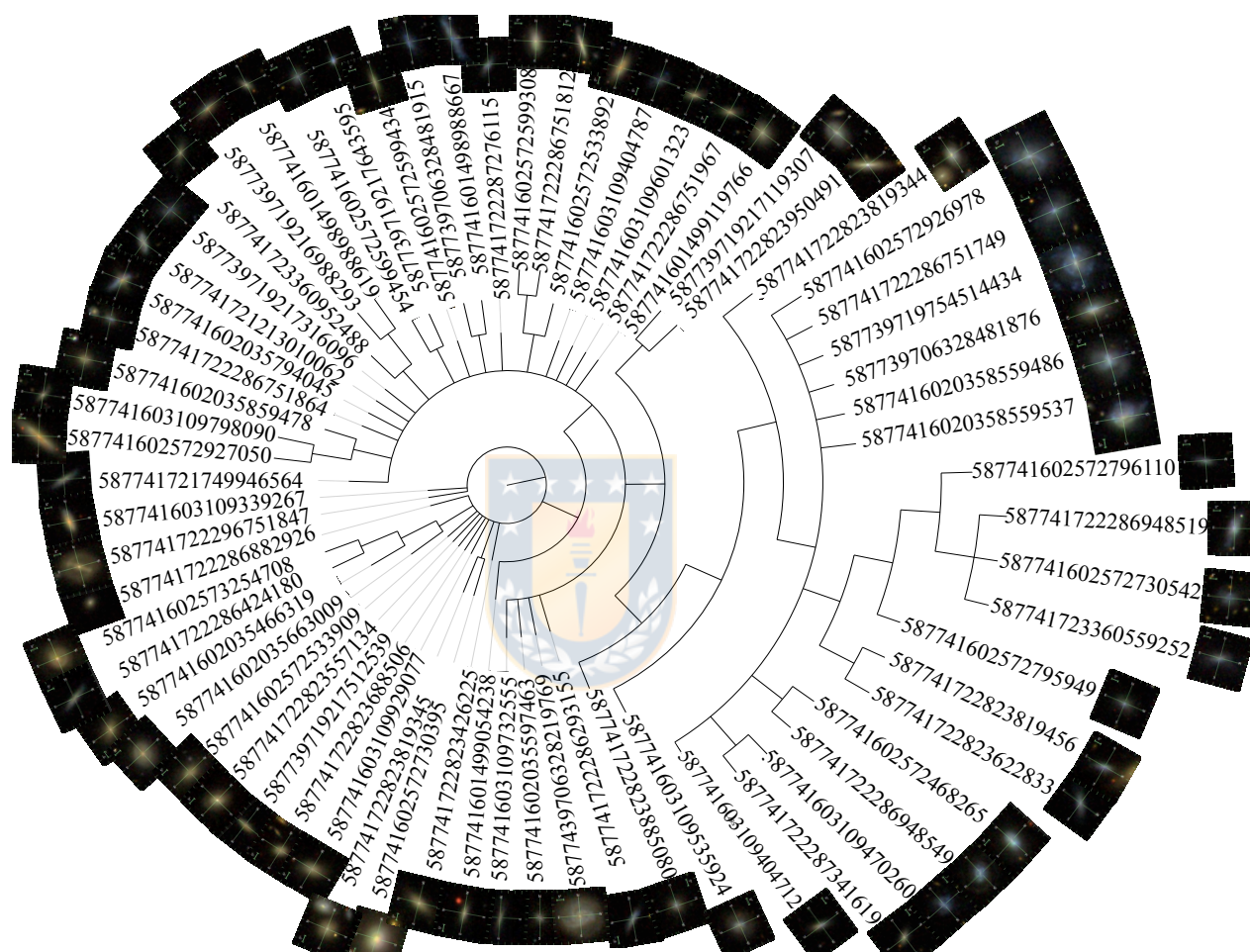


Figure 5.7: Branch B3 of the phylogenetic tree of the Coma Cluster that represents our third galaxy population. Each galaxy has its ID number from the SDSS DR7, and its image from the SkyServer DR15. The number between the nodes is the length of their separation.

galaxies in B2 population are distributed in two clumps in the cluster, one in the center and the other near the galaxy NCG 4839 which is associated with a subgroup of galaxies in-falling into Coma Cluster (Neumann et al. 2001); and 3) galaxies in B3 population do not show any preferential distribution in the cluster.

In order to characterize the common properties of the GP, we compare their stellar mass, sSFR, metallicities and ages in Figure 5.9. From this comparison we can conclude that B1, B2 and B3 are indeed three different GPs. B1 galaxies are quiescent with a high stellar mass ($\log(M/M_{\odot}) = 10,35 - 11,17$), high metallicity ($[Z/H] = 0,24 - 0,59$), low sSFR ($10^{-11,25} - 10^{-10,53} \text{ yr}^{-1}$), however of predominantly S0 morphology and with young-to-intermediate age stellar populations (1,88 – 5,38 Gyr). On the other hand, B2 galaxies correspond to quiescent galaxies with high stellar mass ($\log(M/M_{\odot}) = 10,62 - 11,76$), high metallicity ($[Z/H] = 0,21 - 0,49$), low sSFR ($10^{-11,25} - 10^{-10,53} \text{ yr}^{-1}$), intermediate-to-old-age stellar populations ($< 3,74$ Gyr) and mostly of elliptical morphology. B3 galaxies have low stellar mass ($\log(M/M_{\odot}) = 8,39 - 10,41$) with low metallicity ($[Z/H] = -1,65 - 0,55$) and high sSFR ($10^{-10,25} - 10^{-8,17} \text{ yr}^{-1}$), of predominantly spiral morphology and with stellar populations of a wide range in age (0,12 – 10,89 Gyr).

Now we analyze each branch's hierarchical structure. As we show in section 4.1, the NodeLength works as a chemical index. This chemical index indicates the variation in the chemical composition of galaxies. In order to see how the general properties of mass, sSFR, ages and $[Z/H]$ relate internally in our GP, we add the node length from the root of the tree to the leaves where each galaxy is. See fig. 5.10

We calculate the Pearson and Spearman correlation coefficients between the NodeLength and the stellar mass, metallicity, ages, and sSFR to see which of these properties is the main driver of the measured chemical gradient by determining if any linear or monotonic correlation exists. The results are summarized in table 5.2. We can see that B1 anti-correlates with age and sSFR, while it correlates with metallicity. Also, B3 presents a monotonic correlation with sSFR. This even shows us when our GP have different stellar mass ranges, although stellar mass is not relevant in their hierarchical relationships. Stellar mass and sSFR do not play any role in the hierarchical structure of our phylogenetic tree either.

Study general or average properties is not enough to fully understand what the physical

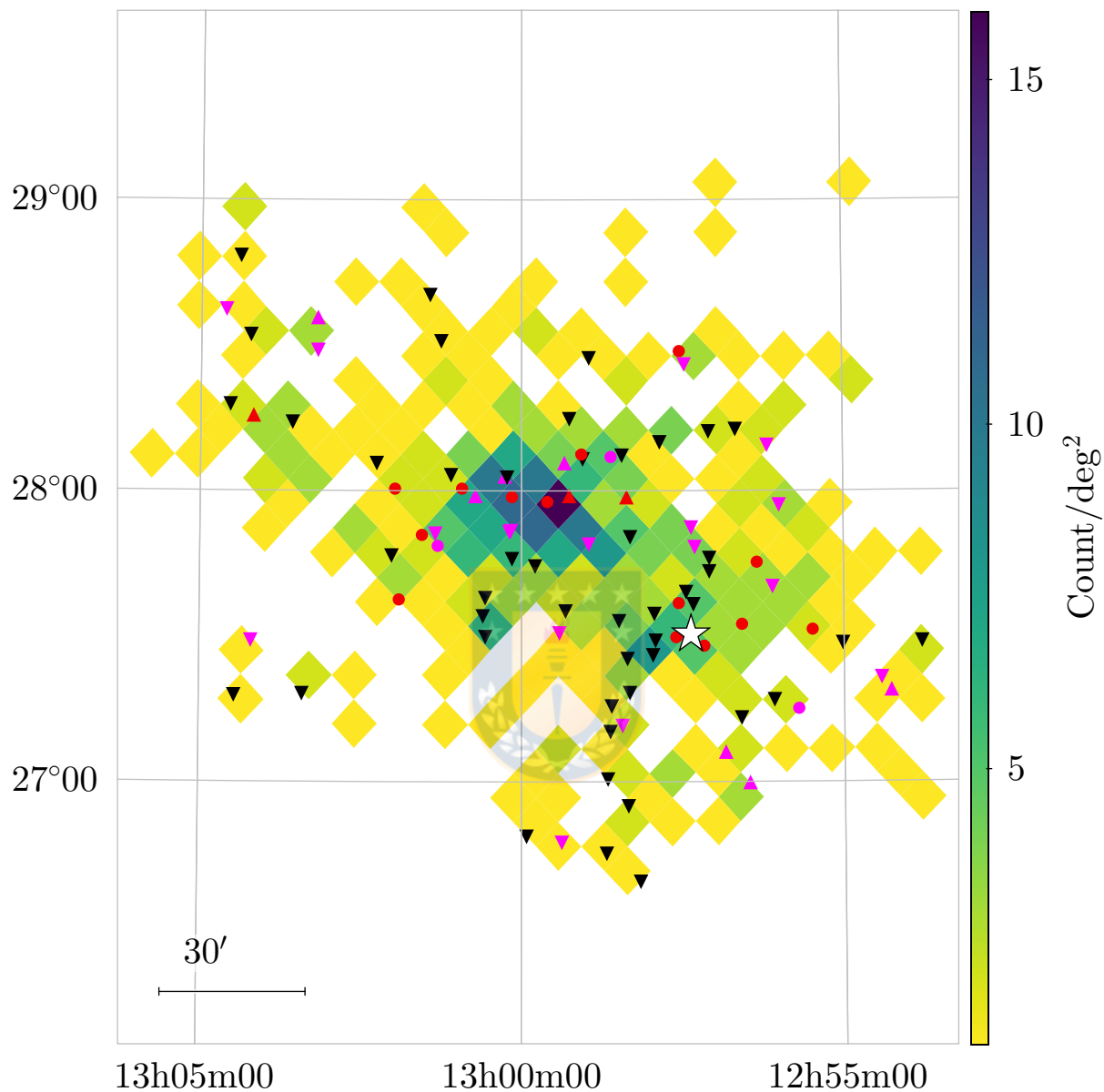


Figure 5.8: Projected sky distribution of galaxies in the Coma Cluster belonging to different branches of the phylogenetic tree. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate the morphology of the galaxy where red corresponds to elliptical galaxies, black to spiral galaxies, and magenta to S0 galaxies. The color scale shows the number count of galaxies by degree squared in the sky for Coma cluster. White star points the galaxy NGC4839.

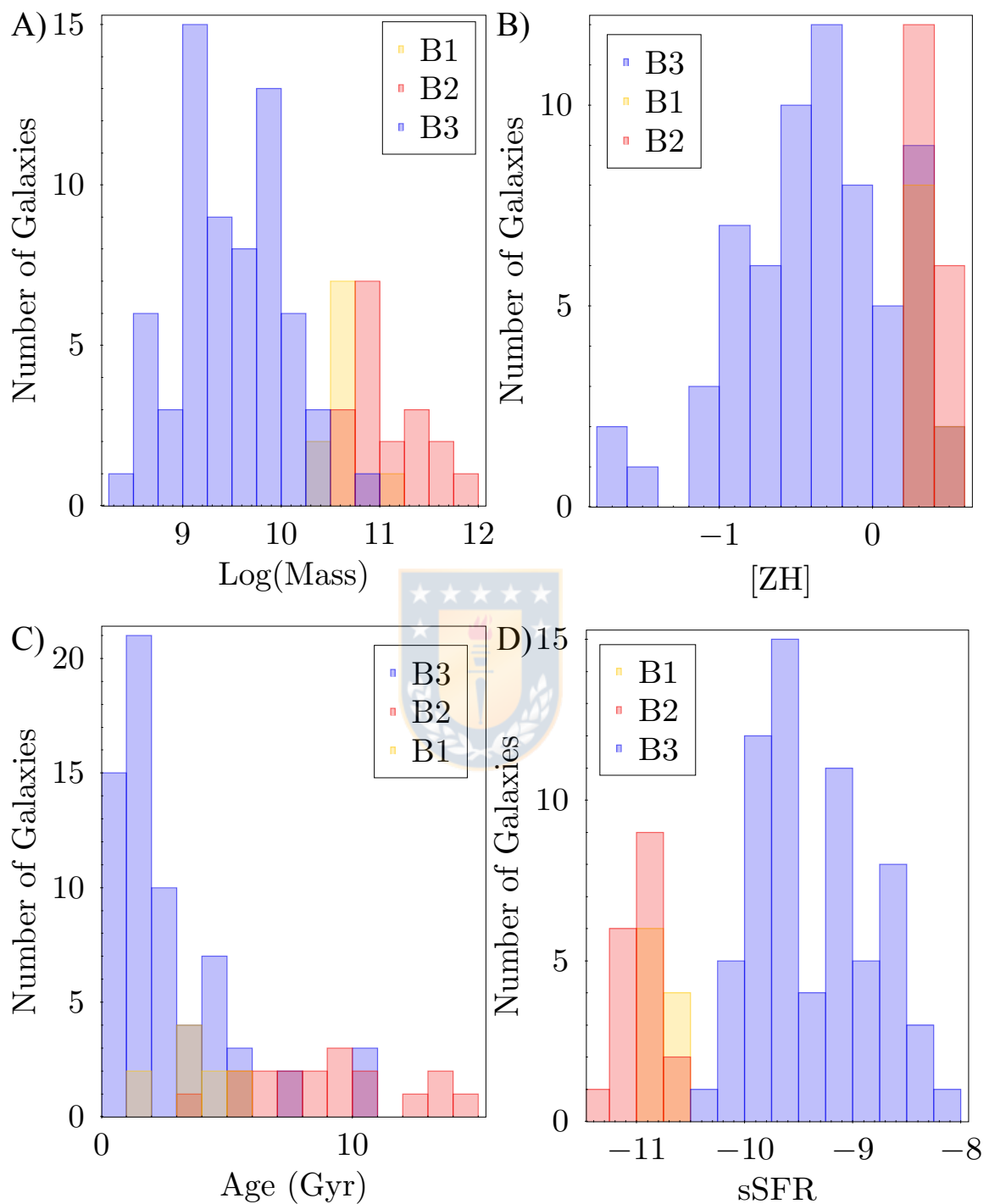


Figure 5.9: The histograms A, B, C and D show the distribution of galaxies from the three different branches B1, B2 and B3 with respect to stellar mass, metallicity, age and sSFR. The yellow, red and blue colors represent galaxies from branches B1, B2 and B3, respectively.

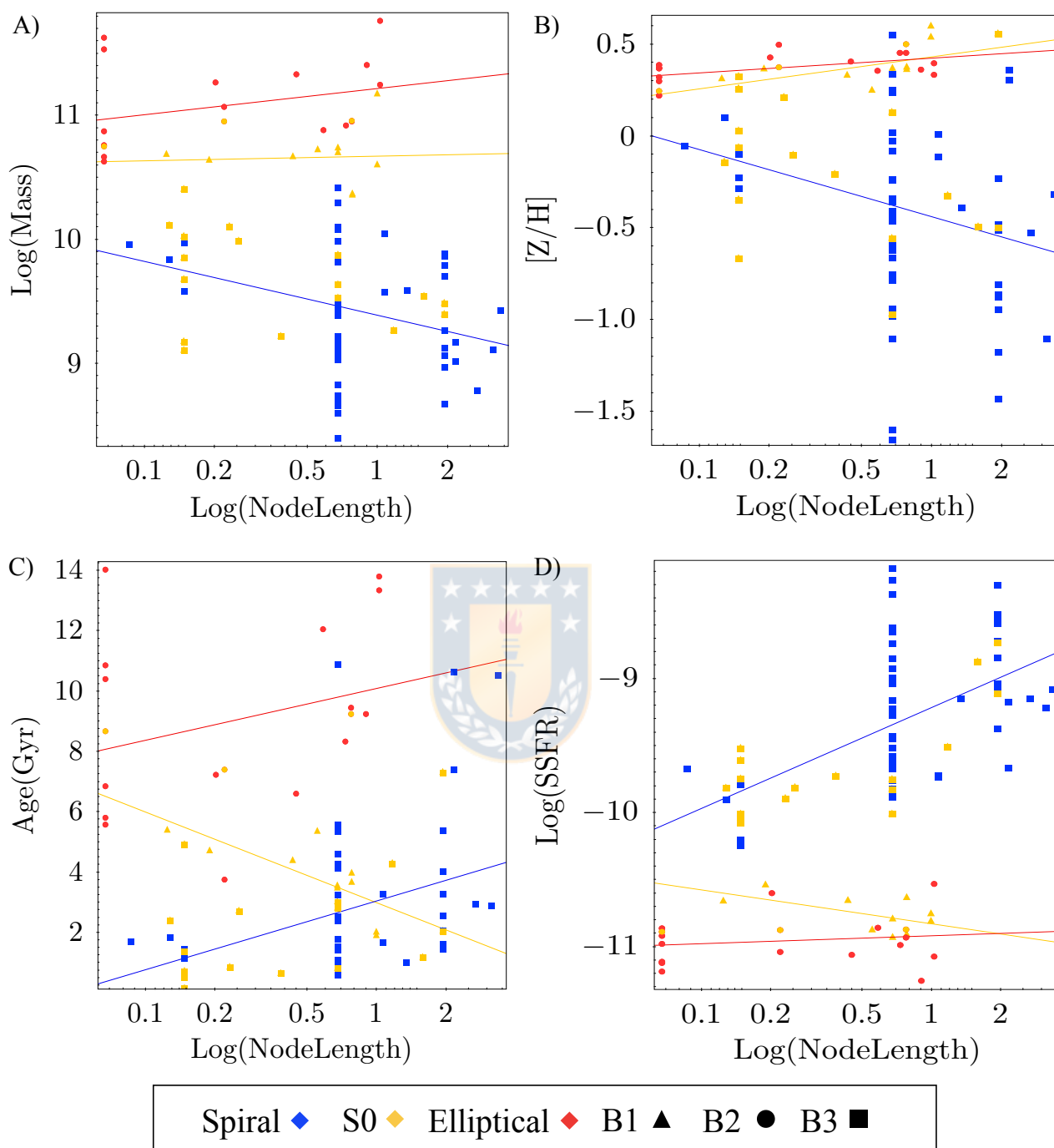


Figure 5.10: Panels A, B, C and D, show a comparison of our chemical index (length between nodes) with stellar mass, metallicity, age of the stellar population, and sSFR, respectively. Triangles, circles and squares represent B1, B2 and B3 GP. The red, yellow and blue colors represent elliptical, lenticular and spiral morphology, respectively.

Cuadro 5.2: Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population

Branch	Pearson coefficient	Spearman coefficient
Length v/s Mass		
B1	0.12	-0.092
B2	0.351	0.436
B3	-0.303	-0.362
Length v/s [Z/H]		
B1	0.681	0.771
B2	0.388	0.444
B3	-0.250	-0.316
Length v/s Age		
B1	-0.861	-0.856
B2	0.439	0.332
B3	0.443	0.462
Length v/s SSFR		
B1	-0.536	-0.355
B2	0.129	0.115
B3	0.490	0.599

differences or common characteristics of galaxies in our GP are that determine the hierarchical structures we observe. Therefore we separate our abundance indices in five classes: (1) absorption Balmer lines with $H\delta_A$, $H\delta_F$, $H\gamma_A$, $H\gamma_F$, $H\beta$ and $H\beta_G$; (2) positive response to α/Fe enhancement with CN_1 , CN_2 , Mg_1 , Mg_2 , Mg_b , $Ca4227$ and $G4300$; (3) negative response to α/Fe enhancement with $Fe4383$, $Fe4531$, $Fe4930$, $Fe5015$, $Fe5270$, $Fe5270S$, $Fe5335$, $Fe5406$ and $Fe5709$; (4) Emission Line $[OIII]\alpha5007$, with $[OIII]_1$ and $[OIII]_2$ (Gonzalez et al. 2005); and (5) insensitive to α/Fe with NaD , $C4668$, $Fe5782$, $Ca4455$, TiO_1 and TiO_2 . Then we recalculate the phylogenetic tree with the same sample of galaxies but without one of these classes. We analyze our re-calculated consensus tree, by plotting color-magnitude diagrams. We graph the distribution of galaxies from our original tree, and that of the galaxies that appears in our re-calculated trees (see figure 5.11). We resume this results in Table 5.3. Now that we can see which classes of indices influence our GP, we can study in more detail the internal structure of our GP, analyzing the relations between each index and the NodeLength. (see figure 5.12, 5.13, 5.14). If an absorption index have a high correlation with NodeLength and at the same time have different range of values for each galaxy populations, this index is responsible for the

hierarchical structure of our branches.

Our main findings are:

- Absorption Balmer lines: from table 5.3 we can see that only the B2 population remains in our tree. Balmer lines decrease in strength as stellar populations get older, which is reflected in their influence on populations B1 and B3 (See figure 5.13-D). From plots D.1 to D.6 in figure 5.14 we see that Balmer lines have a high correlation for each galaxy population, however $H_{\delta A}$ (see figure 5.13-D.1) is the index that has a different value for each galaxy population. This analysis shows us that the $H_{\delta A}$ index is important for breaking the degeneracy between age and horizontal branch morphology Schiavon et al. 2004.
- Insensitive to α/Fe enhancement: this abundance indices are relevant for the B1 population, because as we see in table 5.3 and figure 5.13-B that B1 population is not present in our recalculated tree. Analyzing individual abundance indices, $N_a D$ index (see figure 5.13-B.1)) show a strong correlation with a different value for each GP. This absorption index measures Carbon, Magnesium and Sodium, elements that are produced and released to the interstellar medium by massive stars. This tells us that galaxies in B1 have undergone a period of star formation, and this is the reason for showing up in our phylogenetic tree as a separated galaxy population.
- Negative Response to α/Fe enhancement: at first glance from table 5.3, we see that the re-calculated tree, has tree main GP. But if we look at figure 5.10-A, we can notice that the B3 population is missing its S0 galaxies closer to the red sequence. The $Fe4383$ abundance index (see figure 5.12-A.1), show a correlation and different values for this S0 with other populations. These index measure Carbon, Iron and Magnesium, elements that are produced in Type Ia supernovae, indicating that there is a difference between S0 galaxies from population B1 and B3.
- Positive Response to α/Fe enhancement: From figure C) 5.13 we see that the presence of these lines are not relevant for the identification or hierarchical structures of our GP, because we obtain the same galaxy populations without this absorption indices.
- Emission Line $[OIII]\alpha 5007$, with $[OIII]_1$ and $[OIII]_2$ (Gonzalez et al. 2005): the recalculated tree without this indices, show a few galaxies from B2 and B3 and not show B1

Cuadro 5.3: Remaining galaxy populations for Recalculated phylogenetic trees without a specific set of line indices

Index Class	Branches
Blamer Lines	B2
Positive response to alpha/Fe enhancement	B3 B2 B1
Negative response to alpha/Fe enhancement	BE B3 B1
Insensitive to alpha/Fe	B3 B2
Emission Line $[OIII] \lambda 5007$	B2 B3

population. This indices not present correlations, have similar values for all our GPs (See figure 5.14). Therefore this indices are irrelevant for finding galaxy populations and their internal structure.

From this analysis we can infer that the minimum amount of indices that we need to perform a successful and consistent phylogenetic study is 18 (see figure 5.12, 5.13, 5.14), and those include the following ones: $H\delta_A$, $H\gamma_A$, $H\gamma_F$, $H\beta$, $H\beta_G$, CN_1 , CN_2 , Mg_1 , Mg_2 , Mg_b , $G4300$, NaD , $C4668$, TiO_2 , $Fe4383$, $Fe5270$, $Fe5270S$, $Fe5335$. We investigated the possibility of carrying such analysis with fewer indices. The results showed inconsistencies in the sense that the MC trials with bootstrapping delivered different outcomes in terms of tree structures. Hence, we do not recommend to apply such phylogenetic method with indices other than those indicated above.

5.2.2. The Field

One of the main drivers of galaxy evolution is the environment (Kauffmann et al. 2003, Kauffmann et al. 2003, Baldry et al. 2006, Peng et al. 2010). Therefore, in order to assess the differential effect between high and low density environments, we generated a phylogenetic tree with 438 field galaxies homogeneously distributed in the sky (see figure 5.15), at the coeval redshift range of Coma 0,034 – 0,054. From our consensus tree, we obtain a main branch with 46 galaxies that appears 561 times from 1,000 tree samples, and several minor structures (220 branches with 1, 2, 3, or four nodes and 172 in branches without nodes). We interpret the main branch as our main GP.

This population presents galaxies predominantly of lenticular morphology, that we can observe in more detail in figure 5.16, with high stellar masses ($\log(M/M_\odot) = 10,01 - 11,32$)

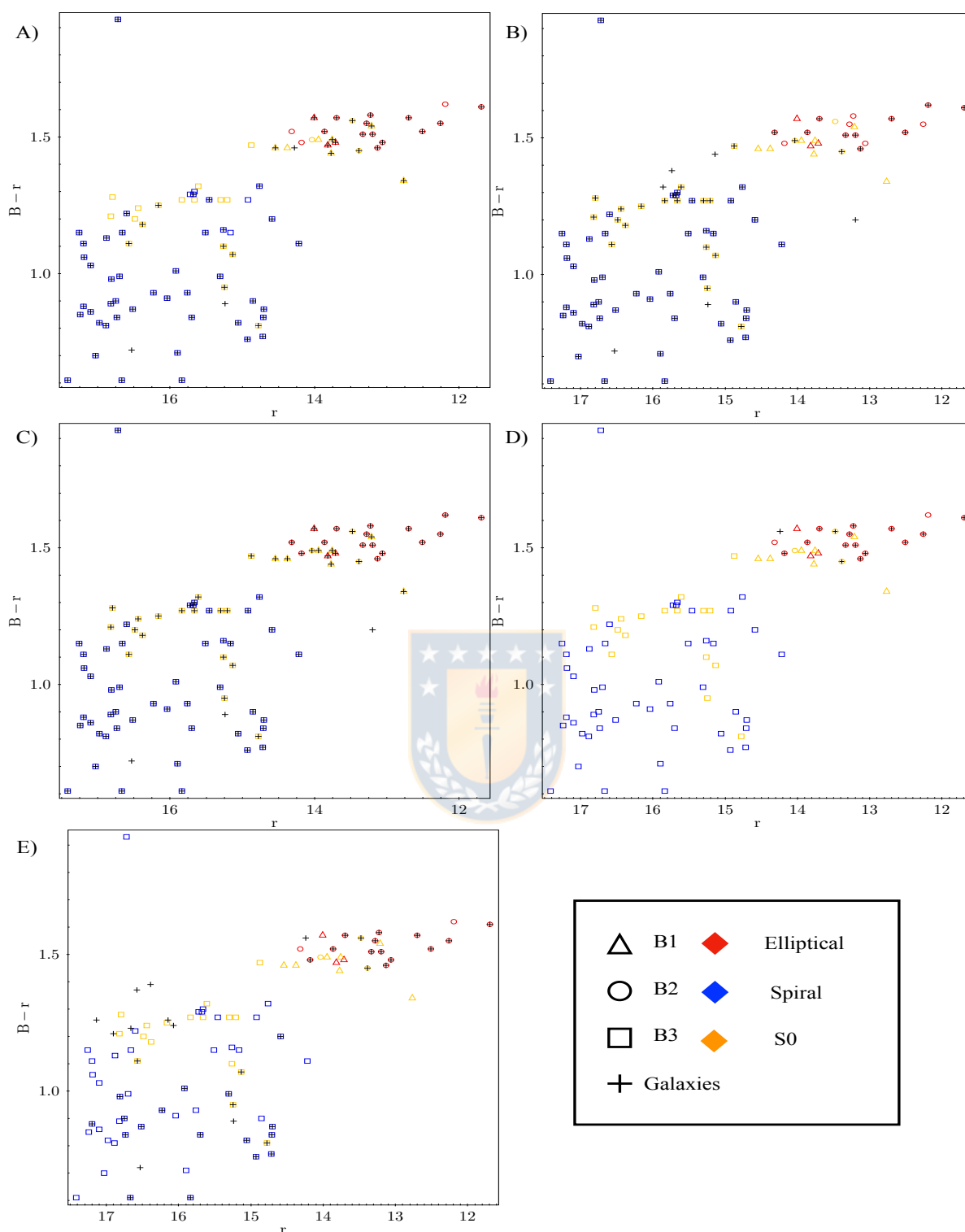


Figure 5.11: Color-magnitude diagram of our Coma cluster galaxy sample with photometric filters Johnson B and Sloan r. Panels A, B, C, D, and E show a recalculated phylogenetic tree without, negative response to α/Fe enhancement, insensitive response to α/Fe enhancement, positive response to α/Fe enhancement, absorption Balmer lines, and absorption line indices $[OIII]$, respectively. The cyan diamonds represent the galaxies that appear in the branches of each recalculated tree.

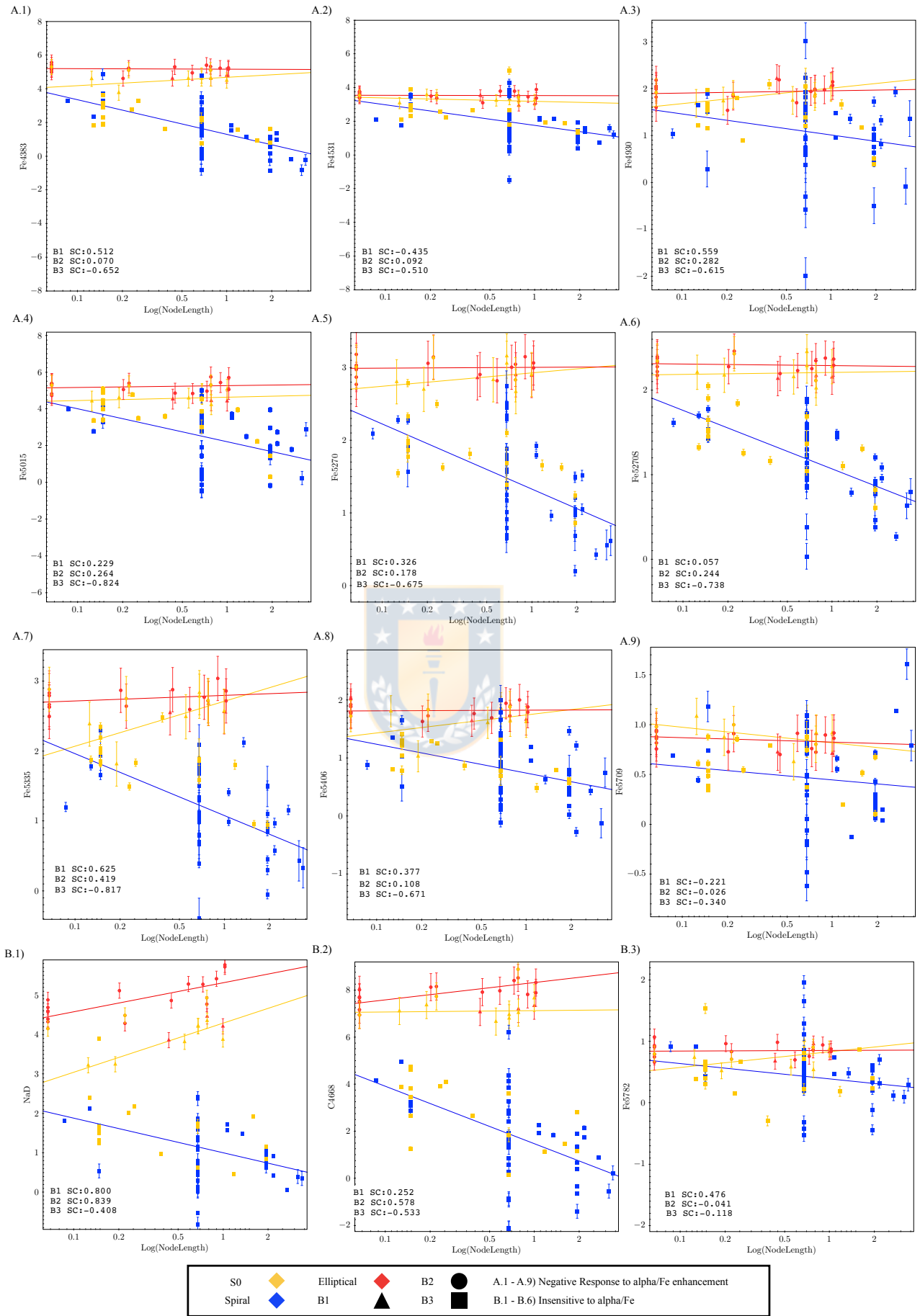
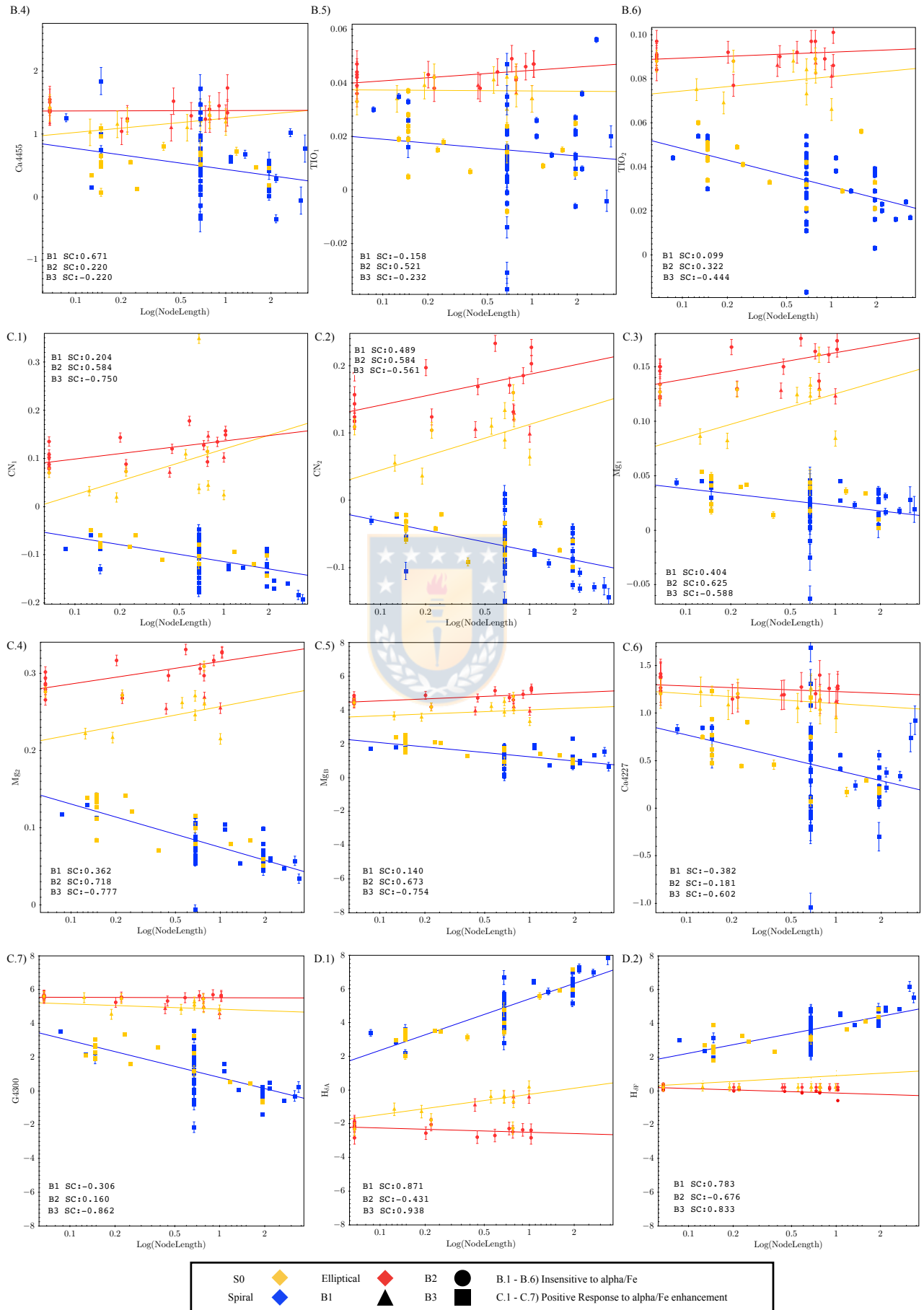


Figura 5.12

Figure 5.13: Continue image



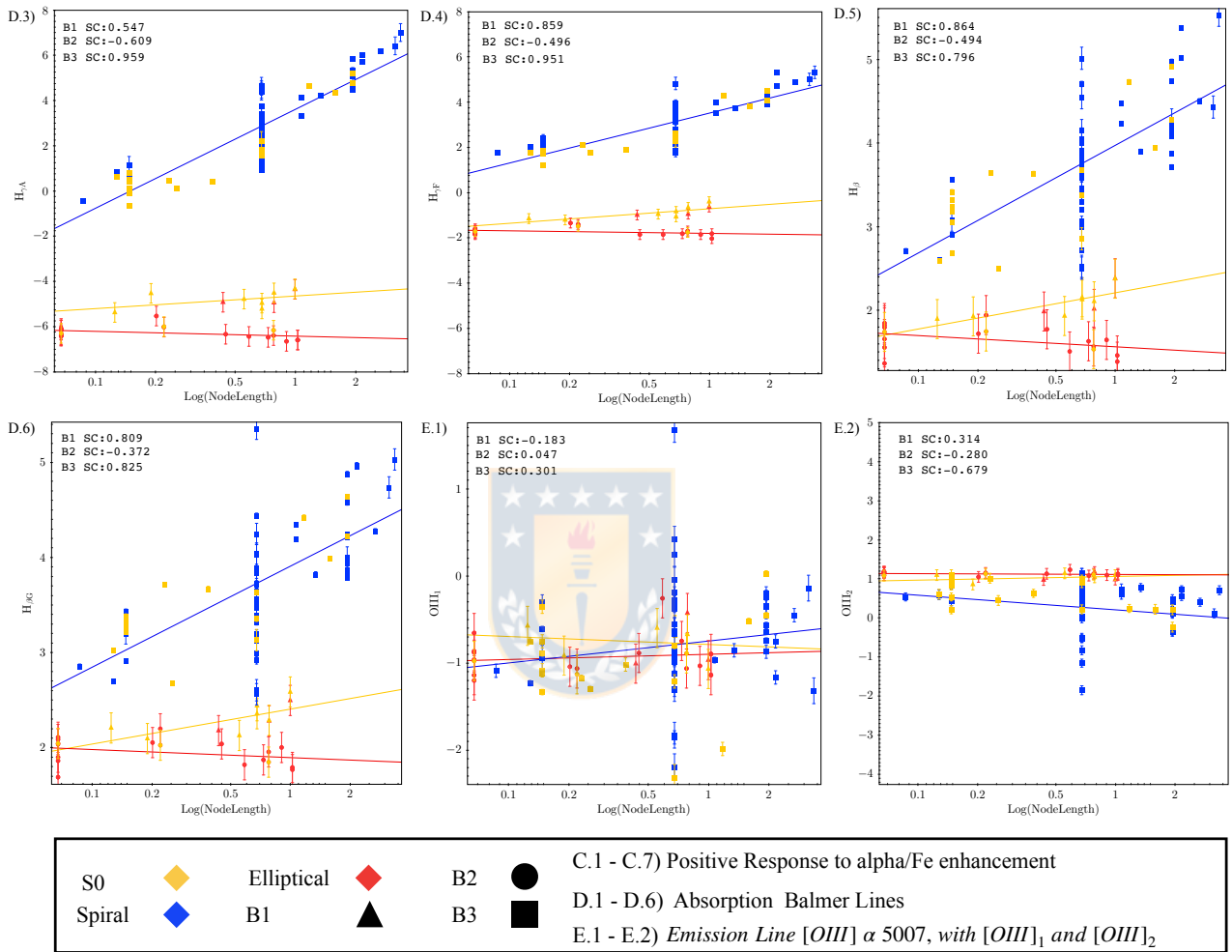


Figure 5.14: Comparison of absorption lines with length between nodes of our GP. Triangles, circles and squares represent the B1, B2, and B3 GP. The red, yellow and blue colors represent elliptical, lenticular and spiral morphology, respectively. Red, yellow and blue lines correspond to linear fits to each galaxy population. The Spearman correlation (SC) index is calculated for each galaxy population.

Cuadro 5.4: Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population

Branch	Pearson Coeff.	Spearman Coeff.
Length v/s Mass	0.087	0.026
Length v/s $[Z/H]$	0.472	0.410
Length v/s Age	-0.191	-0.132
Length v/s SSFR	-0.590	-0.636

and wide ranges of sSFRs ($10^{-10,99} - 10^{-8,72} \text{ yr}^{-1}$), age (3,0 – 14 Gyr) and metallicity ($[Z/H] = -0,33 - 0,34$). In order to analyze the internal structure of our field GP, we follow the same procedure than in section 5.2.1. We compare the lengths between nodes with common galaxy properties (See figure 5.14), and calculate their Spearman and Pearson correlation coefficients summarized in table 6.1.

The internal structure of the main branch shows a weak correlation with sSFR. In order to see whether a specific class of absorption index influences the detection of this GP, we calculate new phylogenetic trees, but excluding different classes of absorption lines, as we can see in figure 5.18. We notice that no particular class influences the outcome of our GP in the tree. We analyze the internal structure of our main GP, running the Spearman correlation test on every absorption index and NodeLength. The results are summarized in table 6.2. The internal structure of our main branch shows a correlation with absorption Balmer lines and lines that have a positive response to α/Fe enhancement.

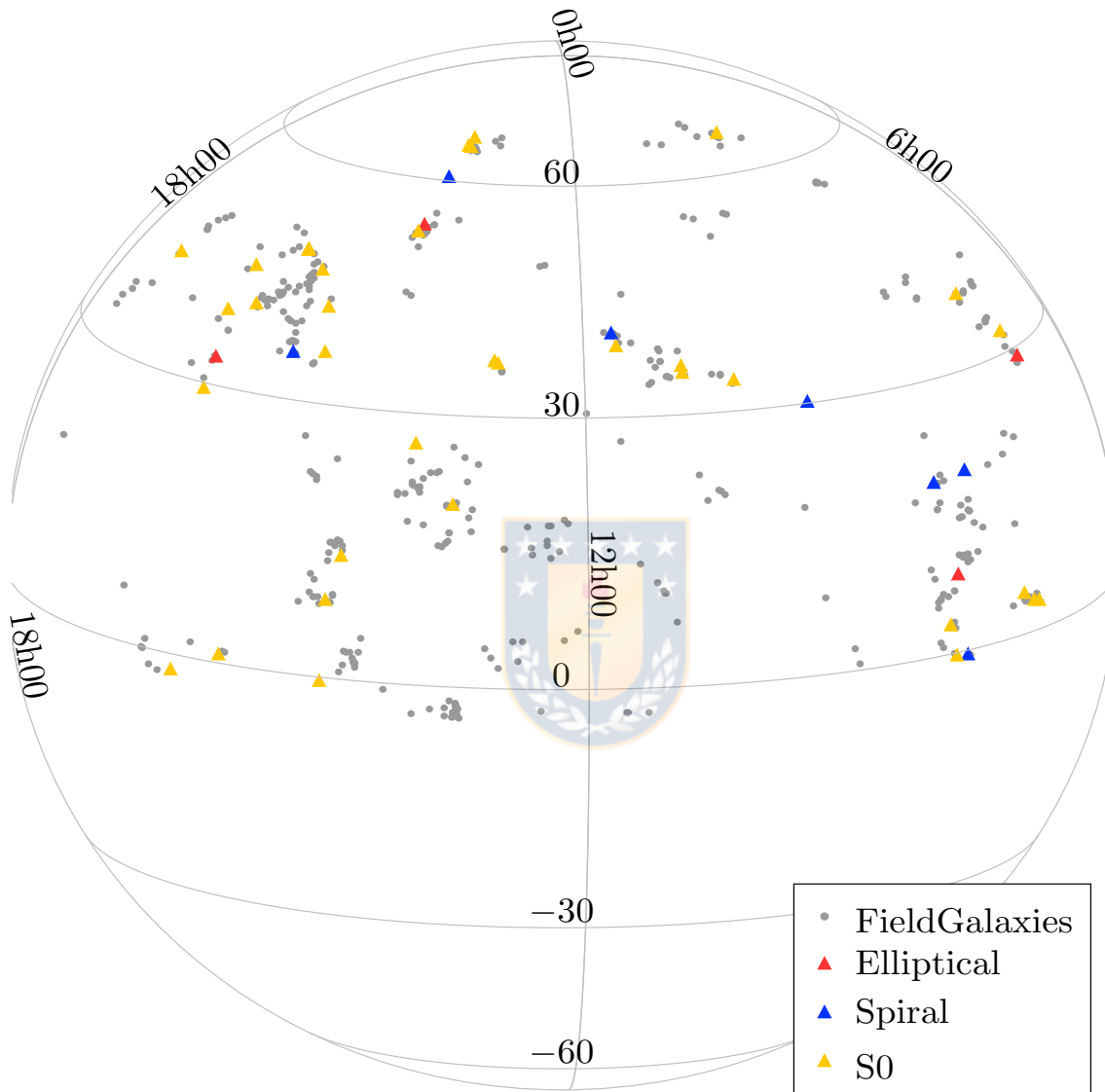


Figure 5.15: Spatial distribution of our field galaxy sample in the sky. Grey circles represent all galaxies in our sample, whereas triangles represent our main galaxy population. Red, yellow, and blue colors indicates, correspondingly, elliptical, S0 and spiral galaxies.

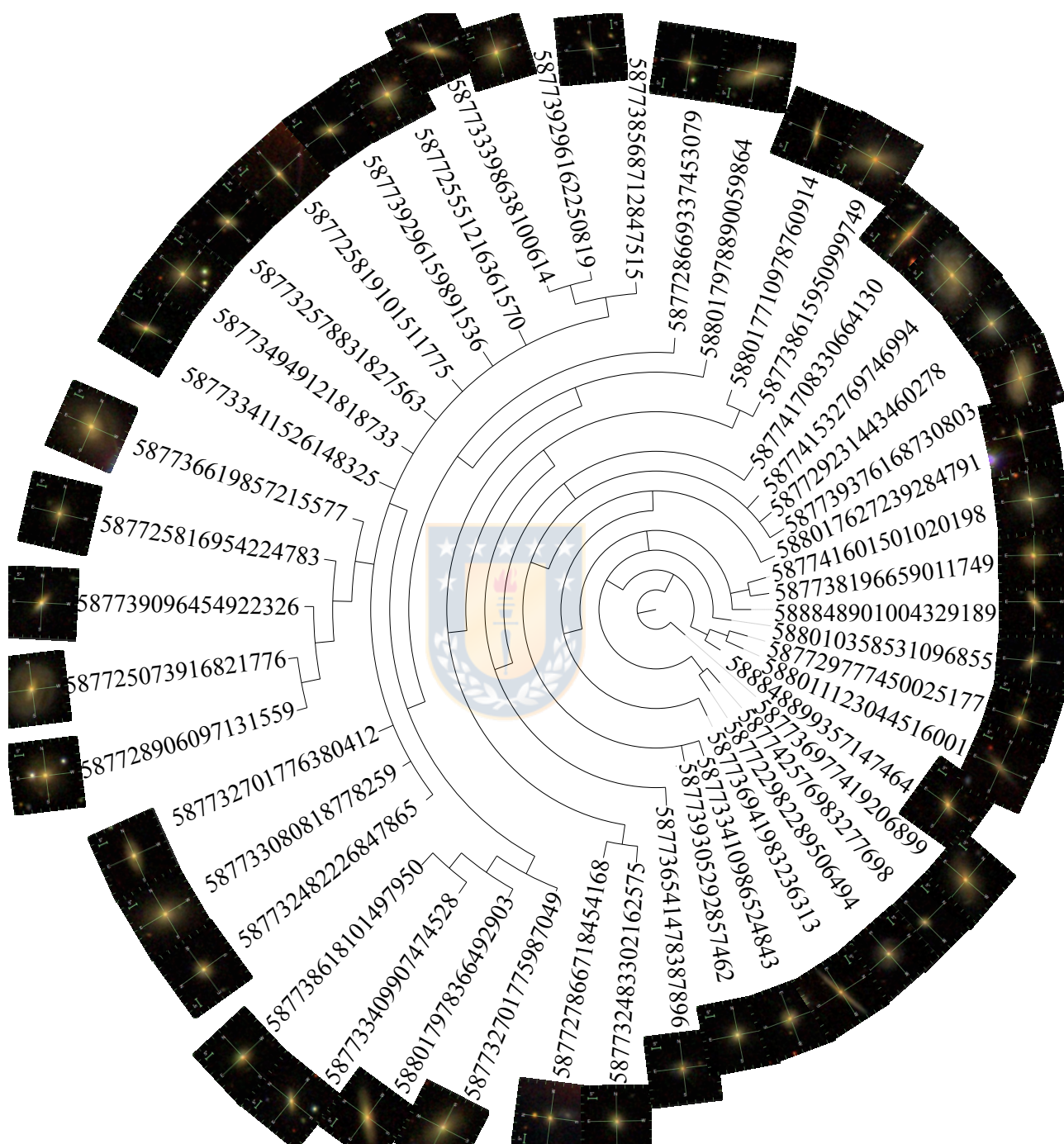


Figura 5.16: Galaxy population obtained in the field, plotted as a single tree. Each galaxy has its ID number from the SDSS DR7, and its image is from the SkyServer DR15.

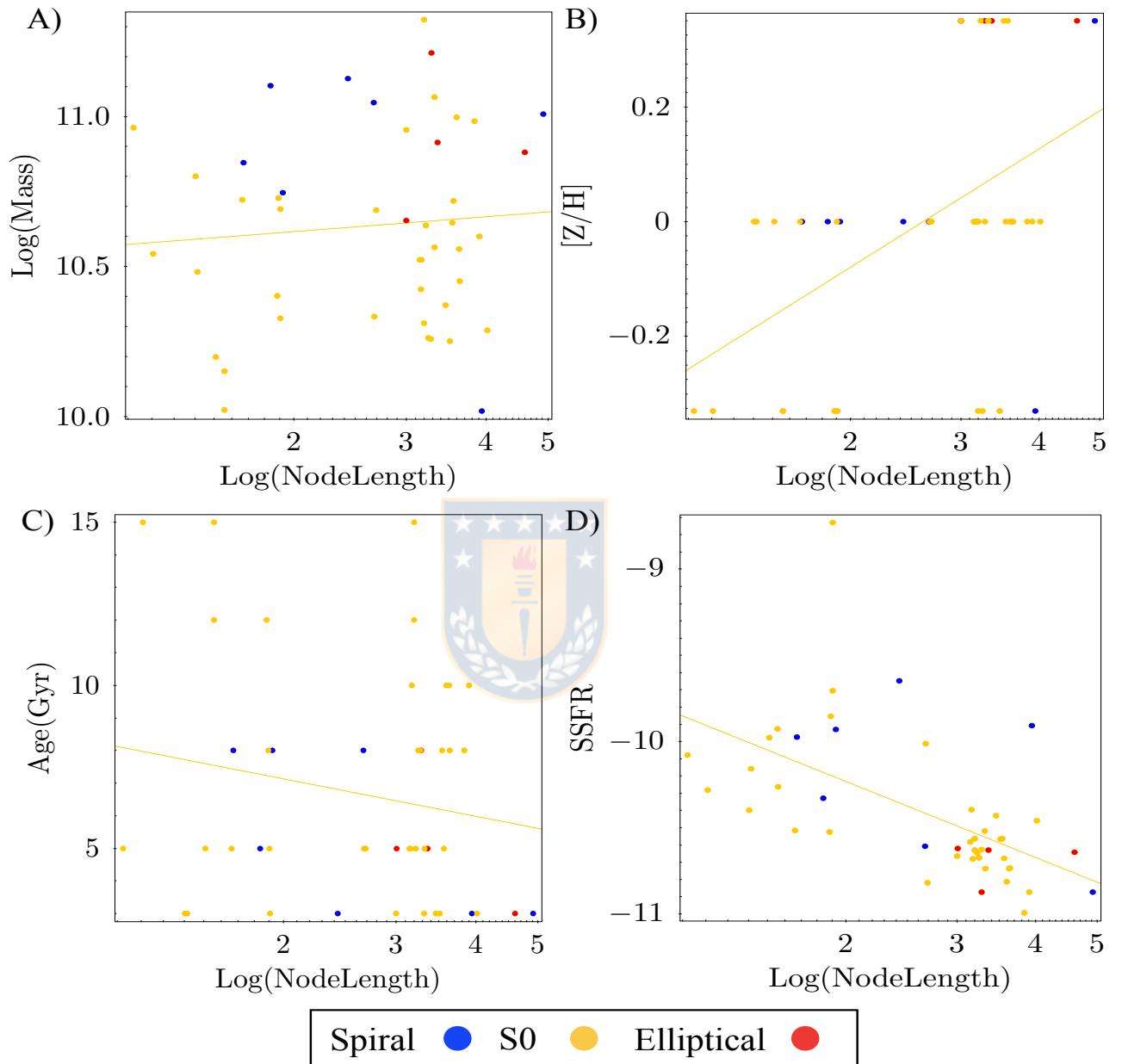


Figure 5.17: Panels A, B, C, and D show a comparison of our chemical index (length between nodes) with stellar mass, metallicity, age of the stellar population and sSFR, respectively. Red, yellow and blue colors represent the elliptical, S0 or spiral morphology, respectively, of our galaxy population in the field.

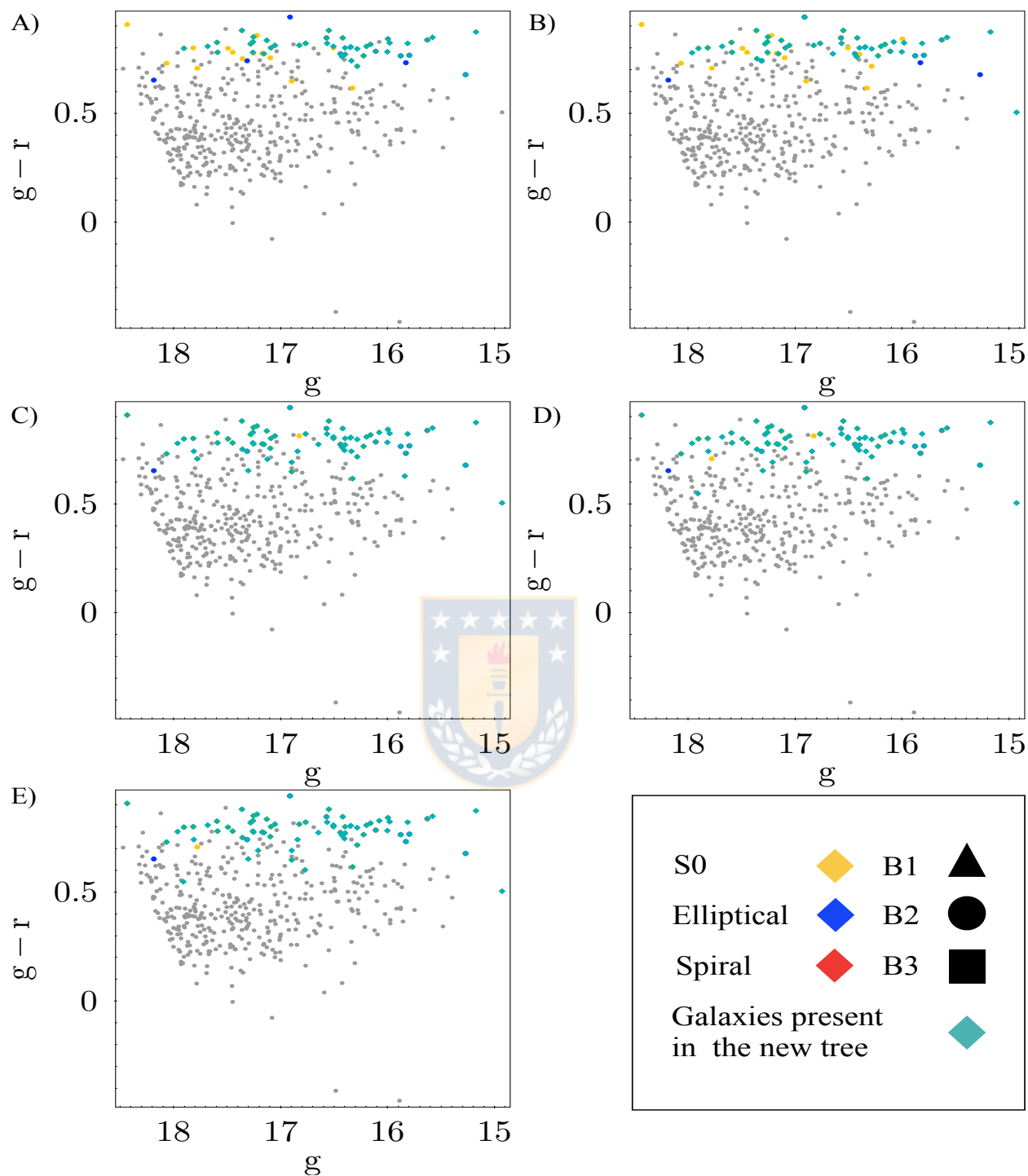


Figure 5.18: Color-magnitude diagram of our field galaxy sample, where panels A, B, C, D, and E show a recalculated phylogenetic tree excluding absorption Balmer lines, lines insensitive to α/Fe , lines with negative response to α/Fe enhancement, lines with positive response to α/Fe enhancement, and Emission Line $[OIII]\alpha 5007$, with $[OIII]_1$ and $[OIII]_2$ respectively. The cyan diamonds represent the galaxies that appear in branches of each recalculated tree.

Cuadro 5.5: Pearson and Spearman coefficients between NodeLength and specific abundance indices for each galaxy population in the field

Index	Spearman Coeff.	Index	Spearman Coeff.
HDEL_A	-0.421	FE5015	0.51
HDEL_F	-0.327	FE5270	0.483
HG_A	-0.951	FE5270S	0.244
HG_F	-0.876	FE5335	0.106
HB	-0.142	FE5406	0.174
HB_G	-0.326	FE5709	-0.135
CN_1	0.405	FE5782	0.254
CN_2	0.16	C4668	0.394
CA42270	0.92	CA4455	0.424
G4300	0.773	NA_D	0.445
MG_1	0.509	TIO_1	0.2
MG_2	0.536	TIO_2	0.437
MG_B	0.622	OIII_1	0.137
FE4383	0.445	OIII_2	0.129
FE4531	0.238		

6. Discussion

The phylogenetic analysis performed in this work, takes, as input, information on the galaxy stellar population content in the form of a pairwise distance matrix whose components are the total difference between absorption indices for a given pair of galaxies. The NJ generates hierarchical structures that retrieve the evolutionary history of galaxies. The branches of the trees represent different GP, and the NodeLength, the chemical variation from one galaxy to another within the corresponding GP or branch.

The Phylogenetic analysis of galaxies is a better tool to identify GP because it allows us to detect them regardless of the morphology, stellar mass, sSFR, metallicity or stellar population properties of the galaxies.

6.1. Comparison with other methods

Chemical tagging is the closest method to our phylogenetic approach. Therefore, we make a principal component analysis (PCA) of the same dataset of 30 abundance indices for the Coma cluster and field samples. Then we apply K-means clustering (Hartigan and Wong 1979), an unsupervised machine learning algorithm, to categorize our data into groups according to centroids. In order to obtain the centroids for the Coma cluster we define 4 groups, so we can compare the result that we have with the phylogenetic approach. Then we use the elbow method (Joshi and Nalwade 2013) to define the ideal number of groups, that in this case are 8 for both samples.

We calculate 10 principal components for the Coma cluster and the field. Figure 6.1 shows the combinations of principal components 1 and 2, and also principal components 1 and 4, which were the only combinations that showed class differentiation in both samples.

For the Coma cluster, the PCA method at a first glance is able to separate two populations, with the combination of principal components 1 and 2. We calculate the centroids for 4 groups, and 8 groups. We see that the centroids are not able to differentiate the three main GP obtained with the phylogenetic approach. On the other hand, the combination of principal components 1 and 4, shows three possible groups. Nevertheless, by calculating

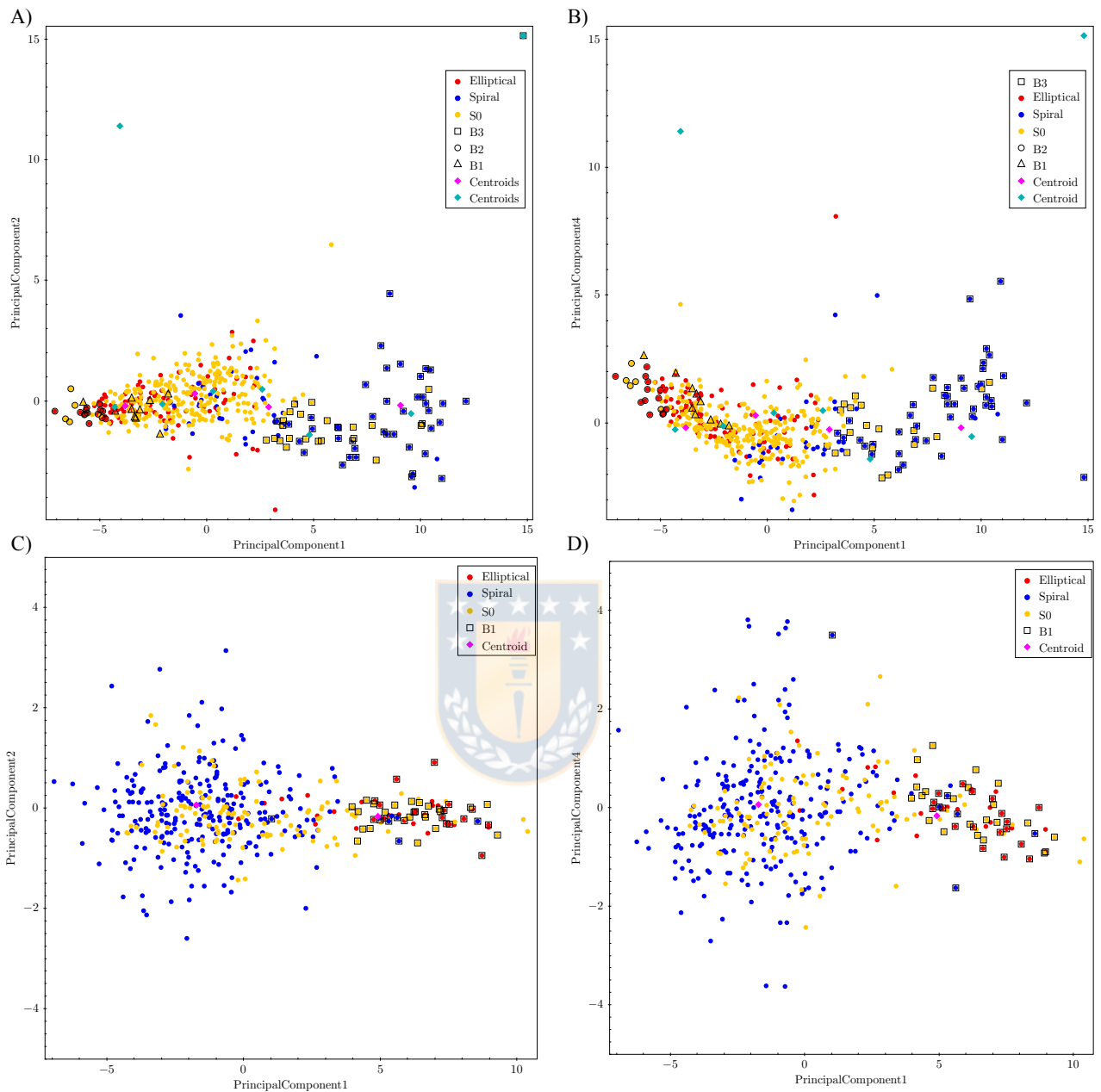


Figure 6.1: Principal component decomposition of the 30 abundance indices for our galaxy samples of the Coma cluster and the field. Figure A present the principal components 1 and 2, and B present the principal components 1 and 4 for the Coma cluster. Figure C present the principal components 1 and 2, and D present the principal components 1 and 4 for the field. Red, blue and yellow dots, indicates the elliptical, spiral and lenticular morphology of the galaxies. Magenta and cyan diamonds indicates the centroids for the PCA decomposition

the centroids for 4 and 8 groups we see that the centroids are unable to differentiate the GP obtained by our phylogenetic tree. PCA decomposition is able to separate our sample into groups, but does not recover the GP obtained from the phylogenetic approach in dense environments like the Coma cluster. On the other hand, PCA lacks the extra parameter given by the phylogenetic approach that is the length between nodes. This new parameter can be used to explore correlations between specific absorption lines in order to understand the physical processes that are occurring in our GP. Also, the determination of groups depends on a K-means analysis which needs, as an input, a predefined number of groups in order to compute the centroids that separate our GP.

6.2. Minor Structures

Our phylogenetic analysis shows the presence of GP in both the Coma cluster and the field. Nevertheless, for the Coma cluster 30 % of our galaxies can be found in minor structures and 49 % in branches without nodes. Likewise for the field 50 % of the galaxy sample are found in minor structures and 39 % in branches without nodes.

For the Coma cluster, the phylogenetic analysis shows three different GP, represented by the three main branches of the tree. The NJ algorithm, in principle, is able to retrieve evolutionary information about their input taxa. Nevertheless, the input information can be incomplete, and thus the consensus tree restricts the possibility of further exploring structures with low statistical significance.

In order to explore the minor structures of our consensus phylogenetic tree, we divided them in to single branch structures, and structures with 1, 2 or 3 nodes. We plot them in color-magnitude space and in the PCA analysis. From figure 6.2-A we can see that these minor structures belong to the red sequence of Coma cluster. Nevertheless both samples of minor structures of Coma cluster, do not show a clear disentanglement from each other. From figure 6.2-B the PCA decomposition is also unable to separate these minor structures as different groups. Therefore for the Coma cluster we can conclude that the red sequence is not homogeneous, it has small groups of galaxies in color-magnitude space belonging to small substructures that appear in our consensus phylogenetic tree. For the field sample we make the same division for minor structures and from figure 5.18

Cuadro 6.1: Pearson and Spearman coefficients between NodeLength and common properties of galaxies for each galaxy population

Branch	Pearson Coeff.	Spearman Coeff.
Length v/s Mass	0.087	0.026
Length v/s [Z/H]	0.472	0.410
Length v/s Age	-0.191	-0.132
Length v/s SSFR	-0.590	-0.636

Cuadro 6.2: Pearson and Spearman coefficients between NodeLength and specific abundance indices for each galaxy population in the field

Index	Spearman Coeff.	Index	Spearman Coeff.
HDEL_A	-0.421	FE5015	0.51
HDEL_F	-0.327	FE5270	0.483
HG_A	-0.951	FE5270S	0.244
HG_F	-0.876	FE5335	0.106
HB	-0.142	FE5406	0.174
HB_G	-0.326	FE5709	-0.135
CN_1	0.405	FE5782	0.254
CN_2	0.16	C4668	0.394
CA42270	0.92	CA4455	0.424
G4300	0.773	NA_D	0.445
MG_1	0.509	TIO_1	0.2
MG_2	0.536	TIO_2	0.437
MG_B	0.622	OIII_1	0.137
FE4383	0.445	OIII_2	0.129
FE4531	0.238		

we can see that PCA decomposition is also unable to separate these minor structures in different groups.

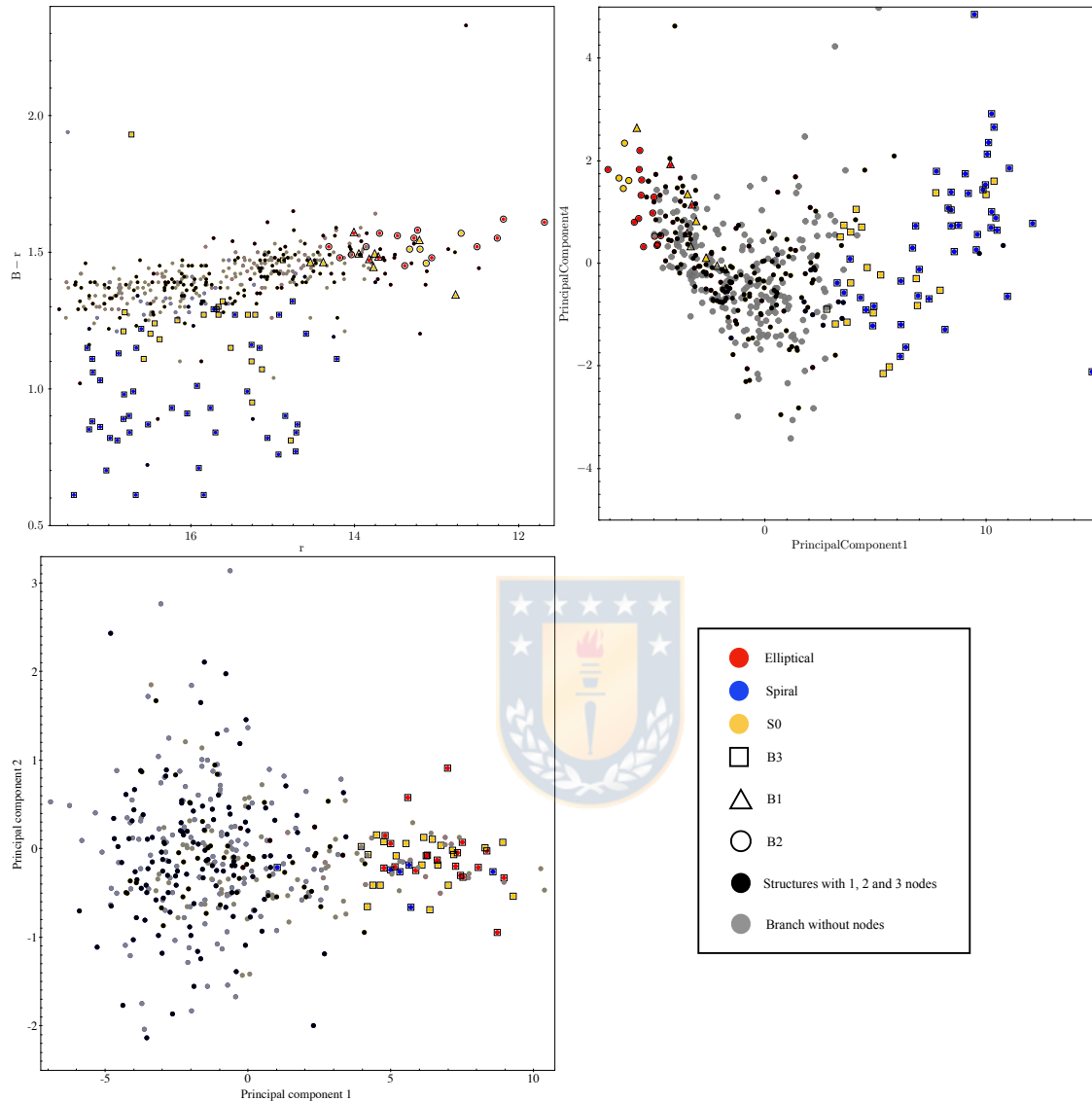


Figure 6.2: Minor structures. Image A show a Color-Magnitude diagram of the Coma Cluster with photometric filters Johnson B and Sloan r, image B show a PCA decomposition of the coma cluster and image C show a PCA decomposition of the field. Triangles, circles and squares represent galaxies belonging to branch 1 (B1), branch 2 (B2) and branch 3 (B3), respectively. The colors indicate galaxy morphology where red corresponds to elliptical galaxies, blue to spiral galaxies, and yellow to S0 galaxies. Black and grey dots indicates the structures with 1, 2 and 3 nodes and single branch respectively.

7. Conclusion

Branches in the consensus phylogenetic tree are interpreted as GPs.

The three main branches in Coma are named B1, B2, and B3. Galaxies belonging to B1 are quiescent with high stellar mass ($\log(M/M_{\odot}) = 10,35 - 11,17$), high metallicity ($[Z/H] = 0,24 - 0,59$), and low sSFR ($10^{11,25} - 10^{10,53} \text{ yr}^{-1}$). However, they are predominantly of lenticular morphology and have young-to-intermediate stellar populations (1,88–5,38 Gyr). On the other hand, galaxies belonging to B2 correspond to quiescent objects with a high stellar mass ($\log(M/M_{\odot}) = 10,62-11,76$), high metallicity ($[Z/H] = 0,21-0,49$), low sSFR ($10^{11,25}-10^{10,53} \text{ yr}^{-1}$), intermediate-to-old stellar populations ($> 3,74$ Gyr) and of mostly elliptical morphology. Galaxies in B3 have a low stellar mass ($\log(M/M_{\odot}) = 8,39 - 10,41$) with low metallicity ($[Z/H] = 1,65 - 0,55$), high sSFR ($10^{10,25}-10^{8,17} \text{ yr}^{-1}$), a predominant spiral morphology and stellar populations with an extended range of ages (0,12–10,89 Gyr). B2 is distributed in two clumps in the cluster, one in the center and the other near the galaxy NCG 4839 which is associated with a subgroup of galaxies in-falling into Coma (Neumann et al. 2001). Minor structures in the Coma cluster belong to the red sequence. Therefore, this one is not homogeneously distributed in color-magnitude space.

The galaxy population in the field is too heterogeneous for us to be able to differentiate more than one population. The field presents a main branch with galaxies that we interpret as our main GP. This one has objects of predominantly lenticular morphology, with high stellar masses ($\log(M/M_{\odot}) = 10,01 - 11,32$) and wide ranges of sSFRs ($10^{-10,99} - 10^{-8,72} \text{ yr}^{-1}$), age ($> 3,0$ Gyr) and metallicity ($[Z/H] = -0,33 - 0,34$).

Our phylogenetic approach introduces a new index that reflects the chemical distance between galaxies in a GP. This index is given by the NodeLength value and allows us to study the hierarchical structures within the tree in terms of general properties of galaxies and specific abundance indices.

The correlations between the NodeLength index and general properties of galaxies in Coma show that the population B1 correlates with stellar age, metallicity and sSFR. The correlation of this index with specific absorption indices show that the B1 population has undergone a period of star formation. B2, on the other hand, is composed by quenched galaxies, and B3 has a S0 population of galaxies that depend on the Fe4383 abundance

index, as opposed to the S0 population from B1 that does not. This shows that, in dense environments regardless of the lenticular morphology, galaxies can have chemical differences. In the field, the correlations between the NodeLength index and general properties of galaxies show a weak dependence on sSFR. No particular class of abundance indices influences the hierarchical structure of our phylogenetic tree.

From the analysis of the NodeLength values and specific absorption indices in both the field and the Coma cluster, we found that just 18 indices are relevant to find GPs with this phylogenetic approach: $H\delta_A$, $H\gamma_A$, $H\gamma_F$, $H\beta$, $H\beta_G$, CN_1 , CN_2 , Mg_1 , Mg_2 , Mg_b , $G4300$, NaD , $C4668$, TiO_2 , $Fe4383$, $Fe5270$, $Fe5270S$, $Fe5335$.

As an overall conclusion of this work, we have that phylogenetics is a promising method to identify and study GPs in different environments. This is due to the fact that it is directly based on chemical information that is determined by the stellar population content in galaxies themselves. Our results indicate that, at least in the case of the Coma cluster, galaxies in that high density region are chemically more heterogeneous than their counterparts in the field. We speculate that these differences may be a consequence of environmental processes that drive the evolution of galaxies at different local densities. The main advantage of the phylogenetic approach is that, by being exclusively based on chemical information, it allows us to perform robust analyses and interpretations without introducing biases by first selecting samples based on a specific characteristic such as stellar mass, morphology, sSFR or metallicity, as done with more traditional techniques. We caution, however, that more studies with more cluster and field samples at higher redshifts are needed to firmly establish the validity of the results obtained in this work.

References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., Bailer-Jones, C. A. L., Barentine, J. C., Bassett, B. A., Becker, A. C., Beers, T. C., Bell, E. F., Belokurov, V., Berlind, A. A., Berman, E. F., Bernardi, M., Bickerton, S. J., Bizyaev, D., Blakeslee, J. P., Blanton, M. R., Bochanski, J. J., Boroski, W. N., Brewington, H. J., Brinchmann, J., Brinkmann, J., Brunner, R. J., Budavári, T., Carey, L. N., Carliles, S., Carr, M. A., Castander, F. J., Cinabro, D., Connolly, A. J., Csabai, I., Cunha, C. E., Czarapata, P. C., Davenport, J. R. A., de Haas, E., Dilday, B., Doi, M., Eisenstein, D. J., Evans, M. L., Evans, N. W., Fan, X., Friedman, S. D., Frieman, J. A., Fukugita, M., Gänsicke, B. T., Gates, E., Gillespie, B., Gilmore, G., Gonzalez, B., Gonzalez, C. F., Grebel, E. K., Gunn, J. E., Györy, Z., Hall, P. B., Harding, P., Harris, F. H., Harvanek, M., Hawley, S. L., Hayes, J. J. E., Heckman, T. M., Hendry, J. S., Hennessy, G. S., Hindsley, R. B., Hoblitt, J., Hogan, C. J., Hogg, D. W., Holtzman, J. A., Hyde, J. B., Ichikawa, S.-i., Ichikawa, T., Im, M., Ivezić, Ž., Jester, S., Jiang, L., Johnson, J. A., Jorgensen, A. M., Jurić, M., Kent, S. M., Kessler, R., Kleinman, S. J., Knapp, G. R., Konishi, K., Kron, R. G., Krzesinski, J., Kuropatkin, N., Lampeitl, H., Lebedeva, S., Lee, M. G., Lee, Y. S., French Leger, R., Lépine, S., Li, N., Lima, M., Lin, H., Long, D. C., Loomis, C. P., Loveday, J., Lupton, R. H., Magnier, E., Malanushenko, O., Malanushenko, V., Mandelbaum, R., Margon, B., Marriner, J. P., Martínez-Delgado, D., Matsubara, T., McGehee, P. M., McKay, T. A., Meiksin, A., Morrison, H. L., Mullally, F., Munn, J. A., Murphy, T., Nash, T., Nebot, A., Neilsen, Eric H., J., Newberg, H. J., Newman, P. R., Nichol, R. C., Nicinski, T., Nieto-Santisteban, M., Nitta, A., Okamura, S., Oravetz, D. J., Ostriker, J. P., Owen, R., Padmanabhan, N., Pan, K., Park, C., Pauls, G., Peoples, John, J., Percival, W. J., Pier, J. R., Pope, A. C., Pourbaix, D., Price, P. A., Purger, N., Quinn, T., Raddick, M. J., Re Fiorentin, P., Richards, G. T., Richmond, M. W., Riess, A. G., Rix, H.-W., Rockosi, C. M., Sako, M., Schlegel, D. J., Schneider, D. P., Scholz, R.-D., Schreiber, M. R., Schwobe, A. D., Seljak, U., Sesar, B., Sheldon, E., Shimasaku, K., Sibley, V. C., Simmons, A. E., Sivarani, T., Allyn Smith, J., Smith, M. C., Smolčić, V., Snedden, S. A., Stebbins, A., Steinmetz, M., Stoughton, C., Strauss, M. A., SubbaRao, M., Suto, Y., Szalay, A. S., Szapudi, I., Szkody, P., Tanaka, M., Tegmark, M., Teodoro, L. F. A., Thakar, A. R., Tremonti, C. A., Tucker, D. L., Uomoto, A., Vanden Berk, D. E., Vandenberg, J., Vidrih, S., Vogeley, M. S., Voges, W., Vogt, N. P., Wadadekar, Y., Watters, S., Weinberg, D. H., West, A. A., White, S. D. M., Wilhite, B. C., Wonders, A. C., Yanny, B., Yocum, D. R., York, D. G., Zehavi, I., Zibetti, S., and Zucker, D. B. (2009). The Seventh Data Release of the Sloan Digital Sky Survey. , 182(2):543–558.
- Abolfathi, B., Aguado, D. S., Aguilar, G., Allende Prieto, C., Almeida, A., Ananna, T. T., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Aragón-Salamanca, A., Argudo-Fernández, M., Armengaud, E., Ata, M., Aubourg, E., Avila-Reese, V., Badenes, C., Bailey, S., Balland, C., Barger, K. A., Barrera-Ballesteros, J., Bartosz, C., Bastien, F., Bates, D., Baumgarten, F., Bautista, J., Beaton, R., Beers, T. C., Belfiore, F., Bender, C. F., Bernardi, M., Bershady, M. A., Beutler, F., Bird, J. C., Bizyaev, D., Blanc, G. A., Blanton, M. R., Blomqvist, M., Bolton, A. S., Boquien, M., Borissova, J., Bovy, J., Bradna Diaz, C. A., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Burgasser, A. J., Burtin, E., Busca, N. G., Cañas, C. I., Cano-Díaz, M.,

Cappellari, M., Carrera, R., Casey, A. R., Cervantes Sodi, B., Chen, Y., Cherinka, B., Chiappini, C., Choi, P. D., Chojnowski, D., Chuang, C.-H., Chung, H., Clerc, N., Cohen, R. E., Comerford, J. M., Comparat, J., Correa do Nascimento, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Cruz-Gonzalez, I., Cunha, K., da Silva Ilha, G., Damke, G. J., Darling, J., Davidson, James W., J., Dawson, K., de Icaza Lizaola, M. A. C., de la Macorra, A., de la Torre, S., De Lee, N., de Sainte Agathe, V., Deconto Machado, A., Dell'Agli, F., Delubac, T., Diamond-Stanic, A. M., Donor, J., Downes, J. J., Drory, N., du Mas des Bourboux, H., Duckworth, C. J., Dwelly, T., Dyer, J., Ebelke, G., Davis Eigenbrot, A., Eisenstein, D. J., Elsworth, Y. P., Emsellem, E., Eracleous, M., Erfanianfar, G., Escoffier, S., Fan, X., Fernández Alvar, E., Fernandez-Trincado, J. G., Fernand o Cirolini, R., Feuillet, D., Finoguenov, A., Fleming, S. W., Font-Ribera, A., Freischlad, G., Frinchaboy, P., Fu, H., Gómez Maqueo Chew, Y., Galbany, L., García Pérez, A. E., Garcia-Dias, R., García-Hernández, D. A., Garma Oehmichen, L. A., Gaulme, P., Gelfand, J., Gil-Marín, H., Gillespie, B. A., Goddard, D., González Hernández, J. I., Gonzalez-Perez, V., Grabowski, K., Green, P. J., Grier, C. J., Gueguen, A., Guo, H., Guy, J., Hagen, A., Hall, P., Harding, P., Hasselquist, S., Hawley, S., Hayes, C. R., Hearty, F., Hekker, S., Hernandez, J., Hernandez Toledo, H., Hogg, D. W., Holley-Bockelmann, K., Holtzman, J. A., Hou, J., Hsieh, B.-C., Hunt, J. A. S., Hutchinson, T. A., Hwang, H. S., Jimenez Angel, C. E., Johnson, J. A., Jones, A., Jönsson, H., Jullo, E., Khan, F. S., Kinemuchi, K., Kirkby, D., Kirkpatrick, Charles C., I., Kitaura, F.-S., Knapp, G. R., Kneib, J.-P., Kollmeier, J. A., Lacerna, I., Lane, R. R., Lang, D., Law, D. R., Le Goff, J.-M., Lee, Y.-B., Li, H., Li, C., Lian, J., Liang, Y., Lima, M., Lin, L., Long, D., Lucatello, S., Lundgren, B., Mackereth, J. T., MacLeod, C. L., Mahadevan, S., Maia, M. A. G., Majewski, S., Manchado, A., Maraston, C., Mariappan, V., Marques-Chaves, R., Masseron, T., Masters, K. L., McDermid, R. M., McGreer, I. D., Melendez, M., Meneses-Goytia, S., Merloni, A., Merrifield, M. R., Meszaros, S., Meza, A., Minchev, I., Minniti, D., Mueller, E.-M., Muller-Sanchez, F., Muna, D., Muñoz, R. R., Myers, A. D., Nair, P., Nandra, K., Ness, M., Newman, J. A., Nichol, R. C., Nidever, D. L., Nitschelm, C., Noterdaeme, P., O'Connell, J., Oelkers, R. J., Oravetz, A., Oravetz, D., Ortíz, E. A., Osorio, Y., Pace, Z., Padilla, N., Palanque-Delabrouille, N., Palicio, P. A., Pan, H.-A., Pan, K., Parikh, T., Pâris, I., Park, C., Peirani, S., Pellejero-Ibanez, M., Penny, S., Percival, W. J., Perez-Fournon, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Pisani, A., Prada, F., Prakash, A., Queiroz, A. B. d. A., Raddick, M. J., Raichoor, A., Barboza Rembold, S., Richstein, H., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Rodríguez Torres, S., Román-Zúñiga, C., Ross, A. J., Rossi, G., Ruan, J., Ruggeri, R., Ruiz, J., Salvato, M., Sánchez, A. G., Sánchez, S. F., Sanchez Almeida, J., Sánchez-Gallego, J. R., Santana Rojas, F. A., Santiago, B. X., Schiavon, R. P., Schimoia, J. S., Schlafly, E., Schlegel, D., Schneider, D. P., Schuster, W. J., Schwobe, A., Seo, H.-J., Serenelli, A., Shen, S., Shen, Y., Shetrone, M., Shull, M., Silva Aguirre, V., Simon, J. D., Skrutskie, M., Slosar, A., Smethurst, R., Smith, V., Soback, J., Somers, G., Souter, B. J., Souto, D., Spindler, A., Stark, D. V., Stassun, K., Steinmetz, M., Stello, D., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Suárez, G., Sun, J., Szigeti, L., Taghizadeh-Popp, M., Talbot, M. S., Tang, B., Tao, C., Tayar, J., Tembe, M., Teske, J., Thakar, A. R., Thomas, D., Tissera, P., Tojeiro, R., Tremonti, C., Troup, N. W., Urry, M., Valenzuela, O., van den Bosch, R., Vargas-González, J., Vargas-Magaña, M., Vazquez, J. A., Villanova, S., Vogt, N., Wake, D., Wang, Y., Weaver, B. A., Weijmans, A.-M., Weinberg, D. H., Westfall, K. B., Whelan, D. G., Wilcots, E., Wild, V., Williams, R. A., Wilson, J., Wood-Vasey, W. M.,

- Wylezalek, D., Xiao, T., Yan, R., Yang, M., Ybarra, J. E., Yèche, C., Zakamska, N., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, C., Zhao, G.-B., Zheng, Z., Zheng, Z., Zhou, Z.-M., Zhu, G., Zinn, J. C., and Zou, H. (2018). The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. , 235(2):42.
- Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., and Budavari, T. (2006). Galaxy bimodality versus stellar mass and environment. , 373:469–483.
- Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., and Szalay, A. S. (2004). Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. , 600:681–694.
- Beijersbergen, M., Schaap, W. E., and van der Hulst, J. M. (2002). Deprojection of luminosity functions of galaxies in the Coma cluster. , 390:817–820.
- Blanco-Cuaresma, S. and Fraix-Burnet, D. (2018). A phylogenetic approach to chemical tagging. Reassembling open cluster stars. , 618:A65.
- Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., and Brinkmann, J. (2004). The physical properties of star-forming galaxies in the low-redshift Universe. , 351(4):1151–1179.
- Burstein, D., Bertola, F., Buson, L. M., Faber, S. M., and Lauer, T. R. (1988). The Far-Ultraviolet Spectra of Early-Type Galaxies. , 328:440.
- Burstein, D., Faber, S. M., Gaskell, C. M., and Krumm, N. (1984). Old stellar populations. I. A spectroscopic comparison of galactic globular clusters, M 31 globular clusters, and elliptical galaxies. , 287:586–609.
- Cardiel, N., Gorgas, J., Sánchez-Blázquez, P., Cenarro, A. J., Pedraz, S., Bruzual, G., and Klement, J. (2003). Using spectroscopic data to disentangle stellar population properties. , 409:511–522.
- Carretta, E. (2015). Five Groups of Red Giants with Distinct Chemical Composition in the Globular Cluster NGC 2808. , 810(2):148.
- Cassata, P., Cimatti, A., Kurk, J., Rodighiero, G., Pozzetti, L., Bolzonella, M., Daddi, E., Mignoli, M., Berta, S., Dickinson, M., Franceschini, A., Halliday, C., Renzini, A., Rosati, P., and Zamorani, G. (2008). GMASS ultradeep spectroscopy of galaxies at $z \approx 2$. III. The emergence of the color bimodality at $z \approx 2$. , 483(3):L39–L42.
- Cowie, L. L., Songaila, A., Hu, E. M., and Cohen, J. G. (1996). New Insight on Galaxy Formation and Evolution From Keck Spectroscopy of the Hawaii Deep Fields. , 112:839.
- Davies, R. L., Burstein, D., Dressler, A., Faber, S. M., Lynden-Bell, D., Terlevich, R. J., and Wegner, G. (1987). Spectroscopy and Photometry of Elliptical Galaxies. II. The Spectroscopic Parameters. , 64:581.
- Davies, R. L., Sadler, E. M., and Peletier, R. F. (1993). Line-strength gradients in elliptical galaxies. *Monthly Notices of the Royal Astronomical Society*, 262(3):650–680.

- Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., and Fischer, J. L. (2018). Improving galaxy morphologies for SDSS with Deep Learning. *Monthly Notices of the Royal Astronomical Society*, 476(3):3661–3676.
- Faber, S. M. (1973). Variations in Spectral-Energy Distributions and Absorption-Line Strengths among Elliptical Galaxies. , 179:731–754.
- Faber, S. M., Friel, E. D., Burstein, D., and Gaskell, C. M. (1985). Old stellar populations. II - an analysis of K-giant spectra. , 57:711–741.
- Fisher, D., Franx, M., and Illingworth, G. (1996). Line Strengths and Line-Strength Gradients in S0 Galaxies. , 459:110.
- Fraix-Burnet, D., Choler, P., and Douzery, E. J. P. (2006). Towards a phylogenetic analysis of galaxy evolution: a case study with the dwarf galaxies of the Local Group. , 455(3):845–851.
- Fraix-Burnet, D. and Davoust, E. (2015). Stellar populations in ω Centauri: a multivariate analysis. , 450(4):3431–3441.
- Fraix-Burnet, D., Davoust, E., and Charbonnel, C. (2009). The environment of formation as a second parameter for globular cluster classification. , 398(4):1706–1714.
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., and Tremonti, C. A. (2005). The ages and metallicities of galaxies in the local universe. , 362(1):41–58.
- Gobat, R., Daddi, E., Onodera, M., Finoguenov, A., Renzini, A., Arimoto, N., Bouwens, R., Brusa, M., Chary, R.-R., Cimatti, A., Dickinson, M., Kong, X., and Mignoli, M. (2011). A mature cluster with X-ray emission at $z = 2.07$. , 526:A133.
- Gonzalez, A. H., Tran, K.-V. H., Conbere, M. N., and Zaritsky, D. (2005). Galaxy Cluster Assembly at $z=0.37$. , 624:L73–L76.
- González, J. J. (1993). *Line strength gradients and kinematic profiles in elliptical galaxies*. PhD thesis, -.
- Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.
- Hubble, E. P. (1926). Extragalactic nebulae. , 64:321–369.
- Huertas-Company, M., Rouan, D., Tasca, L., Soucail, G., and Le Fèvre, O. (2008). A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. I. Method description. , 478:971–980.
- Jofré, P., Das, P., Bertranpetit, J., and Foley, R. (2017). Cosmic phylogeny: reconstructing the chemical history of the solar neighbourhood with an evolutionary tree. , 467(1):1140–1153.
- Joshi, K. D. and Nalwade, P. S. (2013). Modified k-means for better initial cluster centres.
- Jørgensen, I. (1999). E and S0 galaxies in the central part of the Coma cluster: ages, metal abundances and dark matter. *Monthly Notices of the Royal Astronomical Society*, 306(3):607–636.

- Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Brinchmann, J., Bruzual, G., Peng, E. W., Seibert, M., Bernardi, M., Blanton, M., Brinkmann, J., Castander, F., Csábai, I., Fukugita, M., Ivezić, Z., Munn, J. A., Nichol, R. C., Padmanabhan, N., Thakar, A. R., Weinberg, D. H., and York, D. (2003). Stellar masses and star formation histories for 105 galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 341(1):33–53.
- Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Peng, E. W., Seibert, M., Brinkmann, J., Nichol, R. C., SubbaRao, M., and York, D. (2003). The dependence of star formation history and internal structure on stellar mass for 10^5 low-redshift galaxies. , 341(1):54–69.
- Kuntschner, H. (2000). The stellar populations of early-type galaxies in the Fornax cluster. *Monthly Notices of the Royal Astronomical Society*, 315(1):184–208.
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., and Torres-Peimbert, S. (1979). Reprint of 1979A&A...80..155L. Chemical composition and evolution of irregular and blue compact galaxies. , 500:145–156.
- Lidman, C., Rosati, P., Demarco, R., Nonino, M., Mainieri, V., Stanford, S. A., and Toft, S. (2004). Deep near-infrared imaging of RDCS J1252.9-2927 at $z = 1.237$. The colour-magnitude diagram. , 416:829–837.
- Muzzin, A., Wilson, G., Yee, H. K. C., Gilbank, D., Hoekstra, H., Demarco, R., Balogh, M., van Dokkum, P., Franx, M., Ellingson, E., Hicks, A., Nantais, J., Noble, A., Lacy, M., Lidman, C., Rettura, A., Surace, J., and Webb, T. (2012). The Gemini Cluster Astrophysics Spectroscopic Survey (GCLASS): The Role of Environment and Self-regulation in Galaxy Evolution at $z \sim 1$. , 746:188.
- Neumann, D. M., Arnaud, M., Gastaud, R., Aghanim, N., Lumb, D., Briel, U. G., Vestrand, W. T., Stewart, G. C., Molendi, S., and Mittaz, J. P. D. (2001). The NGC 4839 group falling into the Coma cluster observed by XMM-Newton. , 365:L74–L79.
- Oh, K., Sarzi, M., Schawinski, K., and Yi, S. K. (2011). IMPROVED AND QUALITY-ASSESSED EMISSION AND ABSORPTION LINE MEASUREMENTS IN SLOAN DIGITAL SKY SURVEY GALAXIES. *The Astrophysical Journal Supplement Series*, 195(2):13.
- Parikh, T., Thomas, D., Maraston, C., Westfall, K. B., Lian, J., Fraser-McKelvie, A., Andrews, B. H., Drory, N., and Meneses-Goytia, S. (2018). SDSS-IV MaNGA: local and global chemical abundance patterns in early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 483(3):3420–3436.
- Peng, Y.-j., Lilly, S. J., Kovač, K., Bolzonella, M., Pozzetti, L., Renzini, A., Zamorani, G., Ilbert, O., Knobel, C., Iovino, A., Maier, C., Cucciati, O., Tasca, L., Carollo, C. M., Silverman, J., Kampczyk, P., de Ravel, L., Sanders, D., Scoville, N., Contini, T., Mainieri, V., Scodreggio, M., Kneib, J.-P., Le Fèvre, O., Bardelli, S., Bongiorno, A., Caputi, K., Coppa, G., de la Torre, S., Franzetti, P., Garilli, B., Lamareille, F., Le Borgne, J.-F., Le Brun, V., Mignoli, M., Perez Montero, E., Pello, R., Ricciardelli, E., Tanaka, M., Tresse, L., Vergani, D., Welikala, N., Zucca, E., Oesch, P., Abbas, U., Barnes, L., Bordoloi, R., Bottini, D., Cappi, A., Cassata, P., Cimatti, A., Fumana, M., Hasinger, G., Koekemoer, A., Leauthaud, A., Maccagni, D., Marinoni, C., McCracken,

- H., Memeo, P., Meneux, B., Nair, P., Porciani, C., Presotto, V., and Scaramella, R. (2010). Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. , 721:193–221.
- Pérez-Carrasco, M., Cabrera-Vives, G., Martínez-Marin, M., Cerulo, P., Demarco, R., Protopapas, P., Godoy, J., and Huertas-Company, M. (2019). Multiband Galaxy Morphologies for CLASH: A Convolutional Neural Network Transferred from CANDELS. , 131(1004):108002.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425.
- Salim, S., Rich, R. M., Charlot, S., Brinchmann, J., Johnson, B. D., Schiminovich, D., Seibert, M., Mallery, R., Heckman, T. M., Forster, K., Friedman, P. G., Martin, D. C., Morrissey, P., Neff, S. G., Small, T., Wyder, T. K., Bianchi, L., Donas, J., Lee, Y.-W., Madore, B. F., Milliard, B., Szalay, A. S., Welsh, B. Y., and Yi, S. K. (2007). UV Star Formation Rates in the Local Universe. , 173(2):267–292.
- Scarlata, C., Carollo, C. M., Lilly, S., Sargent, M. T., Feldmann, R., Kampeczyk, P., Porciani, C., Koekemoer, A., Scoville, N., Kneib, J.-P., Leauthaud, A., Massey, R., Rhodes, J., Tasca, L., Capak, P., Maier, C., McCracken, H. J., Mobasher, B., Renzini, A., Taniguchi, Y., Thompson, D., Sheth, K., Ajiki, M., Aussel, H., Murayama, T., Sanders, D. B., Sasaki, S., Shioya, Y., and Takahashi, M. (2007). COSMOS Morphological Classification with the Zurich Estimator of Structural Types (ZEST) and the Evolution Since $z = 1$ of the Luminosity Function of Early, Disk, and Irregular Galaxies. , 172:406–433.
- Schawinski, K., Thomas, D., Sarzi, M., Maraston, C., Kaviraj, S., Joo, S.-J., Yi, S. K., and Silk, J. (2007). Observational evidence for AGN feedback in early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 382(4):1415–1431.
- Schiavon, R. P., Rose, J. A., Courteau, S., and MacArthur, L. A. (2004). The Identification of Blue Horizontal-Branch Stars in the Integrated Spectra of Globular Clusters. , 608(1):L33–L36.
- Snyder, G. F., Brodwin, M., Mancone, C. M., Zeimann, G. R., Stanford, S. A., Gonzalez, A. H., Stern, D., Eisenhardt, P. R. M., Brown, M. J. I., Dey, A., Jannuzi, B., and Perlmutter, S. (2012). Assembly of the Red Sequence in Infrared-selected Galaxy Clusters from the IRAC Shallow Cluster Survey. , 756:114.
- Studier, J. A. and Keppler, K. J. (1988). A note on the neighbor-joining algorithm of Saitou and Nei. *Molecular Biology and Evolution*, 5(6):729–731.
- Tanaka, M., De Breuck, C., Venemans, B., and Kurk, J. (2010). The environmental dependence of galaxy properties at $z = 2$. , 518:A18.
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., Bland-Hawthorn, J., Brown, M. J. I., Colless, M., Driver, S., Norberg, P., Robotham, A. S. G., Alpaslan, M., Brough, S., Cluver, M. E., Gunawardhana, M., Kelvin, L. S., Liske, J., Conselice, C. J., Croom, S., Foster, C., Jarrett, T. H., Lara-Lopez, M., and Loveday, J. (2015). Galaxy And Mass Assembly (GAMA): deconstructing bimodality - I. Red ones and blue ones. , 446(2):2144–2185.
- Tempel, E., Stoica, R. S., Martínez, V. J., Liivamägi, L. J., Castellan, G., and Saar, E.

- (2014). Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. *Monthly Notices of the Royal Astronomical Society*, 438(4):3465–3482.
- Tempel, E., Tago, E., and Liivamägi, L. J. (2012). Groups and clusters of galaxies in the SDSS DR8. Value-added catalogues. , 540:A106.
- Terlevich, R. and Melnick, J. (1981). The dynamics and chemical composition of giant extragalactic H II regions. *Monthly Notices of the Royal Astronomical Society*, 195(4):839–851.
- Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., and Silk, J. (2010a). Environment and self-regulation in galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 404(4):1775–1789.
- Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., and Silk, J. (2010b). Environment and self-regulation in galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 404(4):1775–1789.
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S. D. M., Seibert, M., Peng, E. W., Schlegel, D. J., Uomoto, A., Fukugita, M., and Brinkmann, J. (2004). The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. , 613(2):898–913.
- van Dokkum, P. G. (2001). Cosmic-Ray Rejection by Laplacian Edge Detection. , 113:1420–1427.
- Worthey, G. (1994). Comprehensive stellar population models and the disentanglement of age and metallicity effects. , 95:107–149.
- Worthey, G. and Ottaviani, D. L. (1997). H γ and H δ Absorption Features in Stars and Stellar Populations. , 111:377–386.
- York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W. N., Bracker, S., Briegel, C., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M. A., Castander, F. J., Chen, B., Colestock, P. L., Connolly, A. J., Crocker, J. H., Csabai, I., Czarapata, P. C., Davis, J. E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B. R., Evans, M. L., Fan, X., Federwitz, G. R., Fiscelli, L., Friedman, S., Frieman, J. A., Fukugita, M., Gillespie, B., Gunn, J. E., Gurbani, V. K., de Haas, E., Haldeman, M., Harris, F. H., Hayes, J., Heckman, T. M., Hennessy, G. S., Hindsley, R. B., Holm, S., Holmgren, D. J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R. S. J., Kinney, E., Klaene, M., Kleinman, A. N., Kleinman, S., Knapp, G. R., Korienek, J., Kron, R. G., Kunszt, P. Z., Lamb, D. Q., Lee, B., Leger, R. F., Limmongkol, S., Lindenmeyer, C., Long, D. C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R. H., MacKinnon, B., Mannery, E. J., Mantsch, P. M., Margon, B., McGehee, P., McKay, T. A., Meiksin, A., Merelli, A., Monet, D. G., Munn, J. A., Narayanan, V. K., Nash, T., Neilsen, E., Neswold, R., Newberg, H. J., Nichol, R. C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J. P., Owen, R., Pauls, A. G., Peoples, J., Peterson, R. L., Petravick, D., Pier, J. R., Pope, A., Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T. R., Richards, G. T., Richmond, M. W., Rivetta, C. H., Rockosi, C. M., Ruthmansdorfer, K., Sandford, D., Schlegel, D. J., Schneider, D. P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W. A.,

Smee, S., Smith, J. A., Snedden, S., Stone, R., Stoughton, C., Strauss, M. A., Stubbs, C., SubbaRao, M., Szalay, A. S., Szapudi, I., Szokoly, G. P., Thakar, A. R., Tremonti, C., Tucker, D. L., Uomoto, A., Vanden Berk, D., Vogeley, M. S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D. H., Yanny, B., Yasuda, N., and SDSS Collaboration (2000). The Sloan Digital Sky Survey: Technical Summary. , 120:1579–1587.

