

*Annual Review of Astronomy and Astrophysics*Multiple Stellar Populations
in Globular ClustersNate Bastian¹ and Carmela Lardo^{1,2}¹Astrophysics Research Institute, Liverpool John Moores University, Liverpool, L3 5RF, United Kingdom; email: N.J.Bastian@ljmu.ac.uk²Laboratoire d'Astrophysique, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland; email: carmela.lardo@epfl.ch

Annu. Rev. Astron. Astrophys. 2018. 56:83–136

First published as a Review in Advance on May 23, 2018

The *Annual Review of Astronomy and Astrophysics* is online at astro.annualreviews.org<https://doi.org/10.1146/annurev-astro-081817-051839>Copyright © 2018 by Annual Reviews.
All rights reservedANNUAL
REVIEWS CONNECT

- 
- www.annualreviews.org
 - Download figures
 - Navigate cited references
 - Keyword search
 - Explore related articles
 - Share via email or social media

Keywords

stellar abundances, color-magnitude diagram

Abstract

Globular clusters (GCs) exhibit star-to-star variations in specific elements (e.g., He, C, N, O, Na, Al) that bear the hallmark of high-temperature H-burning. These abundance variations can be observed spectroscopically and also photometrically, with the appropriate choice of filters, due to the changing of spectral features within the band pass. This phenomenon is observed in nearly all of the ancient GCs, although, to date, it has not been found in any massive cluster younger than 2 Gyr. Many scenarios have been suggested to explain this phenomenon, with most invoking multiple epochs of star formation within the cluster; however, all have failed to reproduce various key observations, in particular when a global view of the GC population is taken. We review the state of current observations and outline the successes and failures of each of the main proposed models. The traditional idea of using the stellar ejecta from a first generation of stars to form a second generation of stars, while conceptually straightforward, has failed to reproduce an increasing number of observational constraints. We conclude that the puzzle of multiple populations remains unsolved, hence alternative theories are needed.

Contents

1. INTRODUCTION	85
2. OBSERVATIONS OF ABUNDANCE VARIATIONS AND COLOR-MAGNITUDE DIAGRAMS	86
2.1. Abundance Variations	86
2.2. Multiple Populations as Seen Through Color-Magnitude Diagrams	93
2.3. Are There Single Population Globular Clusters?	96
2.4. Global Properties and Correlations	97
2.5. The Role of Cluster Age and Mass	98
2.6. Observational Summary of Multiple Populations	102
3. NUCLEOSYNTHESIS AND MULTIPLE POPULATIONS	103
3.1. Massive Stars	103
3.2. Very Massive Stars	105
3.3. Asymptotic Giant Branch Stars	105
4. THEORIES FOR THE ORIGIN OF MULTIPLE POPULATIONS	107
4.1. The Asymptotic Giant Branch Scenario	107
4.2. Fast-Rotating Massive Stars and Interacting Binaries	110
4.3. The Early Disc Accretion Scenario	111
4.4. Turbulent Separation of Elements During Globular Cluster Formation	112
4.5. Reverse Population Order for Globular Cluster Formation Scenarios	112
4.6. Extended Cluster Formation Event	113
4.7. Very Massive Stars Due to Runaway Collisions	114
5. COMPARING PREDICTIONS WITH OBSERVATIONS	114
5.1. Chemical Abundance Patterns	114
5.2. Discrete Versus Continuous Abundance Spreads	117
5.3. Radial Distributions, Velocity Dispersions, and Binarity	117
5.4. The Mass-Budget Problem	118
5.5. Trends with Cluster Properties	120
5.6. Constraints from Young Massive Clusters	120
5.7. Summary Points of the Comparison Between the Predictions and Observations	122
6. PECULIAR CLUSTERS: IRON SPREADS, CNO, AND S-PROCESS VARIATIONS	123
6.1. Clusters with Multimodal Metallicity Distributions: ω Centauri, M54, and Terzan 5	123
6.2. Clusters with Small Unimodal Iron Spreads and S-Process Bimodality	124
6.3. The Blue Tilt in Cluster Populations	125
7. YOUNG MASSIVE CLUSTERS AND THEIR RELATIONSHIP TO GLOBULAR CLUSTERS	125
7.1. Extended Main Sequence Turnoffs in Young and Intermediate-Age Clusters ..	126
7.2. Split Main Sequences	126
7.3. Chemical Anomalies in Young Massive Clusters?	127
8. MULTIPLE POPULATIONS ON GALAXY SCALES	128

1. INTRODUCTION

The traditional concept of globular clusters (GCs) as simple stellar populations, where all stars share the same age and abundances within some small tolerance, is now a view of the past; it has become clear that (nearly) all GCs host significant abundance spreads within them. Although all GCs show the same basic pattern, i.e., enriched populations in He, N, and Na and populations depleted in O and C, the specifics of each cluster are unique. It is the manifestations of these distinctive chemical anomalies that cause the impressively complex color-magnitude diagrams (CMDs) that have been uncovered with precision *Hubble Space Telescope* (HST) photometry, especially when viewed in the UV and near-UV. These star-to-star abundance variations within clusters are known as multiple populations (MPs).

GCs: globular clusters
MPs: multiple populations

The past decade has seen an impressive amount of observational work on the topic, with ground-based spectroscopic surveys of thousands of stars within samples of GCs tracing the detailed abundance patterns (e.g., Carretta et al. 2009a), and space-based photometry providing unprecedented views of the number and make-up of the different populations within the GCs (e.g., Piotto et al. 2015). In addition to these observational advances, a number of scenarios for the origin of MPs have been put forward, which have begun providing testable predictions. Alongside the coformation/evolution of GC populations in galaxies, the origin of MPs is one of the major unsolved problems in GC and stellar populations research.

The goal of this review is to provide an overview of the present state of observations of MPs along with a critical comparison against theoretical models that have been put forward for their origin. We focus the majority of our attention on results obtained since the last *Annual Review* article on the topic (Gratton et al. 2004) and refer the interested reader to that comprehensive review for the historical developments and status of the field up until that time. Additionally, there have been a number of more recent excellent reviews on the topic, notably Gratton et al. (2012a) and Charbonnel (2016). The field has been growing at a rapid rate, with hundreds of relevant papers published each year, and as such, we are unable to reference all work in the field. Instead, we use typical examples to illustrate broader points and attempt to synthesize all results into a coherent status update of the field.

Although many of the previous reviews have concentrated on the chemistry of MPs, we explore that as only one line of evidence and also consider global properties and correlations, relation to field stars, and the physical properties of both young and old massive clusters.

We define MPs as the presence of star-to-star variations in chemical abundances, which is not expected from stellar evolutionary processes. In particular, as is reviewed below, this means variations in light elements such as He, C, N, O (that can cause complexities in CMDs), Na, Al, and in some cases Mg. This can be contrasted with observations of some young (<2 Gyr) clusters that show unexpected features in their CMDs (e.g., extended main sequence turnoffs or split main sequences), which are not caused by abundance variations but are rather driven by stellar evolutionary processes (i.e., rotationally induced stellar structure changes).

Finally, a note about terminology. Stars within GCs that show enhancements in He, N, and Na and are depleted in O and C have various labels in the literature, e.g., “2nd generation stars,” “2nd population,” “enriched stars.”¹ We chose to use the latter options, as “2nd generation” implies a genetic link to a first population. While such a link is possible, it is by no means established (in fact evidence is currently pointing away from this interpretation), hence the use of more neutral terminology is more natural as the origin of MPs is still unknown. However, when referring to

¹We use the term “enriched” in the “chemical enrichment” sense, meaning that the material appears to be processed through nuclear reactions in stars. We note that some elements are in fact depleted (e.g., O, C).

RGB:

red giant branch

Enriched or second population (2P) star:

a star showing enhanced N, Na, and Al and depleted C and O abundances with respect to the field at the same metallicity [Fe/H]

Primordial or first population (1P) star:

a star having the same abundances as the field at the same metallicity [Fe/H]

MS: main sequence**MSTO:**

main sequence turnoff

models that explicitly invoke multiple generations of stars, we use the “generation” label for clarity. Also, we use “correlation” to refer to a positive correlation between two or more elements, and “anticorrelation” for a negative correlation between abundances.

2. OBSERVATIONS OF ABUNDANCE VARIATIONS AND COLOR-MAGNITUDE DIAGRAMS

2.1. Abundance Variations

MPs with distinctive light element abundance patterns are widely observed in old and massive clusters. Abundance spreads are only rarely associated with star-to-star Fe and heavy element variations, implying that some unique chemical enrichment mechanism, operating only in cluster environments, is responsible for the observed chemistry. The suggestion that the light element anomalies arise from nuclear processing within massive stars from a previous generation born within GCs still remains the only theory that has been quantitatively investigated. Nonetheless, such a hypothesis suffers from several drawbacks and can only account for some of the relevant observations. In the following, we review the status of observations and critically discuss their interpretation in the framework of MPs.

2.1.1. Light element abundance spreads. The presence of chemical inhomogeneities among bright giants in clusters was revealed by pivotal studies in the early 1970s (e.g., Osborn 1971). Stars at the same magnitude along the red giant branch (RGB) were found to display variations in the strengths of CH, CN, and NH blue absorption features, due to underlying star-to-star variations in C and N abundances (Bell & Dickens 1980).² Most of the studied GCs display either a bimodal or multimodal CN distribution (e.g., Norris 1987). The molecular CN (NH) and CH bands were found also to be anticorrelated, with CN-strong stars also characterized by weak CH absorption and vice versa; i.e., N is found to anticorrelate with C.

Although extremely common in clusters, stars characterized by enhanced N and depleted C are rarely found in the field and not present in open clusters (OCs; e.g., Martell et al. 2011, MacLean et al. 2015). However, GCs also contain stars that are characterized by the same abundance pattern observed in field stars of the same metallicity. This has led to the notion that GCs are made up of MPs, one with field-like composition and a second with “anomalous chemistry” unique to GCs. In the following, we refer to the stars with peculiar chemical composition as enriched or second population (2P) and the stars having field-like abundances as primordial or first population (1P). We consider enriched or 2P and primordial or 1P as synonyms, and we use the expressions interchangeably throughout this review.

Evolutionary mixing was originally proposed as the main cause of the C and N inhomogeneities as normal stellar evolution may contribute to the observed N-C anticorrelation in evolved RGB stars (e.g., Denisenkov & Denisenkova 1990). However, such an evolutionary scenario was soon challenged by observations (e.g., Gratton et al. 2004), as mixing theories cannot explain the abundance anomalies seen among nonevolved or scarcely evolved main sequence (MS) and main sequence turnoff (MSTO) stars (e.g., Cannon et al. 1998, Briley et al. 2004), which are characterized by negligible outer convective zones. Even if sufficient mixing could be achieved during MS evolution, it would also result in changes in helium abundances and extended lifetime of stars, e.g., mixing would result in broadening the MSTO region in the CMD, contrary to what is observed (in ancient GCs).

²In a first approximation, the CH absorption at 4300 Å can be regarded as a C sensitive diagnostic, whereas the CN and NH band strengths (at 3839 and 3360 Å, respectively) are proxies for N.

When higher-resolution spectra allowed for direct spectroscopic measurements of Na and O (through atomic lines) in stars where N and C abundances were available, it was found that the N overabundance (C depletion) was associated to enhanced Na (O depletion); i.e., N-Na and C-O are positively correlated (e.g., Sneden et al. 1992). Also, while the individual abundances of C, N, and O show large spreads, the sum C+N+O is generally observed to be constant (e.g., Dickens et al. 1991; see also Section 6.2). Anticorrelated Na and O ranges were found in nearly all the studied clusters, along with variations in Al and (possibly) Mg, which were anticorrelated with each other (e.g., Gratton et al. 2004, 2012a). While O can potentially be depleted in the interiors of low-mass stars through the CNO-cycle reactions, variations in the abundances of heavier elements like Na, Al, and Mg cannot be produced by fusion reactions within low-mass stars. This is because their temperatures are too low for the p-capture reactions to operate through the NeNa- and MgAl-chains (e.g., Prantzos et al. 2007, 2017). Hence, the abundance anomalies are not produced in the course of the evolution of stars we are currently observing, but they were produced elsewhere, potentially within the interiors of more massive stars. See **Figure 1b** for the well-studied cluster NGC 6752, which illustrates some of the typical element (anti)correlations

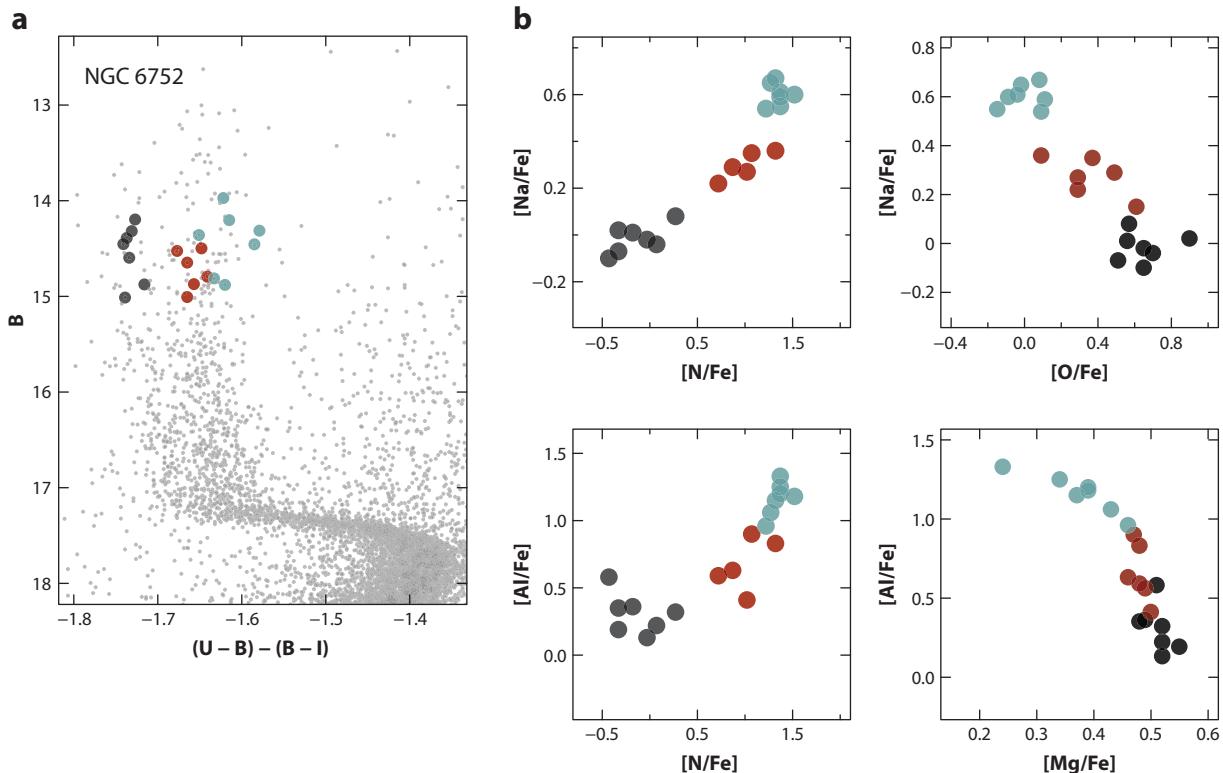


Figure 1

(a) NGC 6752 $C_{U,B,I}$ versus B CMD. Photometry has been kindly provided by Peter Stetson. Spectroscopic targets from Yong et al. (2005, 2008) are also plotted. Colors correspond to a different chemical composition, with green, red, and black symbols having high, moderate, and primordial Na content, respectively. Stars with different light-element composition, which are well mixed along the red giant branch in optical colors, occupy distinct sequences in the $C_{U,B,I}$ versus B CMD. (b) The same stars as in panel *a* are plotted to show light abundance variations. NGC 6752 is representative of many clusters with measured abundance variations. Similar plots have been shown by various authors for this and other clusters using different element ratios and/or different photometric indices. Abbreviation: CMD, color-magnitude diagram.

First generation stars

(1G): in models of MP formation, stars of the first generation

AGB: asymptotic giant branch

FRMS: fast-rotating massive star

VMS: very massive star ($\geq 5,000 M_{\odot}$)

HB: horizontal branch

associated with MPs and their relationship to color variations on the giant branch. Similar plots exist in the literature for this and other clusters using different element ratios and/or different photometric indices.

How this material would then find its way into the low-mass stars observed today is still an open question, as is the exact source of the material. Most models to date have adopted a scenario in which material from first generation stars (1G) pollutes the intracluster medium out of which subsequent generations of stars were born. Several candidate 1P polluters—either intermediate-mass asymptotic giant branch stars (AGB stars, also referred to as AGBs; $3-9 M_{\odot}$), fast-rotating massive stars (FRMSs; $> 15 M_{\odot}$), or very massive stars (VMSs; $\geq 5,000 M_{\odot}$)—have been proposed because they are sites of hot CNO and NeNa processing, and we discuss them in Section 3. A (weak) Si-Mg anticorrelation was observed in a small number of massive and/or metal-poor GCs (e.g., NGC 6752, NGC 2808, M15; Yong et al. 2005, Carretta et al. 2009a), implying that proton burning is occurring in even hotter environments (≥ 75 MK) than that needed for the CNO and NeNa processing.

The presence of anticorrelated CNONaAl abundances has been demonstrated to be nearly universal among old and massive clusters and has even been suggested to be the distinguishing feature between genuine GCs and other stellar associations (e.g., OCs or dwarf galaxies; Carretta et al. 2010b). If stars with high N, Na, or Al abundances are found in the field, they are usually considered to have originated from GCs (unless they are part of a binary system). Spectroscopic studies have estimated that at least 3% of the local-field metal-poor star population was born in GCs (e.g., Carretta et al. 2010b, Martell et al. 2011), under the assumption that all 2P stars must form in GCs.

The shape and the extension of the light element anticorrelations (i.e., their extrema, substructure, and multimodality) vary from cluster to cluster, with some clusters showing both a well-extended Na-O anticorrelation and objects for which both Na and O abundances span very short ranges (e.g., Carretta et al. 2009a,b). In a few cases, the Na-O distribution is clumpy, with the presence of one or more gaps (e.g., Marino et al. 2008, Lind et al. 2011, Carretta 2015). Such quantized distributions may be common, but measurements with very small uncertainties are needed to corroborate this. However, such a multimodality of the blue CN band is nearly universal in metal-rich clusters ($[Fe/H] \geq -1.7$ dex), where errors on CN measurements are small enough to reveal discrete distributions (Norris 1987).

The light element variations span similar intervals in different evolutionary phases (e.g., Gratton et al. 2012b). Observations show that unevolved stars on the MS and evolved RGB stars span the same ranges of chemical anomalies and demonstrate that such light element variations cannot be due to accretion of processed material on already formed stars, as the anticorrelations would be strongly diluted by mixing as the stars evolve (e.g., Gratton et al. 2004). Also, the ratio between 1P and 2P stars along the AGB appears to be consistent with the corresponding ratio found on the RGB and the observed horizontal branch (HB) morphology (e.g., Cassisi et al. 2014, Lapenna et al. 2016, Lardo et al. 2017b).

An Al-Mg anticorrelation is not observed in all the GCs in which the Na-O and N-C variations are detected. There are clusters that are characterized by a single Al abundance, whereas others show wide Al ranges (Carretta et al. 2009a, Mészáros et al. 2015). The majority of the Milky Way (MW) GC stars for which Mg abundances are available have typical Mg abundances in the range of $0.2 \leq [Mg/Fe] \leq 0.5$ dex, implying a very short (if any) Al-Mg anticorrelation. Only a few Galactic GCs have been found to host stars that are significantly deficient in Mg ($[Mg/Fe] \leq 0.0$ dex; e.g., Mucciarelli et al. 2012, Carretta 2014). The extent of the Al-Mg anticorrelation correlates with both cluster mass and metallicity, as massive and metal-poor clusters tend to have larger Al-Mg anticorrelations (e.g., Carretta et al. 2009a,b; Pancino et al. 2017).

Although the N-C and Na-O (and in some cases the Al-Mg) anticorrelations and photometric spreads along the RGBs (see Section 2.2) are distinctive signatures present in (nearly) all ancient GCs, the cluster-to-cluster differences are large in terms of the extreme values, substructure, and multimodality. The evidence that each surveyed GC has its own specific pattern of MPs calls for a high degree of variety (or stochasticity) that must be taken into account when proposing MP formation mechanisms (e.g., Bastian et al. 2015).

To date, there have only been a few stars in a handful of GCs that have been fully characterized in terms of their chemistry (i.e., the full set of varying elements: C, N, O, Na, Al, Mg, He, s-process, etc.; e.g., Smith 2015). Instead, different surveys have focused on different elements, and often even different stars within the same GCs. This is an obvious avenue for future studies, to characterize the exact chemical fingerprint of 1P and enriched 2P stars. We refer the interested reader to the compilation by Roediger et al. (2014) of abundances for a number of elements for stars in GCs.

2.1.2. He variations: main sequence splitting, horizontal branch morphology, and direct measurements. If the CNONaAlMg star-to-star variations are built in the stellar interiors through CNO cycling and p-capture processes at high temperatures, we may also expect He variations (as it is the main product of H-burning). The observational data suggest that N and Na variations are always correlated with some (variable) He enhancement. However, this result is mostly based on indirect evidence as only a handful of studies have provided direct He abundance determinations.³

He enhancement can be inferred from (*a*) direct measurements of He abundances, (*b*) splits or spreads of the MS in optical CMDs, and (*c*) the HB morphology of the clusters. In what follows, we refer to the He mass fraction as Y and denote variations in He as $\Delta Y = Y - Y_p$, where Y_p represents the initial He mass fraction value of $Y_p = 0.244$ (e.g., Cassisi et al. 2003).

Direct Y measurements are difficult to obtain. Temperatures above $T > 8,500$ K are necessary to detect the He photospheric transitions in the optical band. However, hot HB stars—where the He line might appear because of their high temperatures—are also affected by diffusion and preferential settling of elements (Behr 2003). As a result, Y can only be measured in stars with temperatures between $\sim 8,500$ and $11,500$ K, which are hot enough to show the He line but still cooler than the Grundahl jump (e.g., Moehler et al. 2014), the temperature limit above which the original surface abundances are changed by diffusion (Grundahl et al. 1999). Nonetheless, Y measurements from the photospheric HeI line at 5875 Å in HB stars have been obtained for some GCs, and variations have been reported, with typical spreads of $\Delta Y = 0.02\text{--}0.05$ (see Mucciarelli et al. 2014b for a summary). He-rich stars also have been shown to be Na rich, and they are systematically located toward the blue regions of the HB (Villanova et al. 2009).

For FGK-type stars, no photospheric lines exist, and He can only be measured from the purely chromospheric HeI absorption line at 10830 Å. Studies based on this near-IR transition confirm that He enrichment generally correlates with [Na/Fe] and [Al/Fe]. Dupree & Avrett (2013) directly measured He abundances from the 10830 Å line in two giants in ω Centauri. They estimate a He abundance of $Y \leq 0.22$ (below the Big Bang nucleosynthesis value) for the 1P star and $Y = 0.39\text{--}0.44$ for the 2P one, with the He-rich star also enhanced in Al. Similarly, Pasquini et al. (2011) performed a differential analysis between two giant stars of NGC 2808 with different

³Direct measurements of elements like He, O, Na, and Al require high resolution, thus they are limited to the brighter stars in GCs. Conversely, both N and C are generally measured in fainter stars at the base of the RGB, because evolutionary mixing as the star evolves along the RGB can blur the MP chemical signature. Hence abundance determinations for the whole set of elements that vary in GCs are available only for a few stars in a handful of clusters.

Na abundances. They estimated that the 2P star is more He enriched than the Na-poor one by $\Delta Y = 0.17$.

Although the direct spectroscopic evidence of He enhancement is somewhat sparse, several photometric studies provided evidence that such He variations are in place (e.g., Maeder & Meynet 2006, Anderson et al. 2009, Bellini et al. 2013, Nardiello et al. 2015). Photometric estimates of ΔY can be derived by assuming that the observed color dispersions at a given magnitude on the MS in optical colors (i.e., V-I) are due primarily to He spreads. The measure of ΔY spreads from MS isochrone fitting presently appears to be the most reliable method for inferring He dispersions (see Cassisi et al. 2017 and Section 2.2), and recent results from HST photometry reveal that the observed He spreads ΔY strongly correlate with present-day cluster mass and luminosity, with more massive clusters having larger He spreads (e.g., Milone 2015, which is discussed in detail in Section 2.5).

In ω Cen, the presence of a split MS (e.g., Bellini et al. 2010) has been interpreted in terms of a large variation in the abundance of He ($\Delta Y \sim 0.15$; e.g., King et al. 2012). The observation that the bluer MS is also $\simeq 0.3$ dex more metal rich than the redder MS further supports the existence of such large He enhancement, as canonical stellar models would predict the bluer MS to be more metal poor than the red one and only a high He value can explain the color difference between the two MSs (Piotto et al. 2005).

Large He variations are also observed in clusters with homogeneous iron content, as in NGC 2808 where three distinct MSs can be clearly identified in optical CMDs (e.g., Piotto et al. 2007). Given the lack of an iron spread (e.g., Carretta et al. 2006), the MS split is interpreted as being due to three groups of stars with different He (Milone et al. 2015b), which are likely linked to the multimodal HB structure (D'Antona et al. 2005, Dalessandro et al. 2011) and the three chemically distinct groups observed along the RGB (Carretta 2014). In NGC 2808, such He variations are also correlated with light element abundance spreads, in the sense that stars with 1P composition are associated to the red MS with primordial He content, whereas stars with high N, Na, and Al are located on the He-rich, blue MS (Bragaglia et al. 2010b). Variations of He between 1P and 2P stellar groups may also affect the color and luminosity of the RGB bump, as shown in Bragaglia et al. (2010a).

Variations in the abundance of He can have a significant impact on the HB morphology (Rood 1973, D'Antona et al. 2002). This is because He-rich stars evolve faster than those with primordial He and thus, at a given age, He-rich stars at the MSTO are less massive (e.g., Chantereau et al. 2016). Hence, if both He-rich and He-poor stars experience the same mass loss during RGB evolution, they should end up on the HB as stars with different masses, i.e., different colors (see also Norris et al. 1981). Indeed, the HB morphology of several clusters has been modeled in terms of variable He (e.g., Caloi & D'Antona 2007, Cassisi et al. 2009, D'Antona et al. 2010, Dalessandro et al. 2013, Di Criscienzo et al. 2015).

Because He affects the HB morphology both in terms of temperature (due to mass loss) and luminosity (because of the different contribution to the luminosity of the H-burning shell), variations in color (e.g., temperature) along the HB are largely degenerate with mass loss and age. Interestingly, the presence of He-enhanced populations along the blue part of the HB can be inferred without making assumptions about the RGB mass loss when a combination of optical and far-UV magnitudes is used (e.g., Dalessandro et al. 2011, 2013).

Further spectroscopic evidence (not including the measurement of He abundances) strengthens the connection between the HB morphology and the chemical composition (e.g., Lovisi et al. 2012, Gratton et al. 2014, Schaeuble et al. 2015). For example, the extension of the Na-O anticorrelation correlates with the maximum temperature of stars along the HB, indicating that the same physical mechanism responsible for the extreme Na enhancement and O depletion is also responsible for

the morphology of the blue tail at the end of the HB sequence (Carretta et al. 2010b). This correlation is interpreted as evidence that the HB morphology is determined not only by age and metallicity but also by the He abundance, as Na-rich stars are also He rich (e.g., Gratton et al. 2010). More massive clusters also tend to have HBs that are more extended toward higher temperatures (Recio-Blanco et al. 2006). This evidence in turn would again suggest that very massive GCs show larger extents of processing, i.e., very low O and high Na (see Section 2.5).

HBB: hot bottom burning

2.1.3. Lithium variations among globular cluster stars. Lithium traces mixing processes, as it is rapidly destroyed in proton captures at temperatures exceeding ~ 2.5 MK. Thus, if high values of N, Na, and Al are produced through hot H-burning, 2P stars should be depleted in Li. Some studies have revealed an anticorrelation between Na and Li, as expected (Pasquini et al. 2005, Lind et al. 2009, D’Orazi et al. 2015). However, importantly, other works have not found evidence for Li variations among stars with 1P and 2P composition (e.g., Mucciarelli et al. 2011). Because Li is destroyed at relatively low temperatures (i.e., well below temperatures where Na is formed), any material that is enriched in Na should be Li free. In order to explain the presence of some Li in 2P stars, it has been suggested that the polluters’ ejecta (i.e., Li free, Na, N-rich) must be mixed with unprocessed material, i.e., gas that has always been kept cooler than ~ 2.5 MK (Prantzos & Charbonnel 2006). Such models are known as dilution models (see Sections 4 and 5.1.1).

AGBs can potentially produce Li through the Cameron & Fowler (1971) mechanism at the beginning of the hot bottom burning (HBB) phase (e.g., Ventura et al. 2002). However, the finding of exactly the same Li abundance (or barely different) between 1P and 2P stars indicates that if AGB stars were responsible for the observed anomalies, they must have been able to (*a*) produce the same amount of Li previously destroyed by nuclear burning and (*b*) give yields close to the values of primordial nucleosynthesis. This concurrence certainly requires a high degree of fine-tuning and thus this explanation is unsatisfactory. By contrast, both massive star and VMS models require mixing with pristine material to account for the presence of Li in 2P stars because their ejecta are Li free. Thus, the maximum depletion of O in the final enriched composition cannot exceed the depletion of Li (Salaris & Cassisi 2014),⁴ which is contrary to what is observed (Shen et al. 2010). As a matter of fact, all the proposed scenarios have major problems in reproducing the Li content observed in clusters, where small (or no) variations of Li are found associated with large variations of other light elements.

2.1.4. Magnesium and potassium. Mg does not show significant star-to-star dispersion in all but a handful of GCs (Section 2.1.1). In only two clusters (namely NGC 2419 and, to a lesser extent, NGC 2808), low Mg abundances are also correlated with extreme K enhancements (e.g., Mucciarelli et al. 2012, Carretta 2015), whereas star-to-star scatter in K are not generally observed for the bulk of GCs (Takeda et al. 2009). The K overabundance of Mg-poor stars can be produced, under some assumptions, by AGBs (e.g., Ventura et al. 2012). However, both Na and Al are destroyed at the typical temperatures at which K is produced, e.g., Na and K are anticorrelated in stellar ejecta (Prantzos et al. 2017). Thus, the simultaneous Na and K enrichment seen in NGC 2419 and NGC 2808 cannot be explained if the observed Na and K inhomogeneities are produced by the same stellar source. As NGC 2808 and NGC 2419 are unusual in terms of the K-abundance patterns, it is not clear whether this is a promising window into the MP phenomenon or instead a pathological case that confuses the issue.

⁴The processed material is expected to be Li free, but it is only depleted in O.

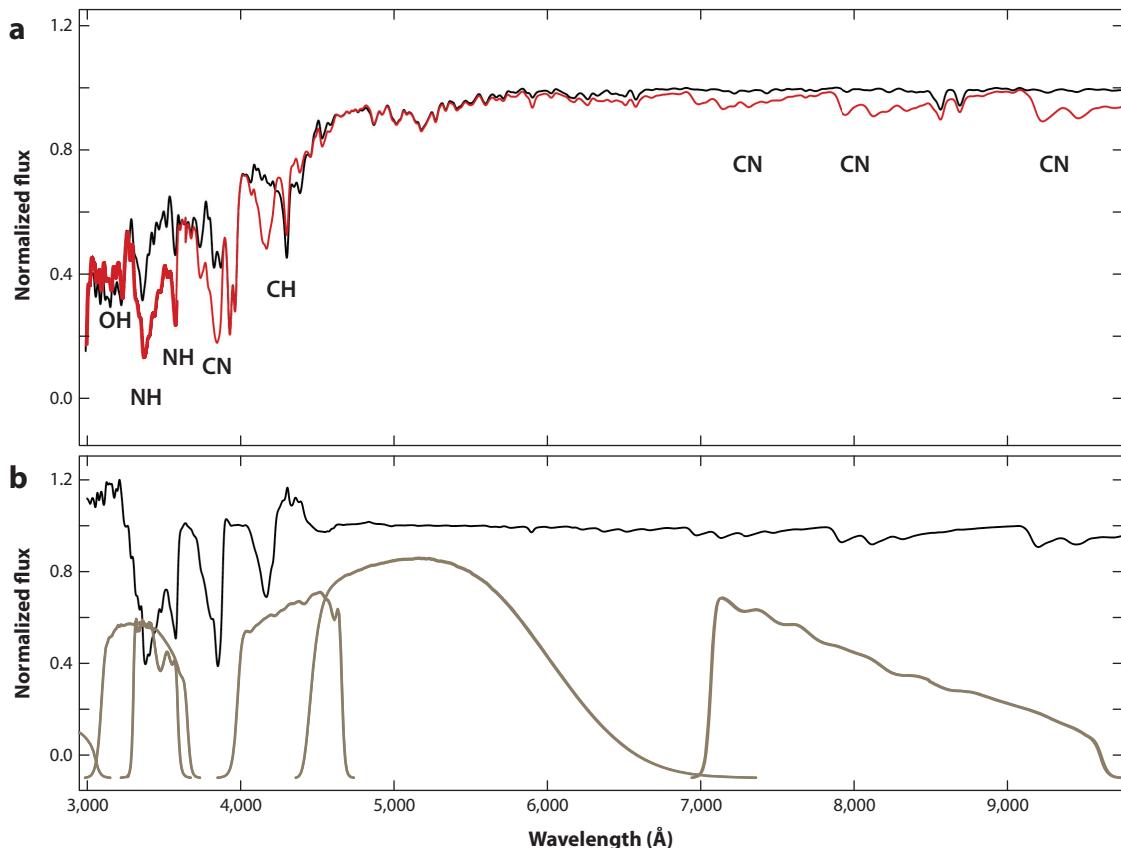


Figure 2

(a) Normalized synthetic spectra of red giant branch stars with 1P (primordial) and 2P (enriched) composition. A number of molecular absorption bands that vary significantly between the two spectra are also labeled. (b) The flux ratio between the two spectra, along with some WFPC3/UVIS filters used in photometric studies to pinpoint the presence and properties of multiple populations (*from left to right*: F336W (U), F343N, F438W (B), F555W (V), F814W (I); the value in parentheses indicates the approximate Cousins–Johnson filter equivalent). See also figure 4 from Sbordone et al. (2011). Abbreviations: 1P, primordial star; 2P, enriched star.

2.1.5. Multiple populations in extragalactic environments. MPs have also been found outside our Galaxy. Star-to-star abundance variations in N, Mg, Na, and Al were reported in extragalactic GCs by Mucciarelli et al. (2009), who studied three ancient GCs in the Large Magellanic Cloud (LMC) (see also Johnson et al. 2006 and Letarte et al. 2006 for earlier studies). They found that these three clusters followed the same Na–O and Al–Mg anticorrelation trends as seen in Galactic GCs. Hollyhead et al. (2017) measured the N and C abundances of stars in the ~8-Gyr Small Magellanic Cloud (SMC) cluster, Lindsay 1, based on low-resolution spectroscopy of cluster members. Using HST imaging in filters that are sensitive to C, N, and O variations (see Figure 2), Larsen et al. (2014b) determined the presence of MPs in four GCs in the Fornax dwarf spheroidal galaxy; they have also been detected in three 6–8-Gyr clusters in the SMC (Niederhofer et al. 2017b), as well as in the only classical GC in the SMC (Dalessandro et al. 2016, Niederhofer et al. 2017a).

There are a number of GCs within the MW that likely originate from accreted dwarf galaxies. These include GCs associated with the Sagittarius dwarf galaxy, for example, M54 (perhaps the nucleus of the galaxy; see Section 6.1), Terzan 7 and 8, Palomar 12, and Arp 2. M54 certainly

LMC: Large Magellanic Cloud

SMC: Small Magellanic Cloud

shows MPs (Carretta et al. 2010a), but the situation is less clear for Ter 7 and 8 and Pal 12 owing to the small samples of stars observed in each (e.g., Cohen 2004). In addition to resolved star studies, integrated light studies have also found strong evidence for MPs to be present in extragalactic clusters by looking for GCs that are strongly enriched in N or Na. These include many ancient GCs in M31 (Schiavon et al. 2013, Colucci et al. 2014, Sakari et al. 2015) and the lone GC associated with the WLM (Wolf–Lundmark–Melotte) dwarf galaxy (Larsen et al. 2014a).

There have also been attempts to search for MPs in extragalactic environments through integrated light photometry in the UV. If (large) He spreads are present within the clusters, an extreme HB may develop, causing significantly more UV emission than if all stars have the nominal He abundance. Such UV excess has been observed in some massive extragalactic GCs in M87, M31, and M81 (e.g., Sohn et al. 2006, Mayya et al. 2013, Peacock et al. 2017).

Based on these studies, along with those of Galactic GCs, it appears that one of the main properties of MPs is their near ubiquity in ancient and massive GCs (see Renzini et al. 2015). However, as is discussed in Section 2.5, this near ubiquity does not appear to apply to the young and intermediate age ($\lesssim 2$ Gyr) massive clusters in the LMC/SMC.

2.2. Multiple Populations as Seen Through Color-Magnitude Diagrams

The peculiar MP chemical composition can also be seen through accurate photometry (e.g., Hartwick & McClure 1972). Imaging allows us to discriminate efficiently between 1P and 2P subpopulations through photometry in samples composed of many thousands of stars while simultaneously covering a wider region in the sky (a result that is difficult to achieve with the most advanced spectroscopic facilities, even for nearby clusters). The relative number ratios between 1P and 2P stars can be inferred, and the radial distribution of the two groups can be investigated in detail by taking advantage of the large number of statistics secured through photometry (e.g., Lardo et al. 2011, Lee 2017). Nonetheless, wide-field photometric observations covering the full extension of the clusters (i.e., out to the tidal radius) are available only for a subset of clusters (Dalessandro et al. 2014, Massari et al. 2016) even if a large amount of archival data are publicly available in the archives.

HST offers high precision and accuracy to effectively sort different subpopulations (Piotto et al. 2015, Milone et al. 2017b, Soto et al. 2017). The *HST UV Legacy Survey of Galactic Globular Clusters* (G. Piotto, principal investigator; Piotto et al. 2015) has had a major impact on the field, allowing for the exploration of MPs and the link with their host cluster in unprecedented precision. However, space-based observations have only a limited spatial coverage.⁵ The less dense outer parts of clusters (where the two-body relaxation timescale is longer and mixing less efficient) can retain imprints of different initial configurations of MPs as differences in their relative spatial distributions or kinematics, hence their study allows us to gain crucial insights on the dynamics in play at the formation of the different subpopulations.

2.2.1. Causes for the complex color-magnitude diagrams and filter dependence. Splits or spreads in cluster CMDs have been used to identify MPs and constrain their properties. The cause of these splits depends on the color (or color combination) used to image clusters and on the specific evolutionary stage considered. Briefly, filters encompassing wavelengths shorter than ~ 4000 Å are very sensitive to individual variations of C, N, and O in the outer layers of stars with

⁵Furthermore, different regions of the clusters are included in the HST field of view, depending on the specific properties of the cluster itself, i.e., core/half-light radii and heliocentric distance.

SGB: sub-giant branch

cooler atmospheres. Conversely, star-to-star variations in He (as well as the CNO sum) impact primarily the stellar structure. As such, they affect mainly optical bands although they have some influence on the UV.

Salaris et al. (2006) first considered the effect of He and light element variations on photometry. They conclude that in the Johnson–Cousins B,V, and I filters only an extreme He enhancement ($Y \geq 0.35$) leads to an appreciable color change of stars with 2P composition as compared to standard 1P stars. A prominent splitting of the MS and the MSTO is produced by relatively large He enhancements, whereas color variations due to He variations are less pronounced in the RGB in optical colors. The CNONa anticorrelations do not affect the evolutionary properties of stars, hence the position of stellar models in the theoretical Hertzsprung–Russell diagram, when the C+N+O sum is kept constant (Sbordone et al. 2011). However, the observed splitting of the sub-giant branch (SGB) into brighter and fainter sequences in some clusters in optical filters can be interpreted as the result of a change in the C+N+O sum (Cassisi et al. 2008, Piotto et al. 2012). Furthermore, 1P and 2P stars also have slightly different luminosity at the RGB bump, and they occupy different regions on the HB when clusters are imaged with optical BVI filters (e.g., Bragaglia et al. 2010a).

Larger color spreads (from the MS up to the RGB, where the effect tends to be larger) are expected in CMDs including near-UV filters, even while leaving the C+N+O sum unchanged (see Pietrinferni et al. 2009 for a comprehensive discussion). C, N, and O individual variations are critical, whereas He enhancement works in the opposite direction of CNONa spreads. This property appears to be shared by any filter encompassing the wavelength range $3000 \leq \lambda \leq 4000 \text{ \AA}$, where most of the NH and CN absorptions are located. In **Figure 2**, we show synthetic spectra of RGB stars with typical 1P and 2P chemical abundances and highlight molecular bands that differ between the spectra. Additionally, in **Figure 2b**, we show the flux ratio between the two spectra and the throughput curves of selected HST filters. Due to these spectral differences the color spread observed in specific color combinations including near-UV filters has been shown to be very sensitive to light element abundances (e.g., Marino et al. 2008). Several combinations of colors have also been introduced to best disentangle the different subpopulations. For example, Monelli et al. (2013) found that all of the 23 clusters in their sample analyzed with ground-based photometry show broadened or multimodal RGBs in the $C_{U,B,I} = (U - B) - (B - I)$ versus V CMDs, where the different branches of the RGBs are tightly linked to their light element content (see **Figures 1** and **2**). Niederhofer et al. (2017a,b) imaged a number of clusters in the LMC in the color index $C_{F336W,F438W,F343N} = (F336W - F438W) - (F438W - F343N)$ to pinpoint the presence of MPs with different C and N abundances, finding evidence for MPs for all observed clusters older than ~ 6 Gyr (see also Hollyhead et al. 2017).

In **Figure 1**, we show an example case of NGC 6752. In **Figure 1a**, we show the $C_{U,B,I}$ CMD showing the split/spread RGB of the cluster in this filter combination. Additionally, we show the position of stars on the RGB, labeled in terms of their chemical abundances (**Figure 1b**). Hence, the position of a star in CMDs, in specific filter combinations, can be used to trace the chemical composition of the stars.

Milone et al. (2017b) used a similar color index to constrain the presence and properties of MPs in 57 Galactic old clusters using the large database of data coming from the HST Large Program, the *HST UV Legacy Survey of Galactic Globular Clusters: Shedding UV Light on Their Populations and Formation* (see Piotto et al. 2015, Soto et al. 2017). UV observations taken in the F275W, F336W, and F438W filters further complement optical HST observations from the *ACS Survey of Galactic Globular Clusters* (e.g., Sarajedini et al. 2007) with WFC3/UVIS images. The defined $C_{F275W,F336W,F435W} = (F275W - F336W) - (F336W - F435W)$ color combination allows one to clearly identify photometric splits/spreads caused by variations in individual elements, namely C,

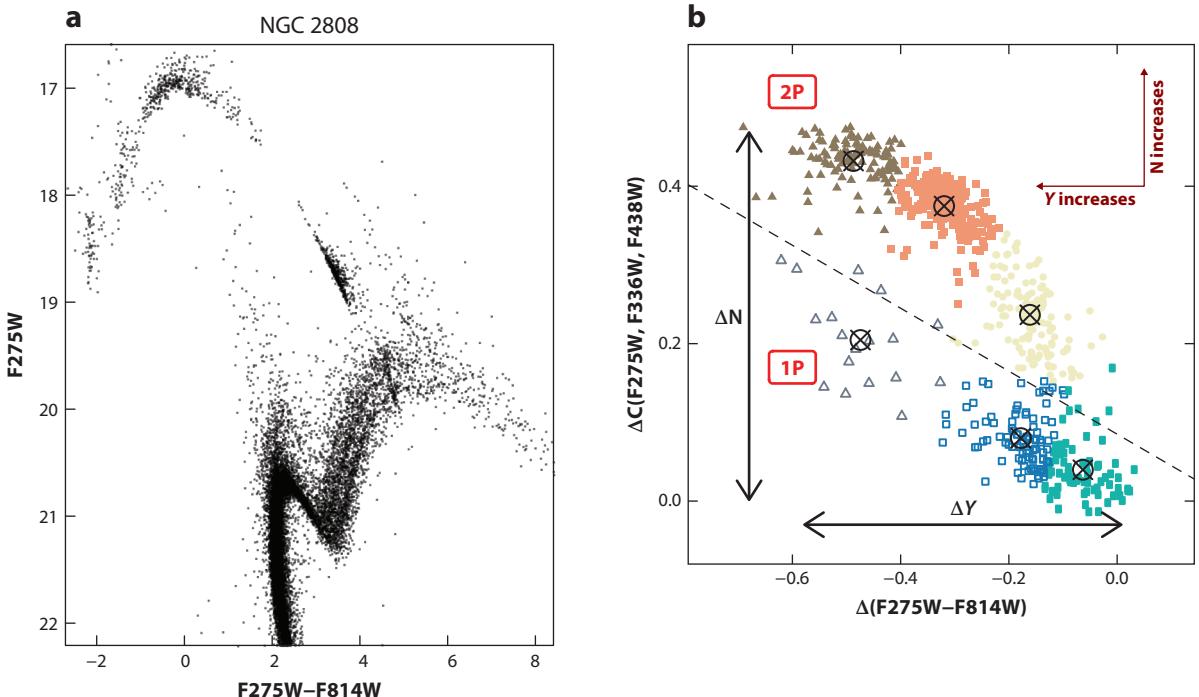


Figure 3

(a) An HST UV-optical color-magnitude diagram of the central regions of NGC 2808 (data are from Piotto et al. 2015 and Soto et al. 2017). Note the distinct multiple RGB stars and the highly structured horizontal branch. This complexity is due to light element abundance variations (He, C, N, and O) between cluster stars. Data are taken from Piotto et al. (2015) and Soto et al. (2017). (b) A “chromosome map” of NGC 2808 (after Milone et al. 2017b) for RGBs (i.e., relative positions of the stars on the RGB in different filter combinations that are sensitive to different abundance variations) where at least six distinct populations can be inferred. Here the x axis is mainly sensitive to variations in He, whereas the y axis is dominated by variations in N (at C, and at O to a lesser extent). Based on the definition of Milone et al. (2017b), stars above the dashed line are considered to be 2P, whereas stars below the same line are 1P. Note that both the 1P and 2P consist of three extended subpopulations. Abbreviations: 1P, primordial star; 2P, enriched star; RGB, red giant branch.

N, and O (see **Figure 2**). Also, the combination of UV CMDs with optical photometry allows He enhancement (ΔY) of the different subpopulations to be seen.

A pseudo color-color diagram (or chromosome map; see **Figure 3**) has also been introduced to identify different subpopulations from the HST UV survey photometry by highlighting subtle chemical differences (in light elements and He) between them (e.g., Milone et al. 2015b). Briefly, two fiducial lines are drawn to fit at the blue and red envelope of the RGB sequence in the F814W versus $C_{F275W,F336W,F435W}$ and F814W versus $(F275W - F814W)$ CMDs. The red and blue fiducial lines are then used to verticalize the RGB sequence in such a way that they translate into vertical lines. A pseudo color-color plot can then be made of the position of each RGB star in the verticalized colors, $\Delta_{C(F275W,F336W,F438W)}^N$ and $\Delta_{F275W,F814W}^N$. An example of such a diagram can be seen in **Figure 3** for NGC 2808 stars, which reveals the presence of at least six subpopulations with distinct chemistry.

With such diagrams, Milone et al. (2017b) were able to efficiently distinguish the 1P and 2P populations for most clusters, although some clusters did display a continuous distribution (see **Figure 3** for the division). These distinctions were confirmed through comparison with the results

of ground-based spectroscopic studies; i.e., 1P stars identified photometrically corresponded to stars with the field abundance patterns of Na and O.

YMC: young massive cluster (a.k.a. a young GC)

With the precision of HST photometry, relatively tight constraints can be placed on any age difference between the populations. Using the *HST UV Legacy Survey* data, Nardiello et al. (2015) selected stars from the 1P and 2P populations based on UV images in the Galactic GC NGC 6352. The authors then estimated the age of each population independently, using optical CMDs (V–I versus I) centered on the MSTO of each population. The optical colors are not strongly affected by MPs (although He variations can affect optical colors as well as nonconstant C+N+O sums), hence any differences would be attributed primarily to age differences (if He variations are taken into account, which the authors did). In this case, the age difference was found to be 10 ± 110 Myr. When all sources of uncertainties are included (including the $[\alpha/\text{Fe}]$ ratio), the authors find that the two populations are coeval with an upper limit of 300 Myr between them. This is consistent with a similar upper age limit found by Marino et al. (2012) for M22. Tighter age constraints can be gotten from younger clusters that show MPs (a young massive cluster or YMC; see Section 5.6).

2.2.2. A spread among 1P stars? An unexpected result of the Milone et al. (2017b) study was that the 1P population displayed a significant spread in some clusters (although no spread was seen in Na and O for these stars) while being quite compact in other clusters. Based on the data provided by Milone et al. (2017b), it appears that $\sim 70\%$ of the GCs in that sample display a significant spread in their 1P stars. Although this appears to be common, many clusters do not show an extended 1P, and it is not clear at present what (if any) cluster property controls the spread in the 1P stars.

Preliminary computations (Lardo et al. 2018) reveal that for intermediate and low metallicities the $\Delta C_{F275W,F336W,F438W}$ color spread essentially traces N (e.g., stars are sorted in order of increasing N abundance from bottom to top in the chromosome map of **Figure 3**). Conversely, the $\Delta_{F275W,F814W}$ color spread is sensitive to He enhancement of the different subpopulations (e.g., in order of increasing He content, from right to left; see **Figure 3b**). The spread in 1P stars is seen predominantly in the F275W–F814W color (UV–I), suggesting that He variations are present within the 1P, which would be very surprising given the lack of Na, N spread, or O variations within this population. This in turn suggests that some stars with little or no N spread show significant enhancement in their He values, which is in conflict with basic nucleosynthesis. Hence something else, other than the recycled by-products of stellar nucleosynthesis, has caused the He variations within 1P stars. This is a particularly promising avenue for future study.

2.3. Are There Single Population Globular Clusters?

Nearly all GCs analyzed at high resolution, with the exceptions of Ter 7, Pal 12 and 3, and Ruprecht 106, show the Na and O variations. Ter 7 and Pal 12 are low-mass members of Sagittarius, and high-resolution abundances exist only for a handful of cluster members (fewer than five stars; Cohen 2004, Tautvaišienė et al. 2004, Sbordone et al. 2007). The same holds true for Rup 106, a slightly more massive ($5 \times 10^4 M_\odot$) cluster with a probable extragalactic origin (nine stars; Villanova et al. 2013), and Pal 3, a distant GC in the outer halo, where the available data (two stars) can neither confirm nor refute the presence of an Na–O anticorrelation (Koch et al. 2009). Increasing the sample of stars studied in these low-mass clusters is essential to determining if there is a lower GC mass limit where MPs are present (e.g., Dalessandro et al. 2014). In this respect searching for MPs through photometric methods can be problematic in these clusters as the low number of RGB stars often makes it difficult to identify MPs there, unless the populations are well separated (i.e., have large N or He variations).

In this respect, the case of the SMC old cluster NGC 121, studied by Dalessandro et al. (2016), is quite illustrative. The authors derived Na and O for five RGBs and found no intrinsic scatter in both elements. However, they detected two RGB sequences in their UV images, meaning that MPs are present. 2P stars were missed in their spectroscopic sample as it was biased (as most spectroscopic samples are) to the outer regions of the cluster, where the fraction of 2P stars is often lower than in the central regions.

Two other old GCs have been claimed not to host MPs based on either ground-based photometry or low-resolution spectroscopy, E 3 (Salinas & Strader 2015) and IC 4499 (Walker et al. 2011), although follow-up HST photometry has detected MPs in IC 4499 (E. Dalessandro, C. Lardo, M. Cadelano, N. Bastian, and A. Mucciarelli, manuscript in preparation). Additional high-resolution studies designed to measure the abundance of the relevant light elements (e.g., Na, O, etc.) for a representative number of stars in such clusters are needed to draw firm conclusions on the presence of MPs.

As is discussed in Section 2.5, a number of high-mass ($\sim 10^5 M_{\odot}$) clusters younger than ~ 2 Gyr have been studied, and so far none have been found to host MPs (e.g., Mucciarelli et al. 2008, 2014a; Martocchia et al. 2017).

2.4. Global Properties and Correlations

2.4.1. Spatial distributions, dynamics, and binary properties of the different populations.

In many cases different stellar subpopulations seem to not share the same radial distribution. Across a range of cluster-centric distance, most studies have found that 2P stars are systematically more concentrated in the innermost region than 1P stars (e.g., Lardo et al. 2011, Simioni et al. 2016). Only a few exceptions to this general trend have been reported, with stars having primordial composition being more centrally concentrated than 2P giants (Larsen et al. 2015, Vanderbeke et al. 2015, Lim et al. 2016) or 1P and 2P stars having the same radial distribution (e.g., Dalessandro et al. 2014, Miholics et al. 2015). Hints that 2P stars have lower velocity dispersion (e.g., Bellazzini et al. 2012, Kučinskas et al. 2014) and more radially anisotropic velocity distribution (Richer et al. 2013, Bellini et al. 2015) have also been reported. The binary properties of 1P and 2P stars may also be different, with 2P stars showing a lower binary fraction (D’Orazi et al. 2010, Lucatello et al. 2015).

2.4.2. Observed population ratios.

Although there are radial trends in the 2P/1P ratios, in most cases large samples of stars are required to demonstrate this statistically. Overall, 2P stars make up the majority of stars in most GCs, although the fraction of 2P stars is seen to be a strong function of cluster mass, with more massive clusters having larger fractions of 2P stars (e.g., Milone et al. 2017b; see **Figure 4**). Bastian & Lardo (2015), using mainly spectroscopic results from the literature that are biased toward the outer regions of clusters, did not find any trends between the enriched fractions ($f_{\text{enriched}} = N_{\text{2P}}/N_{\text{tot}}$) and metallicity or galactocentric distance.⁶ This has been confirmed with HST photometry (Milone et al. 2017b). Hence, the MP phenomenon is not directly linked to the environment in which the cluster forms (e.g., within dwarf galaxies or the bulge of the Galaxy). The trend between population ratios and mass is a key constraint on scenarios for the origin of MPs, which is discussed in Section 5.5.

⁶They also did not find any correlations between f_{enriched} and cluster mass, but found an average value of $f_{\text{enriched}} = 0.68$, which agrees well with the average from HST photometry, although why they did not find a trend with mass is not entirely clear.

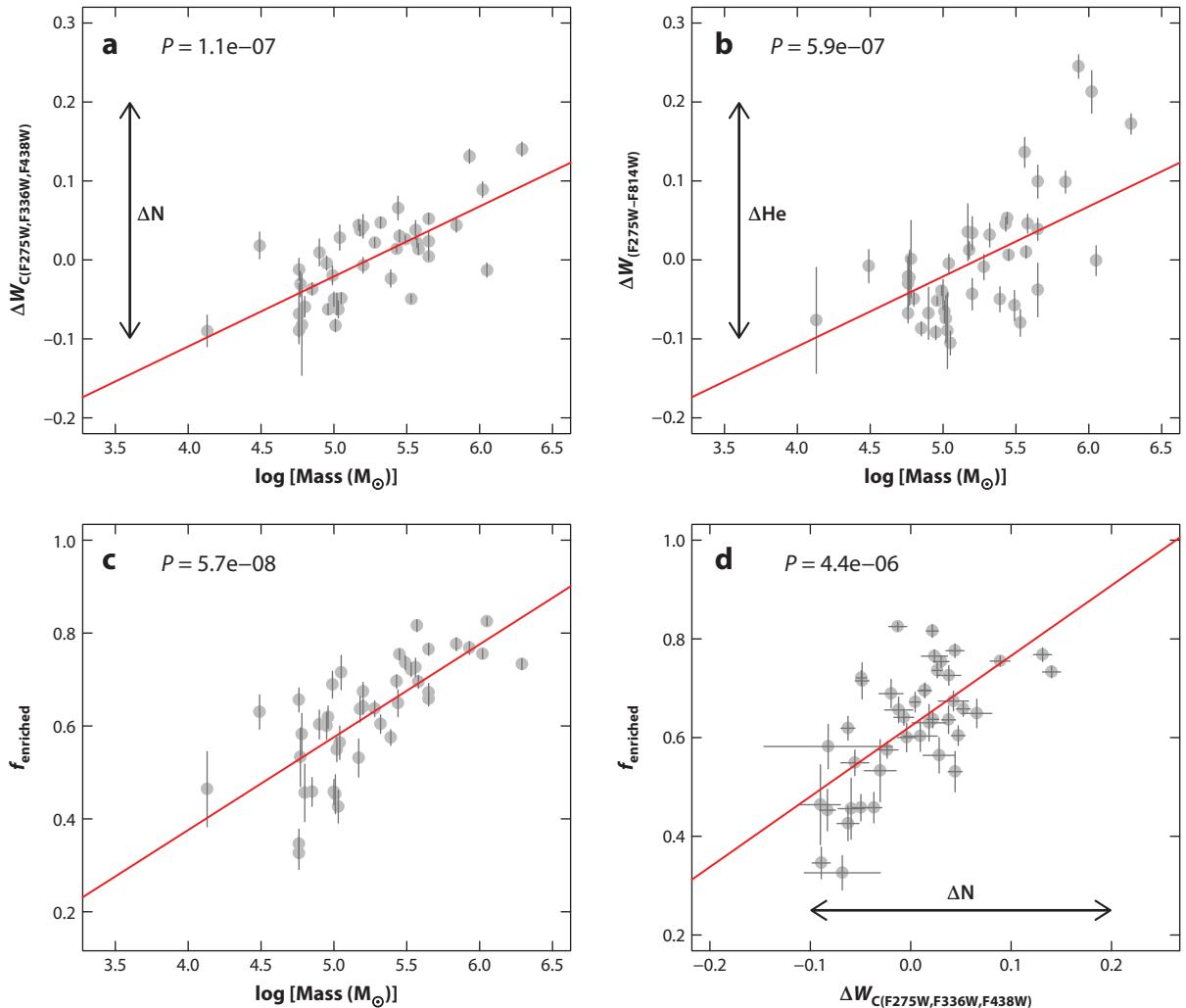


Figure 4

Based on results from the *HST UV Legacy Survey of Galactic Globular Clusters*, we show a summary of how multiple population properties vary with the present-day globular cluster mass (after Milone et al. 2017b). (a) $\Delta W_{F275W,F336W,F438W}$ and (b) $\Delta W_{F275W-F814W}$ are the widths of the RGB in the two colors (or color combinations), which have been corrected for the effect of metallicity, in a first approximation measurement of the amount of N- and He-enrichment (respectively) present in the cluster (i.e., the difference between the most enriched stars and the 1P stars). f_{enriched} is the fraction of 2P stars relative to the total number of stars, as measured on the RGB. f_{enriched} versus (c) cluster mass and (d) $\Delta W_{F275W,F336W,F438W}$. In each of the four panels, the solid red lines give the best linear fit to the data and the probability of no correlation between the points (P) is shown. Self-enrichment scenarios (for standard nucleosynthetic stellar sources) all predict an anticorrelation between f_{enriched} and $\Delta W_{F275W,F336W,F438W}$, which is opposite to the observed trend. All data are from Milone et al. (2017b). Abbreviations: 1P, primordial star; 2P, enriched star; RGB, red giant branch.

2.5. The Role of Cluster Age and Mass

It is still not clear precisely which properties of the clusters determine whether MPs will be present within the cluster. However, with the release of large and homogeneous surveys, we can begin searching for correlations between cluster properties (e.g., age, mass, location) and the presence/absence of MPs as well as their extent in order to glean clues as to the mechanisms responsible

for MPs. In **Figure 5**, we show a collection of clusters from the literature where searches for MPs have been conducted in the age-mass plane and the [Fe/H]-concentration (mass/R_h) plane.

2.5.1. Cluster mass. As it became apparent that (nearly) all of the ancient GCs host MPs and that (so far) none of the OCs do, it was suggested that cluster mass may play a key role (e.g., Carretta et al. 2010b). The general argument is that if clusters host a deep enough gravitational potential well, they may be able to retain the stellar ejecta of a 1G and form second generation (2G) stars with that enriched material. This is generally based on an escape velocity argument, although this often overlooks the role of energetic stellar sources, like high- or low-mass X-ray binaries or ionizing white dwarfs (e.g., D’Ercole et al. 2008, Krause et al. 2012, McDonald & Zijlstra 2015).

Second generation (2G) stars: in models of MP formation, stars of the second generation that show the anomalous chemistry

Cluster mass does appear to be an important parameter for GCs, playing a role in determining whether MPs are present as well as in the properties (i.e., how severe the abundance variations are) of the MPs. The first hints for this came from Carretta et al. (2010b) who used their large sample of stars in 19 GCs to search for correlations between the extent of the Na-O anticorrelation (as measured through the interquartile distribution) and various cluster properties. The strongest relation found was with cluster mass, with higher-mass clusters showing larger Na-O abundance spreads. This is difficult to reconcile with standard stellar evolution, as the stellar ejecta released into the cluster should not depend on cluster properties. For models that invoke dilution, this would require that lower-mass clusters undergo more dilution (though lower-mass clusters would be expected to accrete less gas from their surroundings) or that higher-mass clusters retain a larger fraction of the processed material (although models already adopt that all clusters retain 100% of the processed material). Because models already assume that GCs retain 100% of the material processed through the enriching source (e.g., FRMSs, AGBs, interacting binaries or IBs, etc.), this further exacerbates the mass-budget problem (see Section 5.4).

Similarly, Milone (2015) found that the He spread (ΔY) within Galactic GCs is much larger in higher-mass clusters. Although this was only based on nine GCs, it will be directly tested with a much larger sample from the *HST UV Legacy Survey* (Piotto et al. 2015). In **Figure 4**, we show the results from Milone et al. (2017b) for the width of the RGB in the (F275W–F814W) CMD (corrected for metallicity effects), which is a proxy for He spread (Lardo et al. 2018). This confirms and extends the trend reported by Milone (2015). One of the major results from the *HST UV Legacy Survey* has been the discovery of a strong correlation between cluster mass and the fraction of enriched stars (f_{enriched}) within the cluster (Milone et al. 2017b). Here, f_{enriched} is found in the $\Delta_{\text{F275W,F814W}}$ versus $\Delta C_{\text{F275W,F336W,F438W}}$ color–color plot (see **Figure 3**). The authors note that in some cases the 1P population is made up of multiple groups, hence f_{enriched} may be a lower limit. In **Figure 4**, we show some of the main results from Milone et al. (2017b), namely how N spread ($\Delta C_{\text{F275W,F336W,F438W}}$), He spread ($\Delta_{\text{F275W,F814W}}$), and f_{enriched} vary as a function of mass (after removing the trends with metallicity).

High-mass clusters (e.g., NGC 2808, 47 Tucanae with $M_{\text{cluster}} \sim 10^6 M_{\odot}$) have $f_{\text{enriched}} \approx 0.8$, whereas clusters with masses near $10^5 M_{\odot}$ have $f_{\text{enriched}} \sim 0.4\text{--}0.5$. Note that the enriched population still makes up a substantial fraction of the stars even in low-mass clusters. It is not just the fraction of enriched stars that varies with cluster mass; it is also the extent of the enrichment as well (i.e., larger abundance spreads in higher-mass clusters). This is in agreement with earlier work based on spectroscopic samples (Carretta et al. 2010b). The implications of these results are discussed in Section 5.5.

There have also been studies focused on old OCs, which typically have masses much lower than GCs, e.g., Berkeley 39 (Bragaglia et al. 2012). To date, MPs have not been found in OCs with masses as high as $2 \times 10^4 M_{\odot}$ and ages as old as $\sim 6\text{--}9$ Gyr. Comparison of 6–8-Gyr-old clusters in the SMC having masses of $\sim 10^5 M_{\odot}$ (Hollyhead et al. 2017, Niederhofer et al. 2017b)

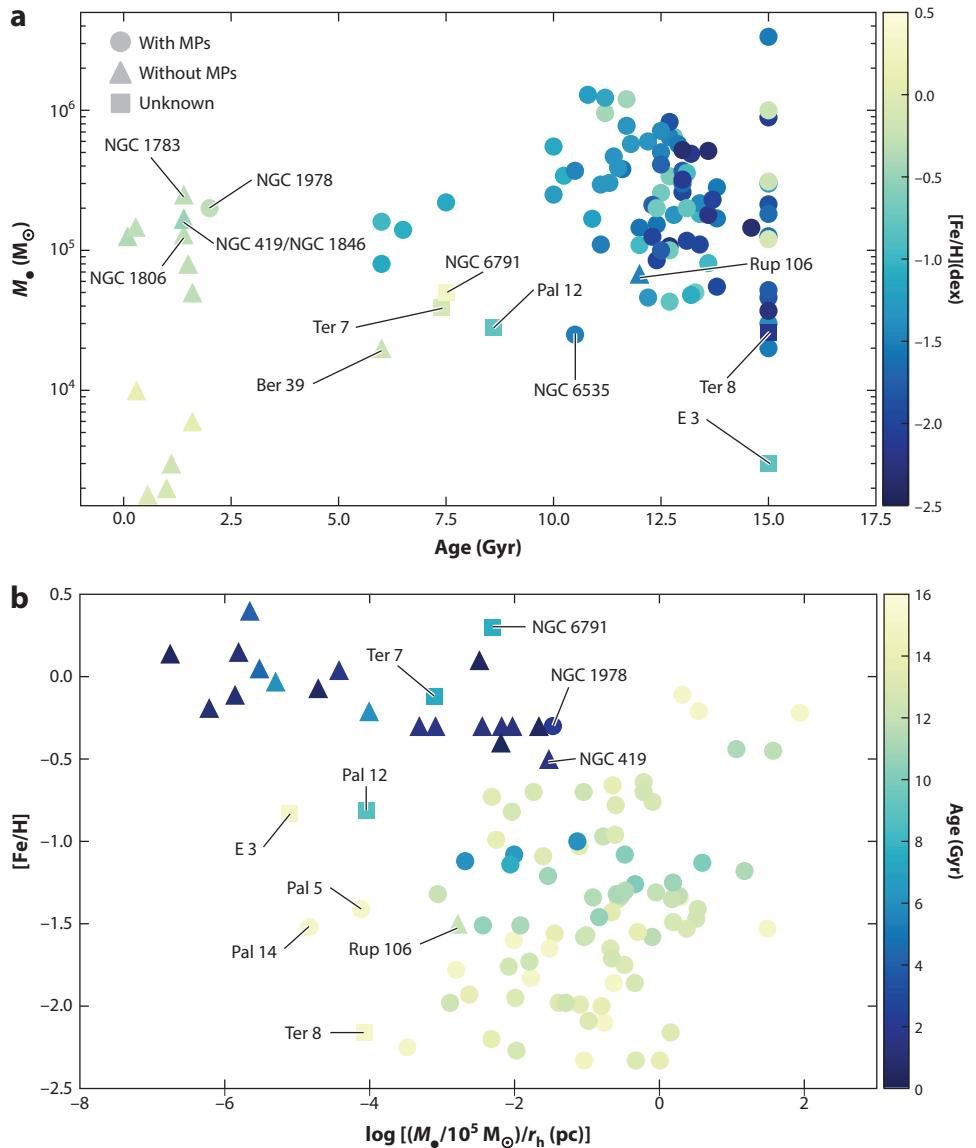


Figure 5

A summary of results from the literature on whether MPs are present within clusters in (a) the age-mass plane and (b) the concentration-[Fe/H] plane. Circles denote clusters where MPs have been unambiguously detected, triangles show where they have not been detected (with large enough samples to suggest a true absence), and squares show ambiguous cases (mainly due to small samples or potentially large observational uncertainties). Some particularly interesting cases are labeled. An age of 15 Gyr has been assigned to clusters for which no age determination has been found in the literature. Whether or not a cluster hosts MPs depends on its mass (or density) as well as its age. The data come from the compilation of Krause et al. (2016) with additional points added from more recent works discussed in this review. Abbreviation: MP, multiple population.

with their lower-mass OC counterparts (e.g., Ber 39) hint that mass may indeed play a role (see **Figure 5**). The SMC clusters do host MPs, whereas OCs do not. However, of course, the formation environment may also have been different.

Recent studies have also targeted low-mass ancient GCs, such as NGC 6362 ($M \sim 5 \times 10^4 M_{\odot}$; Dalessandro et al. 2014) or E 3 ($1.4 \times 10^4 M_{\odot}$; Salinas & Strader 2015), with mixed results. NGC 6362 does host MPs, but based on its orbit and observed stellar mass function, it has likely lost a significant amount of mass during its evolution (e.g., Kruijssen & Mieske 2009). E 3, on the other hand, does not appear to host MPs, based on CN low-resolution spectra. The very extended ($R_h \sim 25$ pc) outer halo cluster Palomar 14 with a mass of only $\sim 10^4 M_{\odot}$ does appear to host MPs (Çalışkan et al. 2012). The current record holder for the lowest current stellar mass cluster that still hosts MPs is NGC 6535 with a mass of a few $\times 10^3 M_{\odot}$ (Bragaglia et al. 2017, Milone et al. 2017b).

A summary of the role of mass (and concentration) in whether a cluster hosts MPs or not is shown in **Figure 5**. There is overlap between ancient GCs that do host MPs and younger clusters that do not. However, the data are consistent with a lower initial mass limit of $\sim 10^5$ where MPs can develop (at least for clusters older than ~ 2 Gyr; see Section 2.5.2).

2.5.2. Cluster age and metallicity. As discussed above, nearly all of the ancient GCs that have been studied in the necessary detail host MPs. However, there are stellar clusters that formed after the peak epoch of GC formation ($z = 2-5$), continuing to form up to the present day, that have masses and densities comparable to, or even significantly above, the ancient GCs. Hence, an obvious question is whether these clusters also host MPs, and if so, can they be used to test the formation scenarios that have been put forward (see Section 4).

There have been a number of studies to search for MPs in massive clusters with ages < 8 Gyr (see Krause et al. 2016 for a recent review). With only a handful of exceptions (discussed above), all massive clusters older than ~ 6 Gyr host MPs (Hollyhead et al. 2017, Niederhofer et al. 2017b), whereas all clusters younger than ~ 2 Gyr do not (e.g., Mucciarelli et al. 2008, 2014a; Martocchia et al. 2017), even with mass being held constant (at $\sim 10^5 M_{\odot}$; see **Figure 5**).

The ~ 6 -Gyr clusters, NGC 339, NGC 416, and Kron 3, all located in the SMC, show clear evidence for MPs (Niederhofer et al. 2017b). This age corresponds to a formation epoch of $z_{\text{form}} = 0.65$, arguing against a cosmological origin of the phenomenon (i.e., special properties of the early Universe that contributed to the formation of MPs). Unexpectedly, however, determined on the basis of HST photometry, another massive cluster in the SMC, NGC 419, with an age of ~ 1.7 Gyr and a similar mass of $\sim 2 \times 10^5 M_{\odot}$, does not host MPs (Martocchia et al. 2017). The youngest cluster found so far to host MPs is NGC 1978, at an age of ~ 2 Gyr (Martocchia et al. 2018a), suggesting that MPs (at least on the RGB) develop in an extremely narrow age range (or alternatively stopped being able to form in the LMC/SMC) between ~ 1.7 and 2 Gyr.⁷ This is shown in **Figure 5**, where clusters like NGC 1783 and NGC 1978 lie on opposite sides of this dividing line in age despite having nearly identical masses. However, there are also older clusters like Ber 39 (a Galactic OC) that do not host MPs, suggesting that mass (and potentially formation environment) plays a strong role as well.

There have also been a number of studies that have searched directly for abundance spreads in young/intermediate-age massive clusters, based on high-resolution spectroscopy (e.g., Mucciarelli

⁷We note, however, that in the 2–8-Gyr clusters, MPs have only been identified through N variations. High-resolution studies to estimate Na and O also in these stars would be a welcome contribution.

et al. 2008, 2014a) of individual stars. No solid evidence for abundance spreads has been found so far for any cluster less than \sim 2 Gyr.

A number of studies have attempted to search for abundance anomalies through integrated light spectral studies (e.g., Colucci et al. 2012, Cabrera-Ziri et al. 2016b, Lardo et al. 2017a). These are mainly focused on finding high mean levels of elements that typically vary due to MPs, namely [Na/Fe] or [Al/Fe]. As with the resolved studies, to date there have not been clear indications for abundance spreads in the young or intermediate-age clusters (<2 Gyr), although the ancient GCs do show the expected trends in integrated light.

Finally, Figure 5 also shows the results from the literature on whether a cluster hosts MPs in [Fe/H] versus concentration (mass/radius) space. There is overlap in both [Fe/H] and concentration where clusters do or do not host MPs. Systematic searches for MPs in diffuse GCs may lead to significant new insights.

2.6. Observational Summary of Multiple Populations

1. MPs, as seen in light element abundance spreads (C, N, O, Na, Al, He, and sometimes Mg), are nearly ubiquitous in old massive GCs, independent of their formation environment (formed within the Galaxy or elsewhere) or metallicity.
2. MPs can be defined through clear correlations and anticorrelations between light elements—the main ones being an Na-O anticorrelation, an N-C anticorrelation, an Na-N correlation, and N and Na being correlated with He. In some clusters Li is correlated with O (and hence anticorrelated with Na), but Li measurements are relatively scarce.
3. In most clusters [Fe/H] is constant between the populations, and the sum of C+N+O is also typically constant within the measurement uncertainties (although there are more clusters with C+N+O spreads than those with [Fe/H] spreads).
4. Observed abundance trends are qualitatively consistent with those expected from the yields of hot H-burning (increase in He, N, Na, and sometimes Al; decrease in C, O, and sometimes Mg); however, no nucleosynthetic source provides a quantitative match to the data simultaneously.
5. It is the spreads in He, C, N, and O (mainly) that cause the complexity observed in high-precision CMDs for the majority of clusters (i.e., not age spreads or Fe spreads).
6. The fraction of enriched stars (ranging from 40–90% in the ancient GCs), the extent of the anticorrelations, and the He spread within the clusters are all strong functions of the cluster mass (all increasing with increasing mass). Hence, the cluster properties appear to play a strong role in the formation of MPs. 2P stars make up the majority of stars in most GCs today, meaning that a substantial amount of processed material is required to form them. This leads to the “mass-budget problem” which is discussed in Section 5.4.
7. The abundance patterns are discrete when high-precision measurements are possible, with many clusters showing the presence of >3 –4 subpopulations.
8. The majority of clusters in the *HST UV Legacy Survey* (\sim 70%) show a spread in their 1P stars, in addition to the spread in the 2P stars. Preliminary modeling suggests that this is mainly due to He variations in 1P stars that are not accompanied by variations in other light elements (e.g., N, Na, O).
9. In most clusters studied to date the 2P stars are either more centrally concentrated than the 1P population or if the cluster is dynamically relaxed, the two populations share the same distribution. However, in a handful of cases the situation is reversed, with the 1P stars being more centrally concentrated than the 2P stars.

10. MPs have been detected in clusters as young as ~ 2 Gyr, which corresponds to a formation redshift of $z = 0.17$, which is well past the peak epoch of GC formation ($z = 2\text{--}5$). Surprisingly, MPs have not been found in massive clusters with ages less than 2 Gyr.
11. MPs are found in the full range of GC metallicities, from $[\text{Fe}/\text{H}] \sim -2.5$ to near solar metallicity.

3. NUCLEOSYNTHESIS AND MULTIPLE POPULATIONS

All elements whose abundances show considerable scatter in GC stars (i.e., C, N, O, Na, Mg, and Al) may participate in hydrostatic H-burning. As a consequence, the presence of the C, N, O, Na, and Al anticorrelated ranges observed in GCs has been interpreted as the results of H-burning through the CNO cycle and the NeNa- and MgAl-chains (e.g., Langer et al. 1993). In the CNO cycle, H is converted into He, and the individual abundances of the C, N, and O are altered while their net sum remains constant (as required by observations; see Section 2.1.1). The CNO cycle is activated at $T \sim 20$ MK, whereas the NeNa chain requires temperatures around ~ 40 MK. Na reaches its equilibrium value at ~ 50 MK and decreases at higher temperature. At higher temperatures ($T \simeq 70$ MK), Al can be produced by proton captures on Mg isotopes (e.g., Denisenkov & Denisenkova 1989, Prantzos et al. 2007).

Three stellar types have been proposed as candidate polluters, because they reach extreme temperatures within their interiors (see also Sections 2.1.3 and 6.2 for additional constraints from elements others than CNO, Na, Al, and Mg). The possible 2P processed material donors are intermediate-mass ($\sim 3\text{--}8 M_{\odot}$) AGB stars experiencing HBB (e.g., D'Antona et al. 2016), massive stars ($\geq 15 M_{\odot}$; de Mink et al. 2009, Krause et al. 2013),⁸ and VMSs ($\sim 10^4 M_{\odot}$; Denissenkov & Hartwick 2014). Scenarios have also been proposed in which the mixed contributions are from different polluters (e.g., Sills & Glebbeek 2010, Bastian et al. 2013b).

As we discuss the characteristics of each of the proposed stellar sources as well as the scenarios developed around them, we keep track of their successes and failures to reproduce key observations in **Figure 6**. When a model matches an observation a green check is used, whereas a green check with an asterisk notes that the model may be consistent with observations under reasonable assumptions. Red crosses indicate when a model is in direct conflict with an observation, and a red cross with an asterisk shows where a model may match an observation but requires a high degree of fine-tuning, or the solving of that problem would violate another constraint.

Several observational constraints can naturally be reproduced within the proposed self-enrichment scenarios. Yet, a number of ad hoc assumptions must be made to explain other MP properties. For the sake of clearness, in what follows, we briefly introduce and discuss candidate stellar polluters for GC self-enrichment (see also Renzini et al. 2015, Charbonnel 2016).

3.1. Massive Stars

Massive ($\geq 15 M_{\odot}$) MS stars reach the high temperatures required to manufacture the observed CNONaAl pattern very early in their MS evolution (e.g., Maeder & Meynet 2006). The fast rotation required by MP models allows for the transport of nuclides from the convective core to the radiative envelope while losing mass through (*a*) a slow outflowing equatorial disc produced by a mechanical wind when the MS star rotates close to critical velocity, and (*b*) a fast radiatively

⁸This happens in the cores of massive stars, so additional processes are necessary to bring the material to the surface. In the case of single stars, rotational mixing has been suggested, the so-called FRMSs. Interactions between massive stars in binary systems can also bring processed material to the surface.

	(Near) ubiquity (Section 2)	Abundances (Section 5.1)	Li correlations (Section 5.1.2)	Variety/Stochasticity (Section 5.1.3)	Mg-Al (Section 5.1.5)	Discreteness (Section 5.2)	^f enriched correlation with GC mass (Section 5.4)	He spread correlation with GC mass (Section 5.5)	YMCs (Section 5.6)	Age trend (Section 5.6.1)
AGB	X	X*	X*	X	✓*	X*	X*	X	X*	X
FRMS	X	X*	X	X	X	X*	X	X	X	X
VMS	X	X*	X	?	X*	X	✓*	✓*	✓	X
EDA	X	X*	X	X	X	X*	X	X	✓*	X
Reverse order	X	X*	X	?	?	X	✓*	X	X	X
eSF period	X	X*	X	X	X	X*	X	X	✓	X

Figure 6

A graphical summary of the comparison between predictions for the proposed models and observations (a.k.a. “truth table”). A red cross shows a direct contradiction; a red cross with an asterisk shows a contradiction that may be avoided with relatively extreme fine-tuning or if the solution to that problem would violate another constraint; a green check mark denotes where the prediction of a model is consistent with observations; a green check mark with an asterisk indicates a situation in which the model can be brought into agreement with observations with a (potentially) reasonable assumption (i.e., some degree of fine-tuning is necessary); and finally a question mark indicates where a model has not been developed enough to make a reliable prediction. As can be seen, no model does particularly well when compared with observations. Abbreviation: AGB, asymptotic giant branch; EDA, early disc accretion; eSF, extended star-formation event; FRMS, fast-rotating massive star; GC, globular cluster; VMS, very massive star; YMC, young massive cluster.

driven wind in the direction unhampered by the disc. The enriched 2G stars are then predicted to form within this outflowing equatorial disc (i.e., a decretion disc).

- The N-C and Na-O anticorrelated pattern is quickly established in massive star interiors, although the details of chemical enrichment depend on the adopted reaction rates. The FRMSs are also able to process some Mg, which results in a production of Al at the expense of ^{24}Mg . However, this requires that the nuclear reaction rates for proton capture on ^{24}Mg are increased by three orders of magnitude (e.g., Decressin et al. 2007b). Using nominal reaction rates, FRMSs would produce a positive Al-Mg correlation, which contradicts the observed anticorrelation. Finally, the temperatures reached in massive star interiors are not high enough to build either the Si-Mg anticorrelation observed in a subset of clusters (Section 2.1.1) or variations in elements heavier than Al.
- Na and Al directly correlate with He, as observed (Section 2.1.2). However, the predicted He enhancement is significantly higher than the value allowed by observations (see Section 2.1.2; e.g., Bastian et al. 2015, Chantereau et al. 2016). However, because the NeNa reaction is very efficient, a large fraction of material in the massive star core does have the correct Na pattern before an extreme He enhancement is produced early in the life of the star. Thus, it is possible to reproduce the observed ΔY if some mechanism is able to increase the mass loss at critical rotation and halt self-pollution before large amounts of He are injected into

Decretion disc:

a disc made up of lost material around the equator of a rapidly rotating star

2P stars (i.e., if the core material can be accessed earlier than models predict). This would, however, introduce a high degree of fine-tuning.

- Discs in which 2P stars are forming must detach at a certain stellar mass/age (which varies from star to star depending on its initial mass and metallicity) to avoid pollution by He-burning products, i.e., to avoid a strong increase of C and O, which has not been seen in observations.
- Massive star ejecta are also Li free, so one must invoke some degree of dilution with unprocessed material to reproduce observations (see Section 2.1.3).
- Rotating massive stars would coexist with the supernovae (SNe) from single stars as well as with other massive stars. Hence, it is not clear how their discs can survive in the crowded central GC regions (e.g., Renzini et al. 2015).
- 2P abundances would have necessarily continuous distribution. The photometric and spectroscopic discreteness observed in some clusters cannot be readily reproduced by massive stars (Krause et al. 2013).

SDU:
second dredge-up
TDU: third dredge-up

3.2. Very Massive Stars

Denissenkov & Hartwick (2014) envisioned a scenario in which the most massive stars in the young cluster sink to the center as a result of dynamical friction. Shortly after they reach the center, the massive stars undergo multiple collisions with each other in a runaway process, eventually forming a VMS. VMSs with masses $\sim 10^4 M_{\odot}$ are predicted to be fully convective with luminosities close to the Eddington limit, allowing for a significant mass loss. Below are some important constraints on VMSs as the polluting stars.

- By the end of their MS lifetimes, VMSs are expected to reach very high He fractions, which would contradict the observed limits of ΔY in GCs today (Section 2.1.2). Hence, in order to stop the overproduction of He, it has been suggested that VMSs fragment (soon after formation), when only a small fraction of H was transformed into He. Thus, hot H-burning should occur only for a limited amount of time during the MS evolution on a VMS to reproduce the observed ΔY distribution, e.g., until the Y has increased to $Y \sim 0.4$.
- Although the observed anticorrelations and the Mg isotopic ratios—contrary to the cases of AGBs and FRMSs—are nicely reproduced, VMS nucleosynthesis cannot account for the observed Li (Section 2.1.3). Therefore, dilution is also required in this model.
- Only stars with masses in the mass range between 2×10^3 – $10^4 M_{\odot}$ have central temperatures that provide the observed GC light element anomalies up to Mg (e.g., Prantzos et al. 2017).
- VMSs have not been observed and their existence is still highly speculative. Also, due to the relativistic conditions required to model them, which in general have not been included in stellar evolutionary codes, their evolutionary and nucleosynthetic yields are also highly uncertain.

3.3. Asymptotic Giant Branch Stars

Processed material with some of the observed 2P chemical composition can be provided by intermediate-mass (~ 5 – $6.5 M_{\odot}$) AGB stars through a complicated interplay of nucleosynthesis and mixing episodes, namely the second dredge-up (SDU), the third dredge-up (TDU), and HBB (e.g., Karakas & Lattanzio 2014). Contribution by lower-mass AGBs should be avoided because AGBs less massive than $\sim 3.5 M_{\odot}$ would release enhanced C+N+O content into cluster ejecta.⁹

⁹Surface C+N+O enhancements are also predicted for rotating AGB stars more massive than $4 M_{\odot}$ (Decressin et al. 2009), which is contrary to observations of the majority of GCs.

COMPARISON OF ASYMPTOTIC GIANT BRANCH MODEL YIELDS

The chemical evolution of AGB star models greatly depends on the adopted input physics. Different treatments for convection and mass-loss recipes lead to variations of the HBB or TDU efficiency (among others) in the AGB models, indirectly changing the chemical yields. As a result, “the predictive power of AGB models is still undermined by many uncertainties” (Ventura & D’Antona 2005a, p. 1090).

Models based on the mixing-length theory (MLT) of low convective efficiency fail to reproduce most of the observed chemical anomalies (e.g., Fenner et al. 2004, Doherty et al. 2014). In particular, they predict HBB temperatures that are too low to allow for efficient ON processing; i.e., AGBs produce too much Na, and they do not provide large O depletion. Also, Mg and Al are positively correlated in the yields. 2P stars would also show an increase in the total CNO, which contradicts observations (Ivans et al. 1999).

Full spectrum of turbulence (FST) models are, compared to the MLT case, more consistent with observations on MPs. The FST model for turbulent convection results in a large convection efficiency, which translates into a very strong HBB episode (e.g., Ventura & D’Antona 2005a,b). Higher temperatures are reached at the base of the convective envelope, and stars evolve to higher luminosities with respect to the MLT case. As a consequence of the high luminosity and larger mass loss, they undergo a limited number of thermal pulses, so that the impact of TDU in changing the surface composition is limited. However, the lack of TDU in the FST models also limits the amount of Na that can be produced in AGB stars with $M \geq 5 M_{\odot}$, which reach temperatures so high that sodium is destroyed, providing a negative sodium yield. The theoretical yields may be reconciled with the observations only if we assume that the (uncertain) cross section of the main channel of sodium destruction is a factor of $\sim 2\text{--}5$ lower than the recommended values (Ventura & D’Antona 2006, D’Antona et al. 2016). Finally, in the FST case, the Mg isotopic ratios are expected to exceed (by far) unity in the more massive stellar models ($M \geq 4 M_{\odot}$), in contrast to what is observed (Yong et al. 2003). This problem is shared by the MLT model.

During the SDU the convective envelope extends into the H-exhausted region and mixes to the surface mostly He and N from the CNO cycling. Ashes from He-burning nucleosynthesis (mostly C and O, as well as Na and Mg) are eventually transported from the interior to the surface by the TDU, leading to an increase of the total C+N+O in the ejecta. Following each TDU episode, the H-burning shell is reignited until the next instability of the He-burning shell develops. This exchange of power between H- and He-burning shells along with the associated TDU episodes occurs many times during the AGB phase, and the overall changes in the surface abundances of AGB stars caused by TDU episodes strongly depend on mass, metallicity, mass-loss, etc.

Intermediate-mass stars also have envelopes that can reach very high temperatures (up to ~ 100 MK, with the maximum temperature reached being a function of the AGB mass) to activate hot H-burning. This process is known as HBB. As a result, the envelope is exposed to regions where hot H-burning takes place, until the temperature at the base of the convective envelope drops below ~ 20 MK (because of the mass loss that removes the envelope) at which point HBB is no longer supported. (See the sidebar titled Comparison of Asymptotic Giant Branch Model Yields.)

A summary of the ability of AGB stars to match observed MP abundances follows:

- Pollution from AGBs qualitatively reproduces some of the light element variations observed in 2P stars. However, it is not possible—without some modifications to the main physical inputs and relevant cross sections—to obtain simultaneous O depletion and Na enrichment and keep the C+N+O sum constant in AGB yields, as required by observations (e.g., Sneden 2000, Charbonnel 2016, Slemer et al. 2017). Indeed, the composition of the material ejected by AGBs through winds critically depends on what mechanism (either TDU or HBB) dominates. The net effect of TDU is the mixing of He-burning products to the

surface, in particular, C, Ne, and O. The HBB destroys O and produces Na by proton captures on the dredged-up Ne (note that the surface Na abundance first increases during the SDU). At the temperatures required to destroy Mg (\sim 100 MK), Na is destroyed again (e.g., Denissenkov & Herwig 2003). Thus, without the Ne dredged up by TDU and converted into Na by HBB, low values of O in the ejecta would lead necessarily to low Na for very high temperatures (e.g., Denissenkov & Herwig 2003). Na production can be increased by invoking an efficient TDU to effectively replenish Na by dredged-up Ne. However, this would lead to an increase of the overall CNO sum that is not corroborated by observations. Alternatively, Na destruction can be lowered by tweaking reaction rates (Renzini et al. 2015, D’Antona et al. 2016).

- The observed Li distribution is not reproduced, and dilution with a large amount of material, characterized by the same pristine GC composition (i.e., the same initial abundances and Fe), is needed (e.g., D’Ercole et al. 2016; see Section 5.1.1). Dilution is also required to obtain the observed anticorrelations (e.g., Na-O). As the cluster is >30 Myr before AGB stars evolve, it is not clear where this material would come from (see Section 5.1.4). The need to dilute AGB ejecta with unprocessed material also requires that material from the first massive stars exploding as SNe should be removed from the cluster, e.g., in order to avoid variable pollution from Fe-rich material resulting in [Fe/H] spreads.
- He-rich material is mixed into the surface via the SDU, whereas the TDU and HBB are responsible for changes in light elements. Thus, He, Na, and Al should not be strictly correlated in AGB yields (e.g., Charbonnel 2016).¹⁰ The He content of the ejecta is predicted to increase with stellar mass, and can reach He values up to $Y \sim 0.38$ in super-AGB stars (e.g., Ventura et al. 2013), which is less than that observed in some GCs.
- Because the temperatures reached during HBB are related to the envelope opacity and thus to the overall metallicity of clusters, the AGB model would naturally explain why the products of extreme nucleosynthesis (Mg depletion and Si and K production) are observed only in metal-poor clusters. However, it is not clear why many metal-poor clusters do not show these trends. The HBB temperature may be high enough to alter Si and K abundances in the most massive AGB models (e.g., Ventura et al. 2012). However, at such temperatures Na would be destroyed; i.e., 2P stars would have low Na abundance (Charbonnel 2016).
- Low-mass AGBs could potentially be responsible for the star-to-star variations in C+N+O and s processes observed in a handful clusters (Section 6). However, they cannot produce light element variations themselves (because of the competition between TDU and HBB).

4. THEORIES FOR THE ORIGIN OF MULTIPLE POPULATIONS

4.1. The Asymptotic Giant Branch Scenario

Early on in the development of this field, AGB stars were suggested as the source of the polluted material (e.g., Cottrell & Da Costa 1981). The “AGB scenario” is arguably the model that has gotten the most attention in the literature, and many aspects of the model have been included in other scenarios, even those that use different enrichment sources. Hence, we begin by discussing this model.

¹⁰Even if some initial Na enrichment during the SDU is expected, Na production due to the burning of dredged-up Ne also contributes to the resulting Na abundance. Thus, an obvious correlation between Na and He is not expected a priori.

IMF: initial mass function

4.1.1. Basic scenario. The model envisions the formation of a massive cluster with a single age and abundance pattern (i.e., a simple stellar population or SSP), representing a 1G of stars. The feedback from high-mass stars and the associated SNe clear any remaining gas from within the cluster, hence all enriched material from the high-mass stars and SNe is lost from the cluster (this is required to avoid Fe spreads). After ~ 30 Myr, stars from the 1G begin to evolve through the AGB phase of stellar evolution, and the winds of these stars, owing to their low velocity (~ 10 – 30 km s $^{-1}$; Loup et al. 1993), are not able to escape the cluster, so a reservoir of polluted gas begins to form in the cluster. This material cools and sinks toward the cluster center, and once a critical density is reached, a 2G of stars begins to form out of this material (e.g., D’Ercole et al. 2008, Bekki 2017). Early versions of the model had the 2G forming more or less continuously until star formation was truncated owing to the onset of rapid Type Ia SNe, which would clear the cluster of any remaining gas, at an assumed age of ~ 100 Myr. After the subpopulations within GCs were found to be largely discrete (e.g., M4; Marino et al. 2008), the model was refined by invoking multiple discrete bursts between the onset of AGB stars (~ 30 Myr) after the formation of the 1G and when Type Ia SNe began (e.g., D’Ercole et al. 2016).

It is worth noting that all AGB models to date do not produce an Na-O anticorrelation but rather a correlation. In order to reproduce the observed anticorrelations, this scenario requires the (re)accretion of large amounts of pristine material (i.e., material that shares the same abundances as the 1G stars) from the surroundings; i.e., dilution of the AGB yields with material that matches the initial chemical composition of 1P stars is required. In Figure 7, we show the basic idea of a dilution model. Combining the yields from the polluting stars (e.g., AGB stars) with material that matches the 1P stars, dilution tracks can be created to explain the run of chemical abundances observed within clusters, where a 2P star’s position is governed by the relative amount of processed material (i.e., AGB yields) and diluting material (1P chemistry) used to form the star.

This accreted material is then mixed with the AGB ejecta and forms 2G stars; this process is known as dilution, hence the 2G of stars would have different Na-O abundances ranging from the pure yields of AGB stars to those of the 1G. An additional problem for yields of AGB stars is that in the mass range of ~ 4 – 9 M $_{\odot}$ ¹¹ some models provide the Na enrichment and O depletion required to match observations (e.g., Ventura & D’Antona 2009), whereas other calculations have found that AGB stars are not able to produce the Na enrichment required (Doherty et al. 2014; see Section 3.3). Additionally, the latter models find that the C+N+O sum is not kept constant at any mass for AGB stars, which is in conflict with the observed properties of MPs in most clusters.

An important aspect of this—and most other—models, is that they can only produce a small fraction of the total cluster mass in 2G stars. This is due to the stellar initial mass function (IMF) of the 1G stars, which only has a small fraction of its total mass in stars in a specific mass range that can produce material to pollute/enrich the 2G of stars (i.e., $f_{\text{enriched,initial}} \sim 0.02$ – 0.1). In order to obtain the observed fractions of 1P and 2P stars ($f_{\text{enriched}} = 0.4$ – 0.8), the model needs to assume that GCs lose substantial fractions of their 1G stars, often up to 95% of their initial masses, while retaining all or most of the 2G stars.¹² This is further discussed in Section 5.4.1.

In the model envisioned by D’Ercole et al. (2008), the gas coming off of AGB stars is able to rapidly cool, mix with material (possibly accreted) with the same chemical abundance pattern as the 1G stars, fall to the center of the cluster, and subsequently form a 2G of stars. However, it is not clear whether such material would be able to cool and remain in the cluster. For example,

¹¹AGB stars of lower masses are generally disregarded as contributing to the formation of the 2G as they do not conserve the C+N+O sum, which contradicts observations (e.g., Ivans et al. 1999).

¹²Heavy mass loss is also required by the FRMS scenario (e.g., Schaerer & Charbonnel 2011).

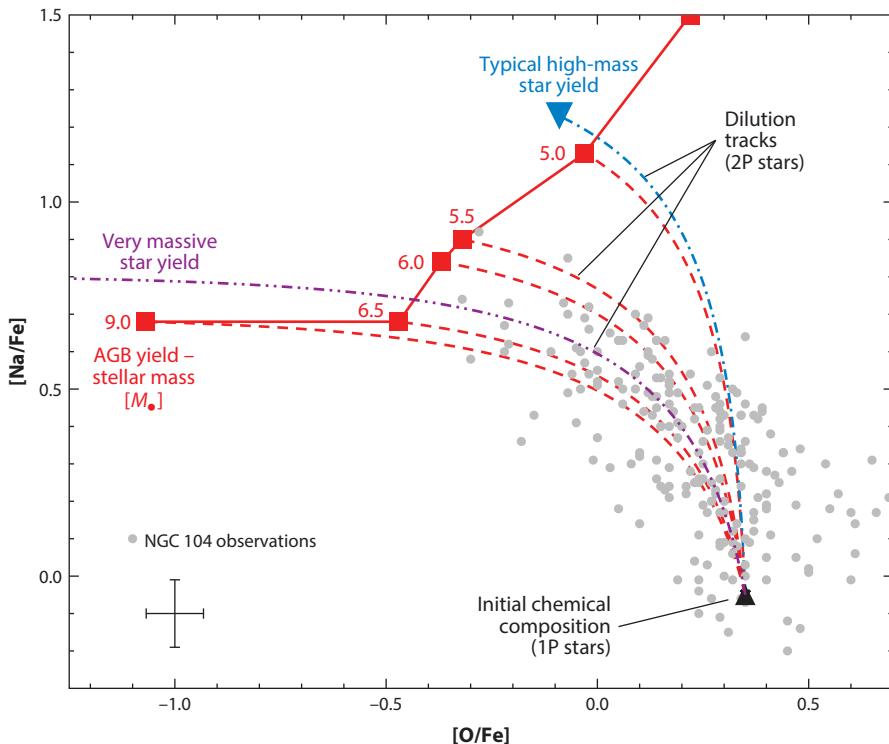


Figure 7

An illustration of a dilution model. The yields of suggested polluter stars are shown: AGB yields (data from D’Ercole et al. 2010) are shown with red squares for different masses (although note that other AGB yields do not show significant Na enhancement; Doherty et al. 2014); typical high-mass star ($\sim 20 M_{\odot}$) yields are given with a blue upside-down triangle (data from de Mink et al. 2009), and very massive star ($\sim 5 \times 10^4 M_{\odot}$) yields are shown (off to the left of the panel; data from Denissenkov & Hartwick 2014). Dilution models use these yields and then dilute them with gas that has the initial chemical composition (i.e., that of the 1P stars). This leads to dilution tracks where the 2P stars are located. All suggested pollution mechanisms require dilution (to various degrees) to explain the observed chemical abundances (i.e., He and Li; see Section 5). Also shown are data from NGC 104 from the compilation of Roediger et al. (2014). Abbreviation: 1P, primordial; 2P, enriched; AGB, asymptotic giant branch.

if the heating of a population of X-ray binaries is included in the simulation, the gas is unable to cool and instead flows out of the cluster. Conroy & Spergel (2011) have shown that the Lyman-Werner photon flux of stars of the 1G is high enough to not allow the gas to cool and sink to the cluster center until an age of 200–300 Myr, delaying the formation of a 2G of stars for a much longer period of time. Such a time delay would be a severe problem for the AGB scenario, as even under optimistic model yields, the C+N+O sum would not be conserved for AGB stars at this mass. Conroy & Spergel (2011) have also shown that, owing to the cluster’s motion within the galaxy, Bondi–Hoyle accretion onto the cluster is expected to be very inefficient, and the authors suggest that clusters can retain a relatively large fraction of their initial gas mass ($\sim 10\%$) to sweep up the interstellar medium (ISM) in order for the cluster to have the necessary primordial gas for dilution. This again ignores the role of heating from X-ray binaries and other mechanisms not included in standard SSP models, whereas if such sources are included clusters would be expected to be gas free, which is a substantial problem for this model (see Section 5.6). It is also not clear

that the material accreted from the surrounding galaxy would match the abundances of the 1P stars to the necessary precision imposed by the lack of Fe spreads in most clusters.

One of the features of AGB stars that make them promising candidates to supply the enriched material is the fact that they can burn H at higher temperatures than MS massive stars; the exact ranges depend on metallicity and mass of the AGB star (see, e.g., Prantzos et al. 2007, their figure 8). This allows them to activate the Al-Mg burning chain, hence, to deplete Mg and increase Al. As discussed in Section 2.1.1, a minority of clusters show significant Mg spreads, and most other potential polluting stars have difficulty producing the spreads without adjusting the nuclear cross sections in an ad hoc manner. By including dilution, the basic AGB model (for some model yields) is able to quantitatively match the observed Na-O anticorrelation with GCs and qualitatively match the increase in He. However, this model does not predict the correct abundance pattern of Li (as material processed through AGB stars should, to first order, be Li free) without invoking and fine-tuning a specific mechanism to produce Li (see Section 2.1.3).

In summary, the basic AGB model, though conceptually simple, has a number of shortcomings that subsequent works have attempted to address. This is explicitly addressed in Section 5.

4.1.2. Alternative versions. To avoid the problems associated with dilution (i.e., accreting the material and the associated timing constraints), Renzini (2013; see also Renzini et al. 2015) suggests that the yields of AGB stars may be very different from that predicted by current theoretical yields. Owing to the many parameters involved in estimating the yield of AGB stars (see Section 3.3), there is significant freedom when adopting AGB yields. The authors speculate that perhaps the true yields of AGB stars result in an Na-O anticorrelation so that no dilution would be necessary, although without dilution it would be very difficult to match the Li abundance patterns. Further work is needed to search the full range of potential parameter space of AGB model yields, but work so far suggests that AGB stars are not able to produce an anticorrelation of Na-O (e.g., Slemer et al. 2017). However, if this was true, it would add an additional factor of ~ 2 to the already strict mass-budget problem (which is discussed in detail in Section 5.4).

It has also been suggested that ancient GCs may have formed embedded in larger dark matter halos, allowing them to hold onto a larger fraction of the material ejected from evolving stars (e.g., Bekki et al. 2007, Trenti et al. 2015). If large/extended dark matter halos were necessary to form MPs, then we would expect that only the oldest ($z_{\text{form}} > 6$) GCs would be able to host MPs, as at lower redshift it would be increasingly unlikely to find a gas-rich dark matter halo that has not undergone significant star formation (where Fe spreads would be expected). The discovery of MPs in clusters younger than 8 Gyr ($z_{\text{form}} < 1$) argues against this type of scenario (Hollyhead et al. 2017, Niederhofer et al. 2017b).

4.2. Fast-Rotating Massive Stars and Interacting Binaries

Massive stars also undergo hot H-burning in their cores, during the MS, and as such are also potential candidates to provide the enriched material needed to form MPs. However, as this happens deep within the stars, it is difficult to bring up the enriched material to the stellar surface where it can be released into the GC intracluster medium. Massive stars that are rapidly rotating can overcome this problem, due to rotationally induced mixing that can cause, in extreme cases, the stars to be (nearly) fully mixed.

Decressin et al. (2007a,b) developed a scenario using FRMSs as the enrichment source. This scenario is similar to that of the AGB scenario, using the enriched material from a 1G of stars to form a 2G, but this happens when the cluster is much younger ($< 10\text{--}20$ Myr). As in the AGB scenario, the ejecta of FRMSs must also be diluted to match the observed abundance patterns

(typical yields and dilution are shown in **Figure 7**). However, because the cluster is still young there is no need to bring the material from outside the cluster, as it is assumed that the cluster has retained a relatively large fraction of gas/dust left over from the formation of the 1G. The winds of the FRMSs then mix with the leftover gas and forms a 2G of stars. The FRMS scenario suffers from the same mass-budget problem discussed for the AGB scenario (e.g., Schaefer & Charbonnel 2011).

FRMSs naturally produce an Na-O anticorrelation, and the enriched material can also be strongly enhanced in He, which helps explain clusters with large He spreads like NGC 2808. However, the high He yields may be a problem for more typical clusters with small He spreads (e.g., Chantereau et al. 2016). FRMSs are not able to activate the Al-Mg chain before the end of the MS, so they are not able to explain the observed Mg spreads in some clusters without ad hoc changes to the nuclear cross sections.

Krause et al. (2013) further developed the FRMS scenario by exploring cases in which a young GC may not be able to expel the leftover gas from the formation of a 1G of stars, even with SNe, allowing the cluster to remain embedded in its natal giant molecular cloud (GMC) for ~ 20 Myr. The authors suggest that the decretion discs might also accrete material from the gas-rich intracluster medium, which would solve the dilution requirements.

Charbonnel et al. (2014) presented a variation on the FRMS scenario to solve the mass-budget problem (see Section 5.4). Here, the 1G stars form with a top-heavy stellar IMF (i.e., only stars that would not be alive today), and the 2G would consist mainly of low-mass stars. In this model, stars with primordial composition (i.e., 1P stars) would be actually 2G stars that formed primarily from material left over from the 1G. Such a model can be tested through carbon isotopic ratios of MS stars.

Another way to release enriched material from the cores of massive stars into the intracluster medium is through binary interactions. de Mink et al. (2009) modeled a binary interaction between a 20- and 15- M_{\odot} star and investigated the yields of the expelled material. They found that the 20- M_{\odot} star shed about $10 M_{\odot}$ worth of material because of the interaction and that the yields matched the observed trends in GCs (i.e., Na enriched, O depleted, etc). Although the overall trends and correlations of the yields should apply to most massive stars, the exact yields depend on a number of parameters, e.g., the time of interaction (i.e., stellar evolutionary state), total mass of the stars, and the mass ratio of the stars. Hence, interacting binaries have the benefit of potentially explaining the observed variations from cluster to cluster but have difficulty matching the discreteness of abundance ratios found in many subpopulations.

A potential problem of scenarios that operate in the first few million years of a cluster's life is that after 3–8 Myr (depending on the cutoff mass for SNe), core collapse SNe begin to explode. The retention of just a small amount of this material results in Fe spreads that are in conflict with observations (Renzini 2008). Hence, processes that are limited in time to the epoch before the first core collapse SNe may be needed in such models.

Szécsi et al. (2018) proposed a variation on this scenario, in which a 2G of stars form in shells around high-mass (150–600 M_{\odot}) red supergiant stars (RSGs). This scenario suffers from problems similar to the FRMS scenario (in terms of abundances, discreteness, and mass budget), but it is also only expected to operate at low metallicity. Because MPs are found in GCs of all metallicities ($-0.3 > [\text{Fe}/\text{H}] > -2.5$), this scenario could only apply to a subset of the known GCs.

GMC: giant molecular cloud

RSG: red supergiant star

4.3. The Early Disc Accretion Scenario

Bastian et al. (2013b) suggested an alternative model for MPs that did not invoke multiple epochs of star formation. Instead, it was driven largely by the constraints posed by YMCs (see Section 5.6).

The model used the enriched material ejecta from high-mass IB stars (de Mink et al. 2009) as well as the FRMSs within the cluster to pollute low-mass stars that formed at the same time as the high-mass stars. The authors suggested that low-mass ($<2 M_{\odot}$) stars may retain the protoplanetary discs around them for ~ 10 Myr, which would sweep up the enriched material as they passed through the cluster core (the authors also assumed that the cluster is mass-segregated from a very early age, so that the high-mass stars are concentrated in the cluster center). The enriched material that was swept up by the discs would then eventually be accreted onto the host star.

Although this scenario matches most observations of YMCs, it has a number of shortcomings as well (see Section 5). In particular, it requires that the accreting stars be fully convective (in order to mix the accreted material throughout the star), which in turn means that the accretion timescales are extremely short (1–3 Myr; Salaris & Cassisi 2014, D’Antona et al. 2015). This minimizes the time that the mechanism could potentially work, which effectively limits the amount of processed material that can be supplied and accreted.

Wijnen et al. (2016) ran hydrodynamical simulations to test this scenario, placing a realistic protoplanetary disc in a “wind” of material (i.e., the ejecta of IB stars, where the “wind” refers to the disc moving through the intracluster ISM). They found that while the disc did indeed accrete material from the ISM, the accreted material had little or no angular momentum, which caused the disc to rapidly accrete onto the star and disappear. Without the disc, no further accretion would be possible. The authors found that this happened on a rapid timescale, $\sim 10^4$ years, which is much shorter than the required 10^7 years for the scenario to work.

4.4. Turbulent Separation of Elements During Globular Cluster Formation

Hopkins (2014) also put forward a potential origin of MPs that did not invoke multiple generations of star formation within GCs. In his scenario, MPs would be the result of cloud physics during the earliest phases of GC formation. In extremely turbulent environments, like those in progenitor clouds of GCs, large dust grains can become aerodynamic and begin to move separately from the gas and small dust grains. Large resonant fluctuations in the dust can then develop. Within these overdense regions, dust is overrepresented, so any stars that form within such regions are enhanced in the elements associated with large dust grains. However, the gas and small dust grains (like Fe grains) are more uniformly distributed. In principle, this mechanism provides a natural and powerful way to separate elements in the early phases of GC formation. Because this mechanism depends on the level of turbulence, it would predict larger abundance spreads in more massive proto-GC clouds, consistent with observations.

However, as noted by the author, Na and O normally occur on the same dust grains, so such fluctuations would predict an Na-O spread but as a correlation instead of the anticorrelation seen in GCs. Also, He is not affected by dust, so an additional mechanism would need to be invoked to explain the inferred He spreads in GCs. Finally, any enhancement in an element in some stars would necessarily lead to a depletion of that element in other stars. We would then expect to see, starting from field star abundance composition, more or less symmetrical spreads around the field star abundance. However, observations show the scatter in a single direction from the position of where halo field stars lie (at a given metallicity).

4.5. Reverse Population Order for Globular Cluster Formation Scenarios

To alleviate the mass-budget problem (which is discussed in Section 5), some authors tentatively investigated formation models in which the abundances of forming stars move from 2P to 1P, as star formation within the cluster proceeds (e.g., Marcolini et al. 2009).

The scenario outlined by Marcolini et al. (2009) envisions GC formation from gas enriched locally by a single Type Ia SN and AGB yields superimposed on an ambient medium pre-enriched by low-metallicity Type II SNe. The star formation of the proto-GC only takes place inside this region, and stars born within the inner volume are depleted in O and Mg (because of the single SN Ia) and enhanced in N, Na, and Al abundances (due to AGB pollution). External to this volume can be found a region with the same composition as the proto-halo gas at the epoch of GC formation. After a new generation of stars is born, associated SNe II begin to pollute and expand the inner volume while mixing with the lower metallicity material from the external shell, i.e., gas with pristine composition. Hence, the [Fe/H] and the CNO sum remain constant during cluster evolution, and the N-C and Na-O anticorrelations can be reproduced. The Al-Mg anticorrelation can only be reproduced assuming that AGBs produce more Al than predicted by models (by a factor of \sim 10–50; e.g., Karakas & Lattanzio 2007).

In a following paper, the authors focus on other elements and achieve some success in reproducing the observed trends (Sánchez-Blázquez et al. 2012). Nonetheless, the dynamical feasibility of the scenario has not been probed with hydrodynamical simulations, and severe assumptions need to be made on the Fe content of the ISM at the epoch of formation as well as on the size of the inner region where the inhomogeneous pollution by the SN Ia and AGBs is confined (e.g., Sánchez-Blázquez et al. 2012). More importantly, this class of models requires very peculiar stellar configurations that are not expected at the present epoch (e.g., Conroy 2012).

4.6. Extended Cluster Formation Event

Elmegreen (2017) has further explored a model put forward by Prantzos & Charbonnel (2006) that invokes the special conditions of galaxies or GMCs at high redshift (namely high density, turbulence, and pressure environments) to foster the formation of MPs before the first SNe occur (<3 Myr). Here, a 1G SSP is born in the core of a massive, dense, and turbulent GMC. Because of the high stellar densities, high-mass stars have their envelopes stripped (and rotating massive stars lose large parts of their envelopes through decretion discs) very rapidly, which (as discussed above) are expected to show many of the observed abundance anomalies. This material mixes with that left over from the formation of the 1G and forms subsequent generations. Low-mass 1G stars are assumed to be ejected owing to two mechanisms: The first is binary dynamics, and the second is that the gravitational potential of the cloud core/cluster is rapidly varying as the gas within it (which dominates the potential) is moved because of stellar feedback. It remains to be seen if the high 1G mass-loss rates (and required low 2G mass-loss rates) required are feasible.

Wünsch et al. (2017), following on Tenorio-Tagle et al. (2005), have suggested that the winds released from massive stars can become so dense in a massive and dense young cluster that they enter a catastrophic cooling regime and can collapse into the cluster center. Here, the material may mix with left over primordial material (i.e., dilute) and form 2G stars. Hence, this is another mechanism (rather than stellar interactions) that can potentially make enriched material from massive stars available for further epochs of star formation within a cluster. This also suffers from the mass-budget problem and would require large fractions of 1G stars to be lost. Lochhaas & Thompson (2017) develop this model further in terms of chemistry and show that the model is not able to simultaneously account for the increasing enriched fraction and increasing chemical spread with increasing cluster mass (see Section 5.5).

As these scenarios invoke massive stars, we include it in our comparisons with observations, in particular the abundance trends, with other scenarios that invoke massive stars (Section 5).

A key aspect of this scenario is that it happens (and terminates) before the first SNe occur within the proto-GC in order to avoid Fe spreads (similar to the FRMS scenario). One potential problem

with the scenario is that it takes high-mass stars some time to increase their He mass through nuclear burning, whereas this model starts using stripped material from the massive stars at $t = 0$. This may be all right for standard clusters with small He spreads (e.g., NGC 104), but it may be difficult for this model to reproduce clusters like NGC 2808, which hosts a large He spread.

Finally, for the limited models available of interacting binaries and fast-rotating mass stars, it is not clear that they will be able to provide the stochasticity (i.e., the specific abundance pattern—extrema, discrete subpopulations—for each GC) required to match the observations. Elmegreen (2017) suggests that subclumps may form within the proto-GC, and each subclump would have its own chemistry due to the exact chain of stellar interactions. However, these subclumps would each be expected to be $>10^4 M_{\odot}$, where the stellar IMF is fully sampled, hence stochastic effects would be expected to be minimized. The Wünsch et al. (2017) scenario suffers from the same problem.

4.7. Very Massive Stars Due to Runaway Collisions

Gieles et al. (2018) have developed a model for MPs that adopts VMSs ($>10^3 M_{\odot}$) as the origin of the processed material. In this model, the proto-cluster undergoes adiabatic contraction due to gas accretion, increasing the stellar density and subsequently the stellar collision rate. A runaway collision process can form a VMS, which releases hot H-burning processes through its stellar wind into the intracluster environment (Denissenkov & Hartwick 2014). This processed material mixes with pristine gas (i.e., gas with the same abundance pattern as the initial protocluster) and forms further generations of stars until the VMSs burn out or potentially explode owing to instabilities within the star. Because the VMS can be continuously rejuvenated through stellar collisions, the amount of processed material ejected by the star can be several times the maximum mass of the star. Although this process leads to multiple generations of stars within the cluster, the expected age spread would be less than ~ 3 Myr.

One major advantage of this model is that it predicts a superlinear scaling between the mass of the VMS and the mass (or density) of the cluster. This naturally produces the observed trend of increasing fractions of enriched stars (and potentially as well as the increasing spreads in N, Na, etc.) as a function of GC mass. This kind of model also does not violate the constraints from YMCs, and much of the expected abundance patterns also appear to match observations.

One of the major drawbacks of the model is that VMSs are still only theoretical, although the authors perform numerical simulations showing that under certain conditions (relevant for GC formation), runaway collisions are likely to take place, even when considering two-body relaxation and the strong stellar mass loss of the massive object due to its stellar wind. This same process is expected to also be at work in clusters today, if they reach the required stellar densities. Hence, it is not clear if the model can explain why NGC 1978 (~ 2 Gyr) hosts MPs while NGC 419 (~ 1.5 Gyr) does not, given their similar masses and radii.

5. COMPARING PREDICTIONS WITH OBSERVATIONS

5.1. Chemical Abundance Patterns

One of the most common tests of the proposed scenarios for MPs is through abundance measurements. However, most tests done to date have just focused on a single element (or pair) in comparison with model yields. In this section, we compare observations with model predictions, concentrating on individual (e.g., Li) and groups of elements that are particularly constraining.

5.1.1. The need for dilution. As discussed in Section 4, the suggested stellar sources for the origin of the polluted material present difficulties in reproducing some of the observed abundance

trends. For example, AGB model yields suggest that Na and O should be correlated, not anticorrelated. Also, wherever nuclear processing of C, N, O, and Na takes place, the resulting material is expected to be Li free, whereas observations show that Li is constant or only slightly varying in GCs from star to star. To address these problems, most models have adopted some form of dilution, i.e., that the enriched material produced by 1P stars is mixed with material that matches the chemistry of the 1P stars (referred to as primordial material). Here, we discuss the predictions of dilution in comparison with observations. A basic illustration of a dilution model is given in **Figure 7**.

5.1.2. Lithium variations. Without dilution, we would expect all 2G stars to be effectively Li free, as any material subjected to hot H-burning will have its Li rapidly destroyed. Observations show, however, that in some GCs Li is constant between 1P or 2P stars, or that it is depressed in 2P stars, i.e., anticorrelated with Na and correlated with O (see Section 2.1.3). The amount of Li would then reflect the amount of diluted material included in the formation of 2P stars. This assumes that Li is not produced by other processes. In principle, AGB stars can produce some amount of Li through the Cameron–Fowler mechanism (Cameron & Fowler 1971), but this requires extreme fine-tuning to match the observe Li variations/constancy (see Section 2.1.3).

Salaris & Cassisi (2014) have pointed out difficulties in such dilution models. Essentially, because the enriched material is expected to be Li free while only depleted in O, the spread in Li should always be larger than the spread in O. However, for at least one cluster, NGC 6752, the spread in Li is smaller than the spread in O. The Li spread (in relation to Na, O, and other light elements) needs to be studied in other clusters, but if these results are confirmed this poses a major problem for all models that use high-mass stars (i.e., $> 15 M_{\odot}$) as well as AGBs.

There are tentative hints that the amount of Li variation is larger in higher-mass clusters, which is similar to what is observed in Na, O, He, and N. Indeed, in high-mass and metal-poor clusters, stars characterized by extreme composition (very high Na and Al enhancement) are also Li poor (e.g., NGC 1904, NGC 2808, NGC 6752, M5, NGC 6397; see D’Orazi et al. 2015, and references therein). The presence of a fraction of 2P stars with depleted Li abundance is surprising, because 2P stars with an intermediate degree of chemical variations share the same Li abundance as 1P stars. If the light element anomalies are produced by nucleosynthesis in the interior of stars, this finding implies that some mechanism (either dilution or Li production by AGBs) should operate to restore the Li abundance of its initial value in 2P stars with intermediate composition without changing Li in extreme 2P stars. Again, such an interpretation requires extreme fine-tuning.

5.1.3. Quantitative abundance trends and the need for stochasticity. Although many studies have compared the observed abundance distributions of specific clusters with the yields of potential polluter stars, few have carried out a more general analysis including multiple elements and comparisons between clusters. Bastian et al. (2015) studied a sample of eight Galactic GCs that all had measurements of their Na–O anticorrelations as well as spreads in He based on HST imaging. With the exception of NGC 2808, the authors conclude that the observed distributions (Na, O, He) were not in agreement with the predicted yields of AGBs, FRMSs, IBs, or VMSs, even when dilution with 1P material was taken into account. Specifically, based on the extent of the Na–O anticorrelations, large He spreads ($\Delta Y > 0.1$) would be expected in all cases, whereas in most cases $\Delta Y_{\text{obs}} < 0.05$.

Bastian et al. (2015) also considered “empirical yields,” i.e., adopting the observed Na–O anti-correlation and He spreads observed for a given cluster, and comparing them to the other GCs in the sample. Surprisingly, even when using the empirical yields a satisfactory fit for the other clusters could not be reached (even when controlling for metallicity). The conclusion is that whatever the polluting source, it needs to produce a high degree of cluster-to-cluster variations in order to

explain the observations. Dilution of a fixed set of yields does not help in explaining the full set of observations. This argues against the stellar sources normally considered (i.e., AGBs or massive stars) being the origin of the enriched material, as none of these can provide the necessary cluster-to-cluster variation. However, the multimodal abundance patterns within GCs suggest that for a given GC, the yield/dilution combination is quite uniform (i.e., taking on only a handful of values within the cluster).

It is beyond the scope of this review to quantitatively compare the yields of each proposed source with observed measurements for each element, especially considering that most works to date have only focused on one or two elements at a time (i.e., not testing whether the yields and required dilution that match a given element are able to match another). We refer the interested reader to, e.g., D'Antona et al. (2016) or Prantzos et al. (2017).

5.1.4. The origin of the diluting material. For models that adopt massive stars as the origin of the enriched material ($>15 M_{\odot}$), it is assumed that a large reservoir of primordial material is leftover within the cluster from the formation of the 1G of stars. However, for models that invoke pollution from AGB stars, the origin of the diluted material is more difficult to explain. Once core-collapse SNe from the massive stars in the 1G begin to explode, all material left over from the formation of the 1G is expected to be removed to large distances (i.e., unbound from the cluster). Hence, the 1P material must then be (re)accreted from the surroundings. This material must also avoid being contaminated with the material (e.g., Fe) from the SNe, or else Fe spreads would be expected in all clusters (e.g., Renzini et al. 2015).

Conroy & Spergel (2011) suggested that this material can be accreted from the host galaxy as the clusters orbit through the ISM. Although accretion due to gravitational focusing is not efficient for the majority of cases, the authors found that if a reservoir of gas already exists within the cluster ($\sim 10\%$ of the stellar mass), it can sweep up material and the reservoir can grow. However, D'Ercole et al. (2011) have shown that this near-constant accretion of new material, when coupled with the adopted AGB yields, does not reproduce the observed abundance distributions. For the AGB model to work, the timing of the dilution needs to be very specific, with nothing being accreted (i.e., no diluting material present) when the most massive AGB stars are shedding their material, and there must be an ever increasing amount of material being accreted after that, until the process is terminated, potentially by the onset of Type Ia SNe.

D'Ercole et al. (2016) have further developed the basic AGB scenario by placing the YMC inside a disc galaxy. In the model, the SNe from the 1G of stars blow a hole in the surrounding ISM, and eventually the expelled material is lost to the host galaxy. These authors adopt the same basic scenario as D'Ercole et al. (2008), that the young cluster can retain the ejecta of AGBs and that this material can cool and form a 2G of stars within the cluster. Eventually, the SNe's blown bubble begins to close (as SNe become less frequent), and material from the galaxy fills the hole, some of which is then accreted back onto the cluster. This scenario requires the surrounding material (out to hundreds of parsecs) to be chemically identical to that of the 1G stars within the cluster. Additionally, this model does not take into account the motion of the cluster within the host galaxy, in particular the high velocity dispersion expected in young galaxies (c.f., Kruijssen 2015), hence it is not clear that the gas would be accreted onto the cluster. Note that massive clusters ($>10^6 M_{\odot}$) in galaxy mergers today do not appear to be able to efficiently accrete material from their surroundings (Cabrera-Ziri et al. 2015, Longmore 2015).

5.1.5. Aluminum-magnesium anticorrelation. Interestingly, the presence of Al and Mg anti-correlated ranges among cluster stars is one of the strongest arguments against the FRMS scenario, as the temperature required to efficiently destroy ^{24}Mg is reached in the core of massive stars only

at the very end of their MS evolution (Decressin et al. 2007b). As a consequence, a large increase (by a factor 1,000) of the $^{24}\text{Mg}(\text{p},\gamma)$ reaction rate around 50 MK with respect to the nominal values is demanded to build the Al-Mg anticorrelation in the stellar core, and even in that case Mg depletion would be associated with a strong He enrichment (up to $Y \sim 0.8$ after dilution with unprocessed material; see Chantereau et al. 2016). Pollution from AGBs would in principle more naturally reproduce the observations, because both the depletion of Mg and the production of Al are sensitive to AGB metallicity, in the sense that more extended Al and Mg variations are expected at low metallicity, as observed (Ventura et al. 2016). However, the resulting (anti)correlations between Na, Mg, Al, and Si are greatly dependent on the mixing with He-burning material; e.g., because of the competition between TDU and HBB (see Section 3). Finally, the observed dependence of Mg depletion and Al production on metallicity can be explained in the VMS scenario if the mass loss leads to the formation of smaller VMSs at higher metallicities (e.g., Vink et al. 2011).

5.2. Discrete Versus Continuous Abundance Spreads

In the majority of GCs, 1P and 2P stars are observed to be distributed continuously in the Na-O plane. However, a number of studies revealed that the O, Na, and Al abundances of different subpopulations are clustered around certain values (e.g., Marino et al. 2008, Lind et al. 2011, Carretta 2014, 2015). Nonetheless, the evidence of multimodality from high-resolution spectra is still sparse. On the contrary, C and N (and CN band strength) multimodality is almost universal among clusters with intermediate to high metallicity (e.g., Norris 1987).¹³ HST photometry, in particular when including UV filters, also shows largely discrete RGBs and MSs in some cases (see Section 2.2 and Figure 3). These findings indicate that the spectroscopic Na-O distributions may also be made up of discrete groups of stars but that errors have blurred the distinction between the groups, causing the distribution to appear continuous (e.g., Carretta et al. 2013, Carretta 2015).

The observed discreteness between two (or more) subpopulations would disfavor formation scenarios based on accretion onto preexisting stars (e.g., EDA scenario) or 2P stars being born within the disk of an FRMS (e.g., FRMS scenario). Such processes would result in a continuous range of abundance variations rather than the discrete distributions demanded by the observations.

5.3. Radial Distributions, Velocity Dispersions, and Binarity

The evidence of a more centrally concentrated 2P (see Section 2.4.1) is in qualitative agreement with most of the proposed scenarios. Also, the higher incidence of binaries with 1P composition (D’Orazi et al. 2010, Lucatello et al. 2015) would again be consistent with a 2P preferentially found toward cluster inner regions. An example is the D’Ercole et al. (2008) scenario, in which the AGB ejecta form a cooling flow and rapidly collect toward the cluster center, forming a concentrated 2P. The system starts with more concentrated 2P stars; as the cluster evolves, the 1P and 2P stars mix. The long-term dynamical evolution of the different subpopulations with initial spatial segregation allows for efficient mixing in the innermost regions, where the local two-body relaxation timescale is shorter, potentially erasing any initial differences between subpopulations on a relaxation timescale (e.g., Vesperini et al. 2013).

However, there are a handful of exceptions to this general behavior, with the 2P stars being less centrally concentrated than 1P stars (Larsen et al. 2015, Vanderbeke et al. 2015, Lim et al. 2016). Although differences in mass between the 1P and 2P stars due to He variations offer a

¹³The bands of bimetallic molecules like CN are weak in metal-poor GCs, because their strength has a quadratic dependence on the metallicity.

potential explanation, the required He spreads are much larger than can be accommodated by the observations (Larsen et al. 2015).

Different formation models may leave unique kinematics imprints that would allow for distinguishing between various scenarios (i.e., different subpopulations showing different flattening; e.g., Mastrobuono-Battisti & Perets 2013). In this regard, the differential rotation of subpopulations provides precious insights, as such an observational property may survive the long-term dynamical evolution of old GCs and would allow us to distinguish different formation scenarios (e.g., Bellazzini et al. 2012, Hénault-Brunet et al. 2015, Cordero et al. 2017).

5.4. The Mass-Budget Problem

A difficulty of all the proposed self-enrichment scenarios that was quickly realized was that because the enriched population within GCs was equal to, or larger than, the primordial population (i.e., $f_{\text{enriched}} > 0.5$), there simply would not be enough material processed through 1P stars to explain the number of 2P stars if standard stellar IMFs are adopted (e.g., Prantzos & Charbonnel 2006). This is known as the mass-budget problem. For example, for a standard IMF, only $\sim 7\%$ of the stellar mass in a 1G is in stars with masses between $5\text{--}9 M_{\odot}$ (i.e., stars that pass through the AGB phase often associated with the AGB scenario). However, low-mass ($<0.8 M_{\odot}$) stars make up $\sim 40\%$ of the initial mass fraction. For a typical GC, 2P stars represent $\sim 67\%$ of cluster stars, whereas 1P stars make up the remaining $\sim 33\%$. Assuming that 100% of the mass of every AGB star gets used to make 2P stars (an extreme assumption) and that the 2P has a standard IMF, AGB stars can only account for 4–5% of the population of 2P stars. If we assume that, on average, 50% of the mass of each 2P star comes from diluting material, then AGB stars can account for 8–10% of the 2P stars.

The commonly invoked solutions to this problem have been (*a*) to apply an ad hoc limit to the mass range of 2G stars to $<0.8 M_{\odot}$, i.e., the mass range observed in GCs today (giving a factor of ~ 2.5 , i.e., accounting for $\sim 20\%$ of the needed mass), and (*b*) to assume that the number of 1G stars was much larger when the cluster formed, and that $\sim 90\text{--}95\%$ of them have been lost during the evolution of the cluster. These lost stars would then populate the field of the host galaxy.

For the AGB scenario, D’Ercole et al. (2008) and Conroy (2012) estimate that GCs must have been at least 10–20 times more massive than observed today. GCs are expected to be 8–25 times more massive in the FRMS scenario as well. Cabrera-Ziri et al. (2015) discuss this problem in detail and conclude that under more realistic assumptions the problem may be a factor of 2–3 times worse than previously suggested (i.e., requiring clusters to have been $\sim 30\text{--}60$ times more massive at birth than they presently are). This is a basic prediction of these scenarios that can be directly tested observationally.

5.4.1. Internal mass-budget problem. We refer to the internal mass-budget problem to mean the relative numbers of 2P and 1P stars within GCs. For a standard stellar IMF, only a small fraction (2–8%) of the 1G mass is processed through a given stellar type (e.g., AGBs, FRMSs, IBs) and released into the intracluster medium, even for optimistic yields. However, the present-day observed f_{enriched} for clusters is 40–90% (e.g., Milone et al. 2017b), and a significant amount of processed mass is needed for each of the enriched stars. The standard solution to this problem is to assume that GCs were 10–100 times more massive at birth than they are currently, and that, because the 2G stars are thought to be born more centrally concentrated, a large fraction of the 1G stars were lost during their evolution.

Vesperini et al. (2010) have simulated the evolution of such a cluster in a Galactic-like potential and found that, in principle, with the right selection of parameters, such extreme mass loss can be reproduced with numerical models. However, in order to obtain such extreme mass loss, the

authors needed to assume that GCs began their lives tidally limited and mass segregated, so that they expand owing to stellar mass loss and lose stars to the galaxy over their tidal boundaries. The clusters would then start their lives with effective radii of tens to hundreds of parsecs (depending on the strength of the tidal field at birth), although it has not been demonstrated that such clusters would resemble the observed Galactic GCs after \sim 10 Gyr of evolution. Present-day GCs and YMCs have much smaller effective radii, with means around \sim 3 pc (Harris 1996, Larsen 2004). Additionally, it is not clear that such a mechanism would work in environments with weaker tidal fields (that display similar f_{enriched} as Galactic GCs) like that of GCs in the LMC/SMC or the Fornax dwarf galaxy.

Bastian & Lardo (2015) and Milone et al. (2017b) both looked at the f_{enriched} as a function of the Galactocentric distance. If large fractions of 1P stars are lost due to the tidal field, even in the case of tidally limited and mass segregated initial cluster conditions, there would be a strong expected relation between f_{enriched} and the Galactocentric radius (see Bastian & Lardo 2015). However, f_{enriched} was not found to depend on the Galactocentric distance (or orbit), in contradiction with predictions from scenarios that invoke heavy mass loss. Milone et al. (2017b) have found that f_{enriched} is a strong function of present-day GC mass, with higher-mass GCs having larger f_{enriched} (see **Figure 4**). This trend is opposite to what would be expected if GCs underwent large amounts of mass loss. Higher-mass clusters are expected to lose a lower fraction of their mass during their evolution, hence they should have enriched fractions closer to the initial value.

Kruijssen (2015) estimated the mass lost from GCs forming and evolving in a cosmological context and found that massive GCs (with initial masses $> 5 \times 10^5 M_{\odot}$) are only expected to lose a relatively small fraction of their initial masses (i.e., potentially being a factor of \sim 2–4 more massive than currently seen). This is largely in agreement with non-MP driven estimates of mass loss from Galactic GCs (e.g., Kruijssen & Mieske 2009) and with constraints from the shape of the lower-mass function in clusters (which is sensitive to mass loss; e.g., Webb & Leigh 2015).

5.4.2. External mass-budget problem. We refer to the external mass-budget problem to mean the number of primordial stars in GCs relative to that of the host galaxy. This is linked to the internal mass budget if one adopts models in which large fractions of 1P stars are lost to the field. In principle, one should find an excess of 1P stars in the halo that came from GCs, at the position of the donor GCs in phase space (i.e., position, velocity, and/or metallicity; e.g., Schaefer & Charbonnel 2011).

The number of GCs found (per unit galaxy mass or luminosity) is known to be high in some dwarf galaxies (e.g., Larsen et al. 2012). It becomes even higher at low metallicity (e.g., $[\text{Fe}/\text{H}] < -1$ dex) when GCs and field stars of the same metallicity are compared (e.g., Harris & Harris 2002). Larsen et al. (2012) have exploited this observation to place some of the strictest constraints on the origin of MPs to date. The authors counted the number of 1P stars in GCs in the Fornax dwarf galaxy below $[\text{Fe}/\text{H}] = -2$ dex and compared that with the number of stars observed in the field in the same metallicity range. They found that GC stars made up \sim 20–25% of the stars in this metallicity range. Even if all stars in this metallicity range formed in clusters, this would mean that these GCs could have only been a factor of 4 or 5 more massive than they currently are, in contradiction with the requirements of models requiring large mass loss. Larsen et al. (2014a) have extended this kind of study to the dwarf galaxies WLM and IKN and found similar results, showing that this is a common phenomenon and not linked to the specific evolutionary history of the dwarf galaxy host.¹⁴

¹⁴In fact, the high specific frequencies observed in many dwarf galaxies (e.g., Harris et al. 2013) argues against these heavy mass-loss scenarios, assuming that GCs in dwarfs also host MPs (i.e., that MPs are ubiquitous).

Khalaj & Baumgardt (2016) have suggested that, in the context of the FRMS scenario, the expulsion of gas (left over from the formation of the 1G and 2G stars) from the young GC could unbind large fractions of stars from the cluster at high velocity. If the stars leave with a large enough velocity they could potentially leave the young galaxy all together. Note that this solution would not be applicable to the AGB scenario, as the cluster would already be gas free when the AGB stars begin to evolve. Although possible, observations of YMCs today do not support the idea that gas expulsion leads to large mass loss within clusters (c.f., Longmore et al. 2014).

5.5. Trends with Cluster Properties

As discussed in Section 2.5, the present-day mass of a GC is directly linked to (*a*) the fraction of enriched stars present and (*b*) the extent of the abundance spreads in N, Na, O, and He (see **Figure 4**), with higher-mass clusters having larger enriched fractions and larger spreads. Assuming that the yields of the polluting source (e.g., AGBs, FRMSs, VMSs, etc.) are not dependent on the GC properties, the link is difficult to explain in the classic scenarios as stellar yields (for a fully sampled IMF) should provide a constant amount of enriched material per unit stellar mass.

The increasing fraction of enriched stars at higher masses is contrary to the expectations of scenarios that invoke heavy mass loss to obtain large (present-day) fractions of enriched stars (e.g., AGB or FRMS scenarios; see Sections 4 and 5.5) as it would require higher-mass clusters to lose larger fractions of their mass (i.e., large numbers of 1P stars), which is opposite to the expectations from basic dynamical considerations (e.g., Kruijssen 2015). Additionally, if GCs did lose large fractions of their initial masses, it would be extremely difficult to maintain these strong correlations with cluster mass (e.g., Schiavon et al. 2013). It is also unexpected that higher-mass clusters should show larger abundance spreads. Though in principle they may hold onto more of the processed material, they also should accrete/retain more primordial material (i.e., diluting material). Additionally, models already assume that all of the processed material is used in the formation of 2P stars (i.e., 100% star-formation efficiency of the processed material).

Although it is difficult to reconcile models to these observations, it is worth noting that no model put forward to date is able to account for both the fraction of enriched stars and the extent of the variations as a function of cluster mass at the same time. This is because, for the polluting sources suggested, the amount of enriched material produced is fixed per unit mass. The model can use that enriched material to create either larger abundance spreads (i.e., putting more of it in 2P stars) or more 2P stars (i.e., increasing f_{enriched}) but not both. The conclusion reached is that the enrichment mechanism must depend on the mass (or density) of the host cluster.

5.6. Constraints from Young Massive Clusters

One of the major discoveries made by the HST was that stellar clusters with masses and densities rivaling (and in some cases, greatly exceeding) GCs are still forming in the local Universe (see Portegies Zwart et al. 2010). These YMCs are commonly referred to as proto-GCs as they have similar properties to those expected for the present-day GCs when they were young (e.g., Kruijssen 2014). Owing to their proximity and relative brightness, we can use YMCs to test the scenarios for the formation of GCs and the MPs within them. The properties of YMCs themselves are discussed in Section 7.

Although MPs have not been found to date within YMCs with ages of < 2 Gyr (e.g., Cabrera-Ziri et al. 2016b), they can still provide useful constraints on the origin of MPs, as most theories put forward so far do not invoke any special physics present only in the early Universe. Most theories simply invoke the gravitational potential of the young GC as being deep enough to hold

onto expelled stellar ejecta. Hence, even if YMCs are not the equivalent of proto-GCs, they can still be used to directly test predictions of the proposed scenarios.

5.6.1. Constraints on age spreads within young massive clusters. One of the key predictions of the AGB scenario is that clusters that are massive enough should be able to retain the ejecta of AGB stars and form subsequent stellar generations. Larsen et al. (2011) studied the resolved CMDs of seven massive (10^5 – $10^6 M_{\odot}$) young clusters in nearby galaxies, and while there were features in the CMDs that were not well described by a standard isochrone, age spreads (of the order of tens of millions of years) were also inconsistent with the observations. In one case, NGC 1313-379, an age spread could not be reliably ruled out.

Following on the work of Peacock et al. (2013), Bastian et al. (2013a) searched for evidence of ongoing star formation within a sample of ~ 140 YMCs with ages between 10 Myr and 1 Gyr and masses in the range of 10^4 – $10^8 M_{\odot}$. They searched for emission lines (i.e., H β and O[III]) from the unresolved clusters and O stars in the CMDs of the resolved clusters. No clusters were found with evidence of ongoing star formation. Cabrera-Ziri et al. (2014, 2016a) took this analysis a step further by estimating the star-formation histories of two massive ($> 10^7 M_{\odot}$) clusters in galactic merger remnants, NGC 34 (S1, ~ 100 Myr) and NGC 7252 (W3, ~ 500 Myr), using high signal-to-noise integrated optical spectra. In both cases, the clusters were best fit by an SSP (i.e., no evidence of a secondary starburst was found).

At an age of ~ 2 Gyr, NGC 1978 is the youngest cluster that shows evidence for MPs (Martocchia et al. 2018a). Owing to its youth, it can be used to place tight constraints on age differences between the subpopulations. Martocchia et al. (2018b) were able to identify two populations on the SGB of the cluster with UV photometry and then compared the positions of the stars in each population in an optical CMD. In optical colors, the position of the stars along the SGB (essentially the vertical placement of the stars) is sensitive mainly to age (and not chemical anomalies). The authors found an age difference of 1 ± 20 Myr between the populations, i.e., that they were coeval.

Taken together the constraints on age spreads in YMCs suggest that they are less than 10 or 20 Myr. This does not directly constrain the FRMS, VMS, or EDA scenarios but does place severe restrictions on scenarios that adopt AGBs as the polluters, as the first AGB stars do not evolve until 30 Myr after the 1G forms.

5.6.2. Constraints on gas and dust reservoirs within young massive clusters. In order for a massive cluster to form 2G stars, it must be able to retain a significant amount of gas within it for an extended period. Longmore (2015) used the predictions of the D’Ercole et al. (2008) model for multiple star-forming events in the context of the AGB scenario to show that the clusters should show extreme extinction in their inner regions, effectively being invisible in the inner ~ 3 pc. He notes that no such massive ring clusters have been observed and that many massive ($> 10^6 M_{\odot}$) clusters have been found with little or no extinction in the age range where the D’Ercole et al. (2008) models predicts that 2G stars should be forming. Cabrera-Ziri et al. (2015) used deep ALMA observations of three massive ($> 10^6 M_{\odot}$) clusters in the Antennae merging galaxies with ages between 50 and 200 Myr to search for any gas within them. Depending on the adopted conversion factor between the observed CO luminosity and total gas/dust mass, the authors could place upper limits of < 1 – 10% of the stellar mass being present in gas within the clusters.

Finally, Bastian et al. (2014) and Hollyhead et al. (2015) have studied a sample of young clusters (< 10 – 20 Myr) with masses between $\sim 10^4$ and $\sim 10^7 M_{\odot}$ to see how long clusters remain embedded in their natal gas cloud. In contradiction to the predictions of the FRMS scenario of Krause et al. (2013), who suggested that massive clusters should remain embedded for ~ 20 Myr,

observations showed that independent of mass (in the range studied) clusters were gas free within the first 2–4 Myr of their lives, probably before the first SNe (for metallicities from 1/5 solar to solar). Whitmore & Zhang (2002) and Reines et al. (2008) studied the nearby starburst galaxies, NGC 4038/39 and NGC 4449, respectively, comparing radio continuum measurements with optical HST colors and magnitudes. Both works conclude that YMCs are largely gas free by an age of 7 Myr and often considerably sooner.

We can conclude from these works that clusters are very efficient at removing (or consuming) any gas within them, from very young (a few million years) to very old ages (>1 Gyr). This applies to very massive clusters, even if simple escape velocity arguments would suggest that they should be able to retain any gas within them. For young clusters the Lyman–Werner flux within the cluster is expected to be very high (e.g., Conroy & Spergel 2011), which will not allow the gas to cool sufficiently to collapse to the cluster center, and the presence of X-ray binaries and other energetic sources (e.g., white dwarfs) and/or ongoing SNe appear to keep the cluster gas free throughout its lifetime. Hence, models that invoke the potential well of clusters to hold onto enriched gas are not supported by observations.

5.6.3. Globular clusters in formation. A major advance in the field may come with the launch of the *James Webb Space Telescope* (JWST), as, in specific circumstances, it will allow us to peer into galaxies at the epoch of GC formation (i.e., $z > 2$). Some initial steps in this direction have already been taken by observing highly lensed galaxies at $z > 3$ and their YMC populations. Vanzella et al. (2017) studied a sample of compact GC-like objects in five highly lensed galaxies including rest-frame UV/optical photometry and spectroscopy from HST and the Multi Unit Spectroscopic Explorer (MUSE) on the Very Large Telescope (VLT). Two of their objects, ID11 ($z = 3.1169$) and GC1 ($z = 6.145$) are particularly interesting because of their young ages (<10 Myr), small effective radii ($\lesssim 50$ pc), and stellar masses ($1\text{--}20 \times 10^6 M_\odot$), which are expected for YMCs. The estimated properties are also similar to those of YMCs forming in nearby galaxies, supporting the idea that YMCs are indeed the equivalent of young GCs.

It is difficult to draw conclusions from such a small sample, but future work on lensed samples (as well as JWST samples) offers a chance to study the population statistics of YMCs. If clusters are 10–30 times more massive when they form than they are currently, JWST would be expected to observe many clusters in excess of $0.5\text{--}1 \times 10^7 M_\odot$ (e.g., Renzini 2017). Alternatively, in models for the evolution of GCs based on the observed properties of YMCs and the conditions expected to be experienced by the clusters throughout their lives (i.e., models not tuned to achieve severe mass loss), only a handful of massive ($>0.5\text{--}1 \times 10^7 M_\odot$) clusters would be expected in each host galaxy (e.g., Kruijssen 2015).

5.7. Summary Points of the Comparison Between the Predictions and Observations

1. The observed positive correlations between f_{enriched} , extent of the abundance spreads, and cluster mass are directly at odds with scenarios that invoke large amounts of cluster mass loss in order to go from a cluster dominated by primordial stars to a cluster dominated by enriched stars.
2. This argues that the observed fractions are imprinted at birth, which essentially rules out all standard nucleosynthetic sources.
3. Quantitative comparison between the observed ranges of Na, O, and He spreads with the predicted yields of suggested polluter stars shows that none (or any combination thereof) can match the observations. While each cluster is unique in the details of its chemistry

- (requiring stochasticity in their formation), most clusters have He spreads that are much too small for the observed Na and O spreads.
4. Li is a problem for all scenarios, as it should be highly depleted in all material that is enriched in Na and He (and depleted in O), whereas observations do not show depletion to the predicted amounts (even including dilution).
 5. YMCs, with properties similar to those expected for young GCs, are still forming today. Studies have found evidence for neither multiple star-forming epochs within the clusters nor large gas/dust reservoirs needed to form further generations of stars. This is in tension with most proposed scenarios for the origin of MPs.
 6. We graphically summarize the comparison between models and predictions in **Figure 6**.

6. PECULIAR CLUSTERS: IRON SPREADS, CNO, AND S-PROCESS VARIATIONS

Although large variations in light element abundances are almost universal among old and massive clusters, the abundances of heavier α (Si, Ca, Ti), Fe-peak (Fe, Ni), and n-capture (Sr, Ba, La, Eu) elements within GCs vary little from star to star. Here we discuss MPs within the more complicated clusters that show abundance variations in heavy elements.

6.1. Clusters with Multimodal Metallicity Distributions: ω Centauri, M54, and Terzan 5

Understanding the formation and evolution of ω Cen, the most massive cluster in the Galaxy, represents a challenge for all the MP scenarios. The presence of a wide metallicity range ($-2.2 \leq [\text{Fe}/\text{H}] \leq -0.6$ dex; e.g., Johnson & Pilachowski 2010) in its stars demands that it was massive enough to retain SN ejecta at very high velocity (or to accrete gas from its surroundings for long periods), allowing for multiple bursts of star formation, with each generation becoming progressively enriched in Fe (e.g., Bedin et al. 2004). This possibly indicates that ω Cen constitutes the remnant of a tidally disrupted dwarf galaxy (e.g., Bekki & Freeman 2003). Although the observational scenario appears far more complex than for normal GCs, ω Cen also displays the key chemical signatures of MPs. Each metallicity subpopulation in ω Cen shows its own Na-O anticorrelation (with the possible exception of the most metal-poor stars), with the more metal-rich, He-rich stars ($Y \geq 0.35$; Joo & Lee 2013) showing an Na-O correlation (Marino et al. 2011a). The extension of the Na-O anticorrelation is also more extended toward higher metallicity, and the fraction of stars with high and intermediate Na also increases with metallicity (Marino et al. 2011a). This is difficult to explain within the AGB scenario framework, as the cooling flow from massive, metal-rich AGB stars would need to be delayed and further enriched by core-collapse SNe to account for more extended Na-O anticorrelation toward higher metallicity (D'Antona et al. 2011).

An increase in the CNO sum and in the s-process elements with $[\text{Fe}/\text{H}]$ is also observed (Johnson & Pilachowski 2010, Marino et al. 2011a). Low-mass AGB stars ($M < 3 M_{\odot}$) are observationally confirmed sites for s-process production, but they evolve on timescales longer (on the order of a gigayear) than the lifetimes of higher-mass AGB stars invoked to be responsible for the Na-O anticorrelation (~ 100 – 200 Myr, in the AGB scenario; see Section 4.1). Also, while AGB stars with masses $\lesssim 3 M_{\odot}$ can produce Na, enhance the C+N+O content, and produce s-process elements, they cannot deplete O or produce He. The most recent age estimates report a maximum relative age spread of only ~ 500 Myr among ω Cen populations (Tailo et al. 2016). Therefore low-mass AGBs that evolve on longer timescales cannot be responsible for the C+N+O and s-process pattern observed in ω Cen.

M54 is the nearest extragalactic GC we can observe and the second most massive GC in the halo. Even though the M54 metallicity distribution has a significantly smaller dispersion than ω Cen (e.g., Carretta et al. 2010a), both clusters have been proposed to represent a snapshot of nuclear star clusters in different stages of evolution. In the case of M54, the associated dwarf galaxy, i.e., the Sagittarius, is still visible, whereas the parent system once hosting ω Cen has been disrupted. Both metallicity groups in M54 display their own Na-O anticorrelation, with the metal-poor group showing a less extended Na-O anticorrelation with respect to the metal-rich stars, as observed for ω Cen (Carretta et al. 2010a).

Terzan 5 is a massive ($\sim 10^6 M_{\odot}$) stellar system located in the Bulge of the Galaxy. The two distinct red clumps in its CMD (Ferraro et al. 2009) have been linked to stellar populations with different metallicities (although see Lee et al. 2015 for an alternative explanation). Indeed, a large and multimodal metallicity distribution ($-0.8 \leq [\text{Fe}/\text{H}] \leq +0.3$ dex) has been reported (Massari et al. 2014); however, there is no consensus on the presence of light element spreads in Ter 5 (e.g., Origlia et al. 2011, Schiavon et al. 2017a). The α -element abundance pattern of the metallicity subpopulations mirrors what is observed for field stars in the Bulge, with α enhancement up to about solar metallicities and a decreasing $[\alpha/\text{Fe}]$ toward the solar ratio at supersolar $[\text{Fe}/\text{H}]$ (Origlia et al. 2011). The presence of two distinct MSTOs suggests that the dominant subsolar metallicity components developed ~ 12 Gyr ago, whereas the supersolar groups formed only ~ 4.5 Gyr ago after a prolonged period of quiescence (e.g., Ferraro et al. 2016). This finding has led to the suggestion that Ter 5 may constitute the remnant core of a dwarf galaxy or perhaps even a surviving fragment of the formation of the original bulge.

6.2. Clusters with Small Unimodal Iron Spreads and S-Process Bimodality

GCs characterized by a dispersion in their s-capture elements (e.g., M22, NGC 1851, M2, NGC 362, M19, NGC 5286) have received growing attention during recent years. The observed s-process bimodal distribution is associated with a split SGB in optical colors (e.g., Piotto et al. 2012) and, when C, N, and O abundances for unevolved stars are available, to variations in the net C+N+O content (e.g., Yong et al. 2015). Each s-process group displays its own Na-O anticorrelation, with the average Na abundance positively correlated with s-process enrichment (Marino et al. 2011b, Yong et al. 2014). Finally, s-rich stars are possibly slightly enhanced in Fe (e.g., Lim et al. 2017).

Because the presence of $[\text{Fe}/\text{H}]$ constrains the potential well in which a stellar system formed, a dispersion in $[\text{Fe}/\text{H}]$ implies that the system was able to retain SN ejecta to host multiple star-formation events.¹⁵ Indeed, it has been speculated that they represent the nuclear remnants of a tidally disrupted dwarf galaxy (e.g., Marino et al. 2015). This leads to the idea that GCs with small Fe variations would have contributed with a significant fraction of stars to the construction of the Galactic Halo, along with their host galaxies.

However, the presence of such small intrinsic Fe variations in a number of GCs is still debated, as they can be artificially introduced by the method used to derive atmospheric parameters of stars (Mucciarelli et al. 2015, Lardo et al. 2016; but see also Lee 2016). For example, very little star-to-star Fe variation is measured when metallicity is measured from FeII lines and the surface gravities are from photometry. Conversely, when gravities are derived by imposing the ionization

¹⁵The average $[\text{Fe}/\text{H}]$ dispersions for MW GCs are significantly smaller than the spectroscopic $[\text{Fe}/\text{H}]$ spreads of ~ 0.3 dex or more for dwarf galaxies, as no GC less luminous than $M_V = -10$ shows a substantial (≥ 0.1 dex) $[\text{Fe}/\text{H}]$ dispersion (Willman & Strader 2012).

equilibrium between the FeI and FeII, the [Fe/H] distribution is broad. Yet, the stellar gravities required to match [FeI/H] and [FeII/H] would lead to stellar masses for giants that are not physical (e.g., Mucciarelli et al. 2015). Interestingly, different FeI and FeII metallicity distributions are only observed in clusters that also show s-process and light element variations. Although the cause of the observed discrepancy between Fe abundances as inferred from FeI and FeII has not yet been determined, this finding suggests caution when measuring abundances using the classical spectroscopic approach on clusters with s-process variations.

Finally, the discrepancy between Fe abundances measured from FeI and FeII lines, which is observed for RGB stars with different s processes in a few clusters, is observed also in GCs with no intrinsic variations in heavy elements in the AGB phase, where FeI lines provide systematically lower abundances than RGBs (e.g., Lapenna et al. 2015, Wang et al. 2017). Currently, there is not an explanation for this effect.

6.3. The Blue Tilt in Cluster Populations

Observations of GC populations, especially around massive early-type galaxies (ETGs) that contain thousands of such clusters, have shown that the metal-poor population of clusters (i.e., the blue GCs) displays an average trend of becoming redder (more metal rich) as a function of increasing brightness (e.g., Harris 2009). The origin of this “blue tilt” is still uncertain, but a popular explanation for the phenomenon is that more massive clusters are able to retain not just the stellar ejecta (i.e., see Section 4) but also the SNe ejecta from 1G stars, and subsequently they form a more metal-rich 2G. The average metallicity of the cluster would then increase with each successive generation (see Strader & Smith 2008, Bailin & Harris 2009). One problem with such scenarios is that it is unclear how a cluster could retain the ejecta from SNe.

An alternative explanation that also accounts for why the the blue tilt is not observed in all GC populations has to do with how the metal-poor GC population is assembled, namely through the accretion of relatively low-mass metal-poor dwarf galaxies and their GC populations. As lower-mass dwarf galaxies have lower ISM pressures than their higher-mass counterparts, they are expected to form fewer high-mass clusters and have lower truncation masses to their mass functions (e.g., Kruijssen 2015). Massive GCs preferentially come from higher-mass dwarf galaxies, which in turn are more likely to be metal rich. This results in an upper envelope in the mass–metallicity plane for GC populations, skewing the mean metallicity to higher values for high cluster masses (Usher et al. 2018) Such a scenario can be tested with the next generation of galaxy-formation simulations that include GC formation and evolution (e.g., Pfeffer et al. 2018).

7. YOUNG MASSIVE CLUSTERS AND THEIR RELATIONSHIP TO GLOBULAR CLUSTERS

Although GCs were historically treated as objects that exclusively formed in the early Universe, it is now clear that objects with properties that are very similar to those expected of YMCs are still forming today. Some of these YMCs have masses and densities well in excess of present-day GCs, and their ages range from forming today to \sim 6–8 Gyr. While such clusters do exist in the Galaxy (with masses up to \sim 10⁵ M_⊙), they are difficult to study owing to the often extreme (differential) extinction and crowding in the Galactic plane. However, we are fortunate that our nearest extragalactic companions, the LMC and SMC, host large populations of such clusters. They are near enough that we can resolve them into their individual stars, especially with HST, and in some cases can obtain high-resolution spectra of individual stars.

A major finding of the past decade is that many of these clusters are not well represented by a single stellar isochrone but instead show features such as dual MSs and extended main sequence

eMSTO: extended main sequence turnoff

turnoffs (eMSTOs) among other unexpected features. The hope has been that these features are related to the MPs observed in the ancient GCs and that they could then be used to pinpoint the physical mechanisms responsible for MPs.

7.1. Extended Main Sequence Turnoffs in Young and Intermediate-Age Clusters

The high-precision photometry achievable with the Advanced Camera for Surveys (ACS) on HST allowed the construction of CMDs of massive young and intermediate-age clusters in the LMC and SMC in unparalleled detail. As is often the case, this increase in detail led to unexpected features that could not be explained within a traditional framework. In this case, it was the discovery of eMSTOs in the intermediate-age clusters (1–2 Gyr) in the LMC and SMC that could not be explained by photometric uncertainties or stellar binarity. This was first reported by Bertelli et al. (2003) and Mackey & Broby Nielsen (2007) and shown to be a general characteristic in subsequent works (e.g., Mackey et al. 2008, Milone et al. 2009, Piatti et al. 2014).

The initial explanation for the eMSTOs was that the clusters were formed in an extended star-forming event, lasting 200–700 Myr (e.g., Milone et al. 2009, Goudfrooij et al. 2014). Owing to this possibility, many works have attempted to link the observations of the eMSTO clusters with those of the ancient GCs hosting MPs (e.g., Goudfrooij et al. 2014). However, subsequent work has shown that the eMSTO phenomenon is unlikely to be caused by an actual age spread within the clusters (see Section 5.6). Subsequent studies have found that YMCs with ages between 20–300 Myr also show eMSTOs and that the inferred age spread was directly proportional to the age of the cluster (Niederhofer et al. 2015). Additionally, studies focused on other regions of the CMDs that should also be affected by age spreads have not been found to be in agreement with the age-spread interpretation (e.g., Li et al. 2016). Finally, at \sim 2 Gyr, NGC 1978 does not show an eMSTO (Martocchia et al. 2018b) despite its relatively high mass.

This points instead toward a stellar evolutionary effect. One such effect is stellar rotation, first proposed by Bastian & de Mink (2009) and subsequently studied in more detail by Brandt & Huang (2015) using the Geneva stellar evolutionary models that include rotation. Such models do well in predicting the relationship between the inferred age spread and the age of the cluster, as well as the lack of eMSTOs in clusters with ages above \sim 2 Gyr owing to magnetic braking of the stars.

Finally, recent high-resolution studies of A and F (1–2.5 M_{\odot}) stars have found evidence for light element abundance (Na, O, Mg) spreads in rapidly rotating stars in OCs (Pancino 2018). The origin of these variations (and their link to GCs) is still unknown, but rotational mixing and diffusion are possible causes.

It is striking that the eMSTO phenomenon disappears at (nearly) the same age that MPs on the RGB begin to be seen (Martocchia et al. 2018a,b). How or whether these two phenomena are related is a rich avenue for future work.

7.2. Split Main Sequences

Another surprising feature that has been found in resolved CMDs of YMCs in the LMC and SMC was that many of them, when viewed in the blue/UV filters, displayed bimodal (i.e., split) MSs (Milone et al. 2015a). At first glance, this appears to be similar to the split MSs in ancient GCs that are due to light element abundance spreads (e.g., He, C, N, and O spreads). However, Milone et al. (2015a) investigated possible causes of the splits, creating stellar models that included the abundance spreads, iron spreads, C+N+O spreads, and also age spreads. They conclude that none of the models were able to explain the split MS observed in clusters like NGC 1856 (\sim 300 Myr, \sim 10⁵ M_{\odot}).

D’Antona et al. (2015) used the SYCLIST stellar models (Georgy et al. 2014) that include rotation (including inclination effects) to model NGC 1856 and showed that rotation could explain the observed MS split if the stellar rotation distribution was bimodal with a minor peak at $\omega < 0.3$ and a dominant peak at $\omega \sim 0.9$. It is interesting to note that in all the YMCs in the LMC studied to date with split MSs, the red MS (corresponding to the rapid rotators) is generally the dominant population (between 42% and 75%; e.g., Milone et al. 2016, 2017a). These stars would be rotating much faster than those typically found in the field or in lower-mass OCs (McSwain & Gies 2005).

Such an extreme rotational distribution should lead to observationally detectable signatures, as a large population of rapid rotators should have a high rate of Be stars, i.e., stars near the critical rotation limit with partially ionized decretion discs. Bastian et al. (2017) looked for such a population of Be stars and indeed found a much higher fraction in the \sim 100-Myr cluster NGC 1850 and the \sim 300-Myr cluster NGC 1856. In both clusters, the authors found Be fractions between 30% and 60% near the MSTO, which is much higher than that found in the field or in lower-mass clusters. These observations confirmed the high fraction of rapid rotators in YMCs, lending support to the idea that the split MS is caused by a bimodal rotational distribution.

However, further observations to measure the actual rotational distribution in YMCs are required to directly test this scenario. Preliminary results appear to confirm the bimodal rotational distribution with a large fraction of rapidly rotating stars (Dupree et al. 2017). If true, the conclusion would be that stars forming in dense/massive clusters would retain a signature of their origin, namely in their rapid rotation rates, although why stars born in clusters would preferentially be born with high rotation rates is currently unknown.

7.3. Chemical Anomalies in Young Massive Clusters?

Although YMCs have provided strong tests for the theories of the formation of MPs, it is not yet clear whether they host such abundance anomalies. As discussed in Section 2.5, initial spectroscopic studies of a limited number of stars in massive young and intermediate-age clusters in the LMC did not find evidence of MPs (Mucciarelli et al. 2008, 2014a). This has been confirmed through photometric studies based on large samples (Martocchia et al. 2017, 2018a).

The young and intermediate-age LMC and SMC clusters are quite massive, relative to their OC counterparts in the Galaxy; however, as discussed in Section 5.6, YMCs with much higher masses (by factors of 10 to 1,000) are known to exist. By contrast, the distances to these extragalactic objects generally make it impossible to obtain high-precision photometry or spectroscopy for individual stars. Hence, some studies have attempted to search for the spectroscopic fingerprint of MPs in integrated light. Cabrera-Ziri et al. (2016b) and Lardo et al. (2017a) have exploited the fact that YMCs are dominated by the light of RSGs at young ages (in the near-IR) and that RSGs all have similar temperatures, meaning that their integrated light can be studied as a single RSG. If MPs are present in these massive YMCs, we would expect that their Al and Na abundances would be higher than that of field RSGs at the same Fe abundance. These authors studied four clusters with masses in the range of $5\text{--}20 \times 10^5 M_{\odot}$ and searched for evidence of Al enhancement, although none was found in any of the clusters despite their high masses. This RGB-focused technique is sensitive to chemical anomalies in stars above $\sim 15 M_{\odot}$ (e.g., Davies et al. 2008), although integrated light spectroscopy can in principle be used to search for MPs at any age, with proper modeling of its stellar populations (e.g., Hernandez et al. 2017).

One potential caveat to note about the previous studies is that they are not comparing like with like, at least in terms of stellar mass. All studies of young and intermediate-age clusters have focused on the evolved portions of the CMD (e.g., the RGB), which at 200 Myr or 2 Gyr corresponds to a stellar mass of $\sim 3.6 M_{\odot}$ and $\sim 1.5 M_{\odot}$, respectively (at $[Fe/H] = -0.7$). At ages of 6 Gyr and

10 Gyr the stellar mass on the RGB is $\sim 1.0 M_{\odot}$ and $\sim 0.9 M_{\odot}$, respectively. Though the MS for the LMC and SMC young and intermediate massive clusters is out of range for spectroscopy with existing instruments, there is potential to use HST to obtain N-sensitive photometry to make comparisons of the same mass range in young and ancient clusters (i.e., $< 0.8 M_{\odot}$). Additionally, future instruments like JWST or the E-ELT (European Extremely Large Telescope) may provide important insights at lower stellar masses.

8. MULTIPLE POPULATIONS ON GALAXY SCALES

Dwarf galaxies have stellar masses ranging from the GC mass scale up to a few $\times 10^9 M_{\odot}$. In many cases, their stellar populations are not too dissimilar from that of certain GCs (like ω Cen and M54), with modest metallicity spreads and a dominant old stellar population (see Section 6.1). It is normally assumed that MPs are not present in the field stars in dwarfs, due to (*a*) the assumption that MPs are restricted to GCs and (*b*) the low fraction ($\sim 3\%$) of 2P stars in the field of the MW halo (e.g., Martell et al. 2011), which is thought to come, at least partially, from accreted satellite dwarf galaxies. We can infer a lack of a large population of stars with large ΔY values within local dwarf galaxies on the basis of the morphology of the HB. The HBs of dwarf galaxies lack, to the “extreme,” stars seen in GCs with large Y spreads (e.g., NGC 2808). For example, detailed modeling of the HB of the Carina Dwarf galaxy did not lead to evidence of Y spreads within the populations (although age and Fe spreads were identified; Savino et al. 2015). Additionally, Norris et al. (2017) searched for MPs in the Carina dwarf galaxies in 63 RGB stars (looking for an Na-O spread) and only found stars with typical abundance patterns, i.e., 1P stars.

Stepping further afield, Strader et al. (2013) studied a very massive ($\sim 2 \times 10^8 M_{\odot}$) and dense ($R_h = 24$ pc) ultracompact dwarf galaxy around the Virgo elliptical galaxy, M60 (M60-UCD1). The authors find evidence for the object to be enriched in N ([N/Fe] = +0.61) and Na ([Na/Fe] = +0.42), hence it likely hosts MPs, with a large population of highly enriched 2P stars.

Although studies of MPs and chemical anomalies have largely focused on massive and dense star clusters, there is growing evidence that they may be present outside clusters, making up a significant fraction of the stars in certain parts of galaxies. Schiavon et al. (2017b) discovered a large population of N-rich stars, which display correlations between [N/Fe] and [Al/Fe], as well as being anticorrelated with [C/Fe]; i.e., they display the same chemical anomalies as stars in GCs. The authors focused on the low-metallicity regime and found that for [Fe/H] < -1 , the chemically anomalous stars make up $\sim 7\%$ of the stars of the Bulge/inner halo. Extrapolating their results to the full Bulge/inner halo, they estimate that the mass of enriched stars is a few $\times 10^8 M_{\odot}$, which is a factor of ~ 8 more than the mass of the entire Galactic GC system. This fact, and the lack of correspondence between the enriched star and GC population metallicity distributions, suggests that the discovered enriched stars in the Bulge/inner halo did not originate from dissolved GCs.

If true, this would suggest that MPs may not be a product of only GCs but may instead be a general feature of certain stellar populations. Though currently still inconclusive, there is tantalizing evidence that MPs may be present in other dense and old stellar populations. For example, the mean [N/Fe] and [Na/Fe] abundances of ETGs increase with increasing velocity dispersion (e.g., Schiavon 2007, Conroy et al. 2014), which could imply that the fraction of enriched stars is an increasing function of velocity dispersion. Recently, van Dokkum et al. (2017) have used high signal-to-noise spatially resolved spectra of massive ETGs and find that the mean [Na/Fe] abundance increases toward the galaxy centers while [O/Fe] decreases, again suggesting that MPs may be present in the centers of such systems. Although high velocity dispersion within ETGs is also positively correlated with high [Mg/Fe] (e.g., Walcher et al. 2015), van Dokkum et al. (2017)

found that relative to the outskirts of the galaxies, [Mg/Fe] was depressed in the central regions. Hence, the centers of ETGs appear to show many of the trends seen in MPs.

Another potential link between MPs and the massive ETGs is through the UV-upturn (e.g., O'Connell 1999). The origin of the UV-upturn is still under debate, but the presence of a large number of extreme HB stars is one possibility. As seen in Galactic GCs, like NGC 2808, the presence of a large He spread among cluster stars is correlated with an extreme population of HB stars (metallicity also affects the fraction of stars that pass through an extreme HB period). Hence, if ETGs do host MPs, it would imply that the UV-upturn is caused by large He spreads, which would be correlated with large Na and N spreads (Chantreau et al. 2018).

Further work is needed to explicitly test if MPs are present within ETGs and, if so, in what fractions. However, if MPs are found to make up a significant fraction of ETG stars, it would have a dramatic effect on our understanding of MPs and their origin. It may imply, for example, that we need to explore noncluster-focused scenarios for the origin of MPs.

FUTURE ISSUES

Throughout this review, we have attempted to highlight topics that are particularly uncertain, and new theoretical and observational studies are likely to lead to important advances. Here, we briefly summarize some of the directions that we feel are likely to be the most fruitful in the next few years.

- While observations of evolved stars in YMCs have not revealed the presence of MPs, it is not clear if MPs are absent or restricted in the stellar mass range in which they can appear. The unexpected transition at ~ 2 Gyr, below which MPs are not found in evolved stars and above which they are, suggests that MPs may be present in many YMCs but only in low-mass stars (i.e., lower-mass MS stars).
- In order to identify the cluster parameter(s) that control whether MPs are present (age, mass, density, metallicity, etc.), the parameter space of clusters should be further sampled. Looking at low-density GCs in the outer Galactic halo, or those that have been accreted, could be particularly fruitful. Also, extending the age range of clusters under study may place stricter limits on the appearance of MPs.
- Further work quantifying how the properties of MPs within clusters depend on the cluster properties would be very beneficial. Is cluster mass or density the controlling factor for the fraction of enriched stars or the degree of abundance spreads within clusters?
- To date, only a handful of GC stars have been fully characterized in terms of their abundances (He, C, N, O, Na, Al, Mg, etc.). Systematic studies of the precise way all these elements are related and of the variety between clusters may help pinpoint the origin of MPs. Dissecting the (pseudo)color–color diagrams of the *HST UV Legacy Survey* may offer an efficient means to search many of these correlations. What causes the spread in the 1P stars in the pseudocolor diagrams in some clusters and not in others? Detailed modeling of the color spreads is needed to characterize the abundance variations in a large sample of GCs (as well as is confirmation through spectroscopic follow-up). If spectroscopy confirms that the color spread among 1P stars is due to He variations (associated with small-or-no C-N-Na-O variations), alternative physical mechanisms for the origin of MPs—other than stellar nucleosynthesis—need to be investigated.

- As discussed in Section 8 there is tentative evidence that MPs may not be restricted to GCs but may be present in other environments as well (dwarf galaxies, bulge/inner halos of galaxies and ETGs). Studies confirming or refuting this may result in a major breakthrough in the field.
- Recent theoretical studies have largely focused on developing existing scenarios, exploring ways in which the models can be changed to provide a better match to observations. We argue that the present observations do not support the traditional theories of self-enrichment through the formation of multiple generations of stars. Hence, new theories for the origin of MPs (e.g., nonstandard stellar evolution, VMSs, etc.) should be encouraged and developed to test against the wealth of observational data now in hand.
- One property of stars that affects stellar evolution, which is dependent on environment, is stellar rotation. Stars in dense/massive young clusters rotate significantly faster than those in the field or lower-mass OCs. Additionally, the age boundary for whether MPs are present (2–2.5 Gyr) in evolved stars is also the boundary (\sim 1.5–1.6 M_{\odot}) at which MSTO and RGB stars would be magnetically braked (i.e., at this age clusters no longer show eMSTOs). Could MPs be caused by a nonstandard stellar evolutionary effect linked to rotation?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to Soeren Larsen, Ivan Cabrera-Ziri, Corinne Charbonnel, Maurizio Salaris, Mark Gieles, Emanuele Dalessandro, Alessio Mucciarelli, Chris Usher, William Chantereau, Elena Pancino, Santi Cassisi, Francesca D’Antona, Henry Lamers, Eugenio Carretta, and Angela Bragaglia for helpful discussions and detailed comments on earlier versions of the manuscript. Additionally, we thank the editor, Sandy Faber, and the copyeditor, Roselyn Lowe-Webb, for suggestions that significantly improved the manuscript. N.B. acknowledges financial support from the Royal Society (University Research Fellowship) and the European Research Council (ERC-CoG-646928, Multi-Pop). C.L. thanks the Swiss National Science Foundation for supporting this research through the Ambizione grant number PZ00P2_168065.

LITERATURE CITED

- Anderson J, Piotto G, King IR, Bedin LR, Guhathakurta P. 2009. *Ap. J. Lett.* 697:L58–62
Bailin J, Harris WE. 2009. *Ap. J.* 695:1082–93
Bastian N, Cabrera-Ziri I, Davies B, Larsen SS. 2013a. *MNRAS* 436:2852–63
Bastian N, Cabrera-Ziri I, Niederhofer F, et al. 2017. *MNRAS* 465:4795–99
Bastian N, Cabrera-Ziri I, Salaris M. 2015. *MNRAS* 449:3333–46
Bastian N, de Mink SE. 2009. *MNRAS* 398:L11–15
Bastian N, Hollyhead K, Cabrera-Ziri I. 2014. *MNRAS* 445:378–84
Bastian N, Lamers HJGLM, de Mink SE, et al. 2013b. *MNRAS* 436:2398–411
Bastian N, Lardo C. 2015. *MNRAS* 453:357–64
Bedin LR, Piotto G, Anderson J, et al. 2004. *Ap. J. Lett.* 605:L125–28

- Behr BB. 2003. *Ap. J. Suppl.* 149:67–99
- Bekki K. 2017. *MNRAS* 467:1857–73
- Bekki K, Campbell SW, Lattanzio JC, Norris JE. 2007. *MNRAS* 377:335–51
- Bekki K, Freeman KC. 2003. *MNRAS* 346:L11–15
- Bell RA, Dickens RJ. 1980. *Ap. J.* 242:657–72
- Bellazzini M, Bragaglia A, Carretta E, et al. 2012. *Astron. Astrophys.* 538:A18
- Bellini A, Bedin LR, Piotto G, et al. 2010. *Astron. J.* 140:631–41
- Bellini A, Piotto G, Milone AP, et al. 2013. *Ap. J.* 765:32
- Bellini A, Vesperini E, Piotto G, et al. 2015. *Ap. J. Lett.* 810:L13
- Bertelli G, Nasi E, Girardi L, et al. 2003. *Astron. J.* 125:770–84
- Bragaglia A, Carretta E, D’Orazi V, et al. 2017. *Astron. Astrophys.* 607:44–61
- Bragaglia A, Carretta E, Gratton R, et al. 2010a. *Astron. Astrophys.* 519:A60
- Bragaglia A, Carretta E, Gratton RG, et al. 2010b. *Ap. J. Lett.* 720:L41–45
- Bragaglia A, Gratton RG, Carretta E, et al. 2012. *Astron. Astrophys.* 548:A122
- Brandt TD, Huang CX. 2015. *Ap. J.* 807:25
- Briley MM, Cohen JG, Stetson PB. 2004. *Astron. J.* 127:1579–87
- Cabrera-Ziri I, Bastian N, Davies B, et al. 2014. *MNRAS* 441:2754–59
- Cabrera-Ziri I, Bastian N, Hilker M, et al. 2016a. *MNRAS* 457:809–21
- Cabrera-Ziri I, Bastian N, Longmore SN, et al. 2015. *MNRAS* 448:2224–31
- Cabrera-Ziri I, Lardo C, Davies B, et al. 2016b. *MNRAS* 460:1869–75
- Çalışkan Ş, Christlieb N, Grebel EK. 2012. *Astron. Astrophys.* 537:A83
- Caloi V, D’Antona F. 2007. *Astron. Astrophys.* 463:949–55
- Cameron AGW, Fowler WA. 1971. *Ap. J.* 164:111
- Cannon RD, Croke BFW, Bell RA, Hesser JE, Stathakis RA. 1998. *MNRAS* 298:601–24
- Carretta E. 2014. *Ap. J. Lett.* 795:L28
- Carretta E. 2015. *Ap. J.* 810:148
- Carretta E, Bragaglia A, Gratton R, Lucatello S. 2009a. *Astron. Astrophys.* 505:139–55
- Carretta E, Bragaglia A, Gratton RG, et al. 2006. *Astron. Astrophys.* 450:523–33
- Carretta E, Bragaglia A, Gratton RG, et al. 2010a. *Ap. J. Lett.* 714:L7–11
- Carretta E, Bragaglia A, Gratton RG, et al. 2009b. *Astron. Astrophys.* 505:117–38
- Carretta E, Bragaglia A, Gratton RG, et al. 2010b. *Astron. Astrophys.* 516:A55
- Carretta E, Gratton RG, Bragaglia A, D’Orazi V, Lucatello S. 2013. *Astron. Astrophys.* 550:A34
- Cassisi S, Salaris M, Anderson J, et al. 2009. *Ap. J.* 702:1530–35
- Cassisi S, Salaris M, Irwin AW. 2003. *Ap. J.* 588:862–70
- Cassisi S, Salaris M, Pietrinferni A, Hyder D. 2017. *MNRAS* 464:2341–48
- Cassisi S, Salaris M, Pietrinferni A, Vink JS, Monelli M. 2014. *Astron. Astrophys.* 571:A81
- Cassisi S, Salaris M, Pietrinferni A, et al. 2008. *Ap. J. Lett.* 672:L115
- Chantereau W, Charbonnel C, Meynet G. 2016. *Astron. Astrophys.* 592:A111
- Chantereau W, Usher C, Bastian N. 2018. *MNRAS* 478(2):2368–87
- Charbonnel C. 2016. In *Stellar Clusters: Benchmarks of Stellar Physics and Galactic Evolution - EES2015*, EAS Publ. Ser., Vol. 80–81, ed. E Moraux, Y Lebreton, C Charbonnel, pp. 177–226. Paris: EAS Publ. Ser.
- Charbonnel C, Chantereau W, Krause M, Primas F, Wang Y. 2014. *Astron. Astrophys.* 569:L6
- Cohen JG. 2004. *Astron. J.* 127:1545–54
- Colucci JE, Bernstein RA, Cameron SA, McWilliam A. 2012. *Ap. J.* 746:29
- Colucci JE, Bernstein RA, Cohen JG. 2014. *Ap. J.* 797:116
- Conroy C. 2012. *Ap. J.* 758:21
- Conroy C, Graves GJ, van Dokkum PG. 2014. *Ap. J.* 780:33
- Conroy C, Spergel DN. 2011. *Ap. J.* 726:36
- Cordero MJ, Hénault-Brunet V, Pilachowski CA, et al. 2017. *MNRAS* 465:3515–35
- Cottrell PL, Da Costa GS. 1981. *Ap. J. Lett.* 245:L79–82
- Dalessandro E, Lapenna E, Mucciarelli A, et al. 2016. *Ap. J.* 829:77
- Dalessandro E, Massari D, Bellazzini M, et al. 2014. *Ap. J. Lett.* 791:L4
- Dalessandro E, Salaris M, Ferraro FR, Mucciarelli A, Cassisi S. 2013. *MNRAS* 430:459–71

- Dalessandro E, Salaris M, Ferraro FR, et al. 2011. *MNRAS* 410:694–704
- D’Antona F, Bellazzini M, Caloi V, et al. 2005. *Ap. J.* 631:868–78
- D’Antona F, Caloi V, Montalbán J, Ventura P, Gratton R. 2002. *Astron. Astrophys.* 395:69–75
- D’Antona F, Caloi V, Ventura P. 2010. *MNRAS* 405:2295–301
- D’Antona F, D’Ercole A, Marino AF, et al. 2011. *Ap. J.* 736:5
- D’Antona F, Di Criscienzo M, Decressin T, et al. 2015. *MNRAS* 453:2637–43
- D’Antona F, Vesperini E, D’Ercole A, et al. 2016. *MNRAS* 458:2122–39
- D’Ercole A, D’Antona F, Ventura P, Vesperini E, McMillan SLW. 2010. *MNRAS* 407:854–69
- D’Ercole A, D’Antona F, Vesperini E. 2011. *MNRAS* 415:1304–9
- D’Ercole A, D’Antona F, Vesperini E. 2016. *MNRAS* 461:4088–98
- D’Ercole A, Vesperini E, D’Antona F, McMillan SLW, Recchi S. 2008. *MNRAS* 391:825–43
- D’Orazi V, Gratton R, Lucatello S, et al. 2010. *Ap. J. Lett.* 719:L213–17
- D’Orazi V, Gratton RG, Angelou GC, et al. 2015. *MNRAS* 449:4038–47
- Davies B, Figer DF, Law CJ, et al. 2008. *Ap. J.* 676:1016–28
- de Mink SE, Pols OR, Langer N, Izzard RG. 2009. *Astron. Astrophys.* 507:L1–4
- Decressin T, Charbonnel C, Meynet G. 2007a. *Astron. Astrophys.* 475:859–73
- Decressin T, Charbonnel C, Siess L, et al. 2009. *Astron. Astrophys.* 505:727–33
- Decressin T, Meynet G, Charbonnel C, Prantzos N, Ekström S. 2007b. *Astron. Astrophys.* 464:1029–44
- Denisenkov PA, Denisenkova SN. 1989. *Astron. Tsirkulyar* 1538:11
- Denisenkov PA, Denisenkova SN. 1990. *Sov. Astron. Lett.* 16:275
- Denissenkov PA, Hartwick FDA. 2014. *MNRAS* 437:L21–25
- Denissenkov PA, Herwig F. 2003. *Ap. J. Lett.* 590:L99–102
- Di Criscienzo M, Tailo M, Milone AP, et al. 2015. *MNRAS* 446:1469–77
- Dickens RJ, Croke BFW, Cannon RD, Bell RA. 1991. *Nature* 351:212–14
- Doherty CL, Gil-Pons P, Lau HHB, et al. 2014. *MNRAS* 441:582–98
- Dupree AK, Avrett EH. 2013. *Ap. J. Lett.* 773:L28
- Dupree AK, Dotter A, Johnson CI, et al. 2017. *Ap. J. Lett.* 846:L1
- Elmegreen BG. 2017. *Ap. J.* 836:80
- Fenner Y, Campbell S, Karakas AI, Lattanzio JC, Gibson BK. 2004. *MNRAS* 353:789–95
- Ferraro FR, Dalessandro E, Mucciarelli A, et al. 2009. *Nature* 462:483–86
- Ferraro FR, Massari D, Dalessandro E, et al. 2016. *Ap. J.* 828:75
- Georgy C, Granada A, Ekström S, et al. 2014. *Astron. Astrophys.* 566:A21
- Gieles M, Charbonnel C, Krause MGH, et al. 2018. *MNRAS* 478(2):2461–79
- Goudfrooij P, Girardi L, Kozhurina-Platais V, et al. 2014. *Ap. J.* 797:35
- Gratton R, Sneden C, Carretta E. 2004. *Annu. Rev. Astron. Astrophys.* 42:385–440
- Gratton RG, Carretta E, Bragaglia A. 2012a. *Astron. Astrophys. Rev.* 20:50
- Gratton RG, Carretta E, Bragaglia A, Lucatello S, D’Orazi V. 2010. *Astron. Astrophys.* 517:A81
- Gratton RG, Lucatello S, Carretta E, et al. 2012b. *Astron. Astrophys.* 539:A19
- Gratton RG, Lucatello S, Sollima A, et al. 2014. *Astron. Astrophys.* 563:A13
- Grundahl F, Catelan M, Landsman WB, Stetson PB, Andersen MI. 1999. *Ap. J.* 524:242–61
- Harris WE. 1996. *Astron. J.* 112:1487
- Harris WE. 2009. *Ap. J.* 699:254–80
- Harris WE, Harris GLH. 2002. *Astron. J.* 123:3108–23
- Harris WE, Harris GLH, Alessi M. 2013. *Ap. J.* 772:82
- Hartwick FDA, McClure RD. 1972. *Ap. J. Lett.* 176:L57
- Hénault-Brunet V, Gieles M, Agertz O, Read JI. 2015. *MNRAS* 450:1164–98
- Hernandez S, Larsen SS, Trager S, Groot P, Kaper L. 2017. *Astron. Astrophys.* 306:119
- Hollyhead K, Bastian N, Adamo A, et al. 2015. *MNRAS* 449:1106–17
- Hollyhead K, Kacharov N, Lardo C, et al. 2017. *MNRAS* 465:L39–43
- Hopkins PF. 2014. *Ap. J.* 797:59
- Ivans II, Sneden C, Kraft RP, et al. 1999. *Astron. J.* 118:1273–300
- Johnson CI, Pilachowski CA. 2010. *Ap. J.* 722:1373–410
- Johnson JA, Ivans II, Stetson PB. 2006. *Ap. J.* 640:801–22

- Joo SJ, Lee YW. 2013. *Ap. J.* 762:36
- Karakas A, Lattanzio JC. 2007. *Publ. Astron. Soc. Aust.* 24:103–17
- Karakas AI, Lattanzio JC. 2014. *Publ. Astron. Soc. Aust.* 31:e030
- Khalaj P, Baumgardt H. 2016. *MNRAS* 457:479–86
- King IR, Bedin LR, Cassisi S, et al. 2012. *Astron. J.* 144:5
- Koch A, Côté P, McWilliam A. 2009. *Astron. Astrophys.* 506:729–43
- Krause M, Charbonnel C, Decressin T, Meynet G, Prantzos N. 2013. *Astron. Astrophys.* 552:A121
- Krause M, Charbonnel C, Decressin T, et al. 2012. *Astron. Astrophys.* 546:L5
- Krause MGH, Charbonnel C, Bastian N, Diehl R. 2016. *Astron. Astrophys.* 587:A53
- Kruijssen JMD. 2014. *Class. Quantum Gravity* 31:244006
- Kruijssen JMD. 2015. *MNRAS* 454:1658–86
- Kruijssen JMD, Mieske S. 2009. *Astron. Astrophys.* 500:785–99
- Kučinskas A, Dobrovolskas V, Bonifacio P. 2014. *Astron. Astrophys.* 568:L4
- Langer GE, Hoffman R, Sneden C. 1993. *Publ. Astron. Soc. Pac.* 105:301–7
- Lapenna E, Lardo C, Mucciarelli A, et al. 2016. *Ap. J. Lett.* 826:L1
- Lapenna E, Mucciarelli A, Ferraro FR, et al. 2015. *Ap. J.* 813:97
- Lardo C, Bellazzini M, Pancino E, et al. 2011. *Astron. Astrophys.* 525:A114
- Lardo C, Cabrera-Ziri I, Davies B, Bastian N. 2017a. *MNRAS* 468:2482–88
- Lardo C, Mucciarelli A, Bastian N. 2016. *MNRAS* 457:51–63
- Lardo C, Salaris M, Bastian N, et al. 2018. *Astron. Astrophys.* Accepted. arXiv:1805.09599
- Lardo C, Salaris M, Savino A, et al. 2017b. *MNRAS* 466(3): 3507–12
- Larsen SS. 2004. *Astron. Astrophys.* 416:537–53
- Larsen SS, Baumgardt H, Bastian N, et al. 2015. *Ap. J.* 804:71
- Larsen SS, Brodie JP, Forbes DA, Strader J. 2014a. *Astron. Astrophys.* 565:A98
- Larsen SS, Brodie JP, Grundahl F, Strader J. 2014b. *Ap. J.* 797:15
- Larsen SS, de Mink SE, Eldridge JJ, et al. 2011. *Astron. Astrophys.* 532:A147
- Larsen SS, Strader J, Brodie JP. 2012. *Astron. Astrophys.* 544:L14
- Lee JW. 2016. *Ap. J. Suppl.* 226:16
- Lee JW. 2017. *Ap. J.* 844:77
- Lee YW, Joo SJ, Chung C. 2015. *MNRAS* 453:3906–11
- Letarte B, Hill V, Jablonka P, et al. 2006. *Astron. Astrophys.* 453:547–54
- Li C, de Grijs R, Bastian N, et al. 2016. *MNRAS* 461:3212–21
- Lim D, Hong S, Lee YW. 2017. *Ap. J.* 844:14
- Lim D, Lee YW, Pasquato M, Han SI, Roh DG. 2016. *Ap. J.* 832:99
- Lind K, Charbonnel C, Decressin T, et al. 2011. *Astron. Astrophys.* 527:A148
- Lind K, Primas F, Charbonnel C, Grundahl F, Asplund M. 2009. *Astron. Astrophys.* 503:545–57
- Lochhaas C, Thompson TA. 2017. *MNRAS* 470:977–91
- Longmore SN. 2015. *MNRAS* 448:L62–66
- Longmore SN, Kruijssen JMD, Bastian N, et al. 2014. In *Protostars and Planets VI*, ed. H Beuther, FS Klessen, CP Dullemond, T Henning, pp. 291–314. Tucson: Univ. Ariz. Press
- Loup C, Forveille T, Omont A, Paul JF. 1993. *Astron. Astrophys. Suppl.* 99:291–377
- Lovisi L, Mucciarelli A, Lanzoni B, et al. 2012. *Ap. J.* 754:91
- Lucatello S, Sollima A, Gratton R, et al. 2015. *Astron. Astrophys.* 584:A52
- Mackey AD, Broby Nielsen P. 2007. *MNRAS* 379:151–58
- Mackey AD, Broby Nielsen P, Ferguson AMN, Richardson JC. 2008. *Ap. J. Lett.* 681:L17
- MacLean BT, De Silva GM, Lattanzio J. 2015. *MNRAS* 446:3556–61
- Maeder A, Meynet G. 2006. *Astron. Astrophys.* 448:L37–41
- Marcolini A, Gibson BK, Karakas AI, Sánchez-Blázquez P. 2009. *MNRAS* 395:719–35
- Marino AF, Milone AP, Karakas AI, et al. 2015. *MNRAS* 450:815–45
- Marino AF, Milone AP, Piotto G, et al. 2011a. *Ap. J.* 731:64
- Marino AF, Milone AP, Sneden C, et al. 2012. *Astron. Astrophys.* 541:A15
- Marino AF, Sneden C, Kraft RP, et al. 2011b. *Astron. Astrophys.* 532:A8
- Marino AF, Villanova S, Piotto G, et al. 2008. *Astron. Astrophys.* 490:625–40

- Martell SL, Smolinski JP, Beers TC, Grebel EK. 2011. *Astron. Astrophys.* 534:A136
- Martocchia S, Bastian N, Usher C, et al. 2017. *MNRAS* 468:3150–58
- Martocchia S, Cabrera-Ziri I, Lardo C, et al. 2018a. *MNRAS* 473:2688–700
- Martocchia S, Niederhofer F, Dalessandro E, et al. 2018b. *MNRAS* 477:4696–705
- Massari D, Lapenna E, Bragaglia A, et al. 2016. *MNRAS* 458:4162–71
- Massari D, Mucciarelli A, Ferraro FR, et al. 2014. *Ap. J.* 795:22
- Mastrobuono-Battisti A, Perets HB. 2013. *Ap. J.* 779:85
- Mayya YD, Rosa-González D, Santiago-Cortés M, et al. 2013. *MNRAS* 436:2763–73
- McDonald I, Zijlstra AA. 2015. *MNRAS* 446:2226–42
- McSwain MV, Gies DR. 2005. *Ap. J. Suppl.* 161:118–46
- Mészáros S, Martell SL, Shetrone M, et al. 2015. *Astron. J.* 149:153
- Miholics M, Webb JJ, Sills A. 2015. *MNRAS* 454:2166–72
- Milone AP. 2015. *MNRAS* 446:1672–84
- Milone AP, Bedin LR, Piotto G, Anderson J. 2009. *Astron. Astrophys.* 497:755–71
- Milone AP, Bedin LR, Piotto G, et al. 2015a. *MNRAS* 450:3750–64
- Milone AP, Marino AF, D'Antona F, et al. 2016. *MNRAS* 458:4368–82
- Milone AP, Marino AF, D'Antona F, et al. 2017a. *MNRAS* 465:4363–74
- Milone AP, Marino AF, Piotto G, et al. 2015b. *Ap. J.* 808:51
- Milone AP, Piotto G, Renzini A, et al. 2017b. *MNRAS* 464:3636–56
- Moehler S, Dreizler S, LeBlanc F, et al. 2014. *Astron. Astrophys.* 565:A100
- Monelli M, Milone AP, Stetson PB, et al. 2013. *MNRAS* 431:2126–49
- Mucciarelli A, Bellazzini M, Ibata R, et al. 2012. *MNRAS* 426:2889–900
- Mucciarelli A, Carretta E, Origlia L, Ferraro FR. 2008. *Astron. J.* 136:375–88
- Mucciarelli A, Dalessandro E, Ferraro FR, Origlia L, Lanzoni B. 2014a. *Ap. J. Lett.* 793:L6
- Mucciarelli A, Lapenna E, Massari D, et al. 2015. *Ap. J.* 809:128
- Mucciarelli A, Lovisi L, Lanzoni B, Ferraro FR. 2014b. *Ap. J.* 786:14
- Mucciarelli A, Origlia L, Ferraro FR, Pancino E. 2009. *Ap. J. Lett.* 695:L134–39
- Mucciarelli A, Salaris M, Lovisi L, et al. 2011. *MNRAS* 412:81–94
- Nardiello D, Piotto G, Milone AP, et al. 2015. *MNRAS* 451:312–22
- Niederhofer F, Bastian N, Kozhurina-Platais V, et al. 2017a. *MNRAS* 464:94–103
- Niederhofer F, Bastian N, Kozhurina-Platais V, et al. 2017b. *MNRAS* 465:4159–65
- Niederhofer F, Georgy C, Bastian N, Ekström S. 2015. *MNRAS* 453:2070–74
- Norris J. 1987. *Ap. J. Lett.* 313:L65–68
- Norris J, Cottrell PL, Freeman KC, Da Costa GS. 1981. *Ap. J.* 244:205–20
- Norris JE, Yong D, Venn KA, et al. 2017. *Ap. J. Suppl.* 230:28
- O'Connell RW. 1999. *Annu. Rev. Astron. Astrophys.* 37:603–48
- Origlia L, Rich RM, Ferraro FR, et al. 2011. *Ap. J. Lett.* 726:L20
- Osborn W. 1971. *Observatory* 91:223–24
- Pancino E. 2018. *Astron. Astrophys.* Accepted. arXiv:1802.06654
- Pancino E, Romano D, Tang B, et al. 2017. *Astron. Astrophys.* 601:A112
- Pasquini L, Bonifacio P, Molaro P, et al. 2005. *Astron. Astrophys.* 441:549–53
- Pasquini L, Mauas P, Käufel HU, Cacciari C. 2011. *Astron. Astrophys.* 531:A35
- Peacock MB, Zepf SE, Finzell T. 2013. *Ap. J.* 769:126
- Peacock MB, Zepf SE, Kundu A, Chael J. 2017. *MNRAS* 464:713–20
- Pfeffer J, Kruijssen JMD, Crain RA, Bastian N. 2018. *MNRAS* 475:4309–46
- Piatti AE, Keller SC, Mackey AD, Da Costa GS. 2014. *MNRAS* 444:1425–41
- Pietrinferni A, Cassisi S, Salaris M, Percival S, Ferguson JW. 2009. *Ap. J.* 697:275–82
- Piotto G, Bedin LR, Anderson J, et al. 2007. *Ap. J. Lett.* 661:L53–56
- Piotto G, Milone AP, Anderson J, et al. 2012. *Ap. J.* 760:39
- Piotto G, Milone AP, Bedin LR, et al. 2015. *Astron. J.* 149:91
- Piotto G, Villanova S, Bedin LR, et al. 2005. *Ap. J.* 621:777–84
- Portegies Zwart SF, McMillan SLW, Gieles M. 2010. *Annu. Rev. Astron. Astrophys.* 48:431–93
- Prantzos N, Charbonnel C. 2006. *Astron. Astrophys.* 458:135–49

- Prantzos N, Charbonnel C, Iliadis C. 2007. *Astron. Astrophys.* 470:179–90
- Prantzos N, Charbonnel C, Iliadis C. 2017. *Astron. Astrophys.* 608:28–40
- Recio-Blanco A, Aparicio A, Piotto G, de Angeli F, Djorgovski SG. 2006. *Astron. Astrophys.* 452:875–84
- Reines AE, Johnson KE, Goss WM. 2008. *Astron. J.* 135:2222–39
- Renzini A. 2008. *MNRAS* 391:354–62
- Renzini A. 2013. *Mem. Soc. Astron. Ital.* 84:162
- Renzini A. 2017. *MNRAS* 469:L63–67
- Renzini A, D’Antona F, Cassisi S, et al. 2015. *MNRAS* 454:4197–207
- Richer HB, Heyl J, Anderson J, et al. 2013. *Ap. J. Lett.* 771:L15
- Roediger JC, Courteau S, Graves G, Schiavon RP. 2014. *Ap. J. Suppl.* 210:10
- Rood RT. 1973. *Ap. J.* 184:815–38
- Sakari CM, Venn KA, Mackey D, et al. 2015. *MNRAS* 448:1314–34
- Salaris M, Cassisi S. 2014. *Astron. Astrophys.* 566:A109
- Salaris M, Weiss A, Ferguson JW, Fusilier DJ. 2006. *Ap. J.* 645:1131–37
- Salinas R, Strader J. 2015. *Ap. J.* 809:169
- Sánchez-Blázquez P, Marcolini A, Gibson BK, et al. 2012. *MNRAS* 419:1376–89
- Sarajedini A, Bedin LR, Chaboyer B, et al. 2007. *Astron. J.* 133:1658–72
- Savino A, Salaris M, Tolstoy E. 2015. *Astron. Astrophys.* 583:A126
- Sbordone L, Bonifacio P, Buonanno R, et al. 2007. *Astron. Astrophys.* 465:815–24
- Sbordone L, Salaris M, Weiss A, Cassisi S. 2011. *Astron. Astrophys.* 534:A9
- Schaerer D, Charbonnel C. 2011. *MNRAS* 413:2297–304
- Schaeuble M, Preston G, Sneden C, et al. 2015. *Astron. J.* 149:204
- Schiavon RP. 2007. *Ap. J. Suppl.* 171:146–205
- Schiavon RP, Caldwell N, Conroy C, et al. 2013. *Ap. J. Lett.* 776:L7
- Schiavon RP, Johnson JA, Frinchaboy PM, et al. 2017a. *MNRAS* 466:1010–18
- Schiavon RP, Zamora O, Carrera R, et al. 2017b. *MNRAS* 465:501–24
- Shen ZX, Bonifacio P, Pasquini L, Zaggia S. 2010. *Astron. Astrophys.* 524:L2
- Sills A, Glebbeek E. 2010. *MNRAS* 407:277–84
- Simioni M, Milone AP, Bedin LR, et al. 2016. *MNRAS* 463:449–58
- Slemer A, Marigo P, Piatti D, et al. 2017. *MNRAS* 465:4817–37
- Smith GH. 2015. *Publ. Astron. Soc. Pac.* 127:1204–17
- Sneden C. 2000. In *Proc. 35th Liège Int. Astrophys. Colloq., Liège, Belgium, Jul. 5–8, 1999*, ed. A Noels, P Magain, D Caro, E Jehin, G Parmentier, AA Thoul, pp. 159–75. Liege: Univ. Liege
- Sneden C, Kraft RP, Prosser CF, Langer GE. 1992. *Astron. J.* 104:2121–40
- Sohn ST, O’Connell RW, Kundu A, et al. 2006. *Astron. J.* 131:866–88
- Soto M, Bellini A, Anderson J, et al. 2017. *Astron. J.* 153:19–29
- Strader J, Seth AC, Forbes DA, et al. 2013. *Ap. J. Lett.* 775:L6
- Strader J, Smith GH. 2008. *Astron. J.* 136:1828–36
- Szécsi D, Mackey J, Langer N. 2018. *Astron. Astrophys.* 612:A55
- Tailo M, Di Criscienzo M, D’Antona F, Caloi V, Ventura P. 2016. *MNRAS* 457:4525–35
- Takeda Y, Kaneko H, Matsumoto N, et al. 2009. *Publ. Astron. Soc. Jpn.* 61:563–76
- Tautvaišienė G, Wallerstein G, Geisler D, Gonzalez G, Charbonnel C. 2004. *Astron. J.* 127:373–79
- Tenorio-Tagle G, Silich S, Rodríguez-González A, Muñoz-Tuñón C. 2005. *Ap. J.* 628:L13–16
- Trenti M, Padoan P, Jimenez R. 2015. *Ap. J. Lett.* 808:L35
- Usher C, Pfeffer J, Bastian N, et al. 2018. *MNRAS*. Submitted
- van Dokkum P, Conroy C, Villaume A, Brodie J, Romanowsky A. 2017. *Ap. J.* 841:68–91
- Vanderbeke J, De Propis R, De Rijcke S, et al. 2015. *MNRAS* 451:275–81
- Vanzella E, Calura F, Meneghetti M, et al. 2017. *MNRAS* 467:4304–21
- Ventura P, D’Antona F. 2005a. *Astron. Astrophys.* 431:279–88
- Ventura P, D’Antona F. 2005b. *Astron. Astrophys.* 439:1075–91
- Ventura P, D’Antona F. 2006. *Astron. Astrophys.* 457:995–1001
- Ventura P, D’Antona F. 2009. *Astron. Astrophys.* 499:835–46
- Ventura P, D’Antona F, Di Criscienzo M, et al. 2012. *Ap. J. Lett.* 761:L30

- Ventura P, D'Antona F, Mazzitelli I. 2002. *Astron. Astrophys.* 393:215–23
- Ventura P, Di Criscienzo M, Carini R, D'Antona F. 2013. *MNRAS* 431:3642–53
- Ventura P, García-Hernández DA, Dell'Agli F, et al. 2016. *Ap. J. Lett.* 831:L17
- Vesperini E, McMillan SLW, D'Antona F, D'Ercole A. 2010. *Ap. J. Lett.* 718:L112–16
- Vesperini E, McMillan SLW, D'Antona F, D'Ercole A. 2013. *MNRAS* 429:1913–21
- Villanova S, Geisler D, Carraro G, Moni Bidin C, Muñoz C. 2013. *Ap. J.* 778:186
- Villanova S, Piotto G, Gratton RG. 2009. *Astron. Astrophys.* 499:755–63
- Vink JS, Muijres LE, Anthonisse B, et al. 2011. *Astron. Astrophys.* 531:A132
- Walcher CJ, Coelho PRT, Gallazzi A, et al. 2015. *Astron. Astrophys.* 582:A46
- Walker AR, Kunder AM, Andreuzzi G, et al. 2011. *MNRAS* 415:643–54
- Wang Y, Primas F, Charbonnel C, et al. 2017. *Astron. Astrophys.* 607:135
- Webb JJ, Leigh NWC. 2015. *MNRAS* 453:3278–87
- Whitmore BC, Zhang Q. 2002. *Astron. J.* 124:1418–34
- Wijnen TPG, Pols OR, Pelupessy FI, Portegies Zwart S. 2016. *Astron. Astrophys.* 594:A30
- Willman B, Strader J. 2012. *Astron. J.* 144:76
- Wünsch R, Palouš J, Tenorio-Tagle G, Ehlerová S. 2017. *Ap. J.* 835:60
- Yong D, Grundahl F, Johnson JA, Asplund M. 2008. *Ap. J.* 684:1159–69
- Yong D, Grundahl F, Lambert DL, Nissen PE, Shetrone MD. 2003. *Astron. Astrophys.* 402:985–1001
- Yong D, Grundahl F, Nissen PE, Jensen HR, Lambert DL. 2005. *Astron. Astrophys.* 438:875–88
- Yong D, Grundahl F, Norris JE. 2015. *MNRAS* 446:3319–29
- Yong D, Roederer IU, Grundahl F, et al. 2014. *MNRAS* 441:3396–416

Contents

Cosmology Paradigm Changes <i>Jaan Einasto</i>	1
High-Mass Star and Massive Cluster Formation in the Milky Way <i>Frédérique Motte, Sylvain Bontemps, and Fabien Louvet</i>	41
Multiple Stellar Populations in Globular Clusters <i>Nate Bastian and Carmela Lardo</i>	83
Dynamical Evolution of the Early Solar System <i>David Nesvorný</i>	137
Origins of Hot Jupiters <i>Rebekah I. Dawson and John Asher Johnson</i>	175
Chemodynamical History of the Galactic Bulge <i>Beatriz Barbuy, Cristina Chiappini, and Ortwin Gerhard</i>	223
Multiconjugate Adaptive Optics for Astronomy <i>François Rigaut and Benoit Neichel</i>	277
Extreme Adaptive Optics <i>Olivier Guyon</i>	315
The Pluto System After <i>New Horizons</i> <i>S. Alan Stern, William M. Grundy, William B. McKinnon, Harold A. Weaver, and Leslie A. Young</i>	357
Weak Lensing for Precision Cosmology <i>Rachel Mandelbaum</i>	393
The Connection Between Galaxies and Their Dark Matter Halos <i>Risa H. Wechsler and Jeremy L. Tinker</i>	435
Atomic and Ionized Microstructures in the Diffuse Interstellar Medium <i>Snežana Stanimirović and Ellen G. Zweibel</i>	489
Debris Disks: Structure, Composition, and Variability <i>A. Meredith Hughes, Gaspard Duchêne, and Brenda C. Matthews</i>	541

Rubble Pile Asteroids <i>Kevin J. Walsh</i>	593
Obscured Active Galactic Nuclei <i>Ryan C. Hickox and David M. Alexander</i>	625
The Interstellar Dust Properties of Nearby Galaxies <i>Frédéric Galliano, Maud Galametz, and Anthony P. Jones</i>	673

Indexes

Cumulative Index of Contributing Authors, Volumes 45–56	715
Cumulative Index of Article Titles, Volumes 45–56	718

Errata

An online log of corrections to *Annual Review of Astronomy and Astrophysics* articles may be found at <http://www.annualreviews.org/errata/astro>