Linking Chemical Signatures of Globular Clusters to Chemical Evolution

Francesca D'Antona
Paolo Ventura
INAF – Astronomical Observatory of Rome, Italy

The majority of Globular Clusters (GC) show chemical inhomogeneities in the composition of their stars, apparently attributable to a second stellar generation in which the forming gas is enriched by hot CNO-cycled material processed in stars belonging to a first stellar generation. We review the reasons why the site of the nucleosynthesis can be identified with hot-bottom burning in the envelopes of massive asymptotic giant branch (AGB) stars and super-AGB stars. The analysis of spectroscopic data and photometric signatures, such as the horizontal branch morphology, shows that the percentage of 'anomalous' stars is 50 % or more in most GCs examined so far. If anomalies are the rule and not the exception, then they clearly are closely related to the dynamical way in which GCs form and survive. We show a possible solution obtained by a hydrodynamical model followed by the N-body evolution of the two stellar populations, and propose that most GCs survive thanks to the formation of the second stellar generation.

The most massive AGB stars as sites of nucleosynthesis of CNO, ^{23}Na, ^{27}Al and ^{7}Li

The presence of the CN dichotomy, and of the Na-O and Mg-Al anticorrelations among the stars of practically all GCs examined in the Milky Way, show very clearly that the matter of 'anomalous' stars must have been processed through the hot CNO cycle, i.e. it has not only experienced CN, but also ON cycling. This process occurs deep in stellar interiors during the H-burning stage or at the 'hot' bottom of the convective envelopes of massive asymptotic giant branch stars. The so-called metals (iron, but also calcium and the other heavy elements) do not show significant star-to-star variations in most GCs (Gratton et al., 2004), so that any process of production of the anomalous gas does not involve supernova ejecta. Associated 'anomalies' include carbon depletion and nitrogen enhancement in a fraction of the cluster stars, and it has been shown that the increase in nitrogen is not only due to CN cycling, but must also include N production at the expense of oxygen (see, e.g., Cohen et al., 2002; 2005), thus relating the Na-O with the C-N anticorrelations.

If neither supernovae that are due to the core collapse of massive stars (SNe II), nor supernovae that occur much later in the life of the clusters (SNe Ia) had a role to play in the process of self-enrichment, the most obvious source is the massive AGB stars, as already proposed in the 1980s, notably by Cottrell & Da Costa (1981). The hot CNO-processed matter of the AGB envelopes is injected into the cluster at low velocity by stellar winds and planetary nebula ejection, and the stellar remnants are quiet massive white dwarfs (WDs). The very high temperatures at the bottom of the convective envelopes (T_{bce}) of these stars are exemplified in Figure 1, and no peculiar extra mixing needs be invoked to obtain the necessary very hot CNO processing. Hot-bottom burning (HBB) was recognised as an important physical process in these stars in the 1970s, and ^{3}He burning with ^{4}He, followed by the non-instantaneous mixing of the resulting ^{7}Be (Cameron-Fowler hypothesis) could explain the presence of abundant lithium, due to beryllium decay, in very high luminosity M giants that are above the highest luminosity limit for carbon stars. More recently, the possibility that the anomalous gas comes out from fast-rotating massive stars has been explored by the Geneva group (e.g., Decressin et al., 2007), but this model has not yet been fully explored from the dynamical point of view.

In some recent work, the AGB enrichment scenario, although appealing for the dynamical reasons quoted above, has often been considered inadequate to explain the features of the second stellar generation (SG). This is attributable to the fact that the results of stellar modelling of massive AGB stars obtained by different groups differ greatly from each other.

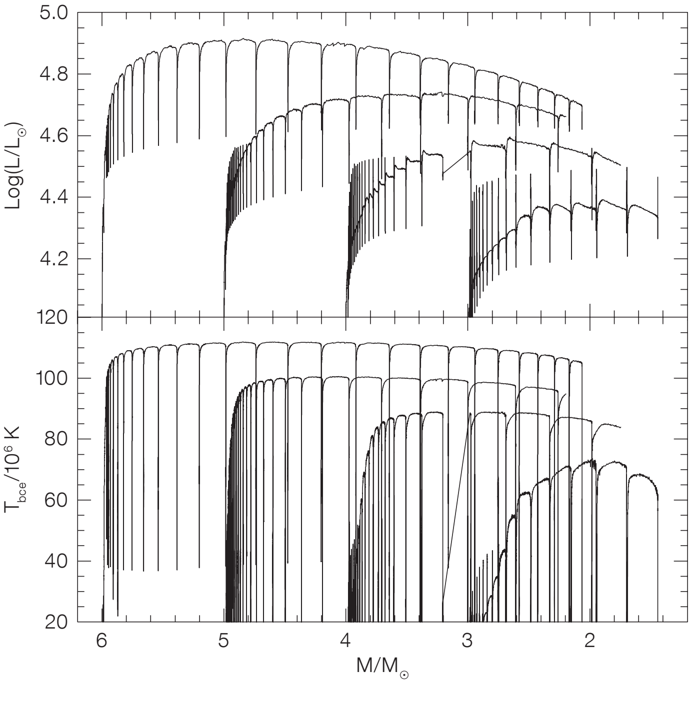

Figure 1. Luminosity (upper) and temperature at the bottom of the convective envelope T_{bce} (lower) in models of 6, 5, 4 and 3 $M_\odot$ with metallicity $Z = 0.001$. The abscissa is the current mass of the evolutionary track, which decreases due to mass loss (from the Ventura and D'Antona models). The very high T_{bce} values allow hot-bottom burning in all models down to 4 $M_\odot$, and only marginally in those of 3 $M_\odot$. More massive stars have faster nucleosynthesis and evolution, a smaller number of thermal pulses and third dredge-up episodes, so their evolution is able to account for the chemistry of the second generation stars.

The uncertainty in the nuclear cross-sections has often been emphasised in the work of, e.g., Lattanzio, Karakas, Izzard and Ventura. In particular, the very important sodium yield is made up by a series of events: it first increases in the envelope due to the second dredge-up (e.g., Iben & Renzini, 1983); then increases due to HBB of the ^{22}Ne dredged up in this same event; afterwards, it increases in the third dredge-up episodes. If T_{bce} is high, sodium is destroyed by p-captures. Notably, however, the ^{23}Na(p,α)^{20}Ne and, especially, the ^{22}Ne(p,γ)^{23}Na cross-sections are very uncertain in the range of temperatures of interest for HBB, by up to a factor 10^3 for the latter (compare Hale et al., 2002 with the 'standard' NACRE cross-sections found in Angulo et al., 1999). By choosing a high rate for this cross-section, the ^{23}Na yield of massive AGBs, where HBB is actively depleting oxygen, is high and compatible with the observations. Also the ^{27}Al production, by proton capture on ^{25}Mg, ^{26}Mg and ^{26}Al, is crucially dependent on the relevant cross-sections, and selection of the highest possible rates in the NACRE compilation allows the observed Mg-Al anticorrelation to be reproduced today.

The highest uncertainty in modelling is recognised to be the treatment of convection: the discrepancy, e.g., between the yields obtained by Karakas & Lattanzio (2007) and ours is mainly due to this modelling (Ventura & D'Antona, 2005). We adopt a very efficient convection model, proposed by Canuto et al. (1996), the Full Spectrum of Turbulence model, or FST. This results in a higher T_{bce} and more efficient nuclear processing, higher luminosities, higher mass-loss rates and consequently a lower number of third dredge-up episodes. Thus the oxygen reduction is not nullified by the third dredge-up, and the sodium is not increased too much by the dredge-up of ^{22}Ne (the ultimate product of ^{14}N burning in the helium intershell) and its consequent burning by p-captures.

Nevertheless, the sodium and oxygen yields in the matter expelled from AGB stars are directly correlated: lower initial masses, with smaller T_{bce}, longer AGB lifetimes and a higher number of third dredge-up episodes, have higher oxygen and sodium abundances. This is shown schematically in Figure 2: the three diagonal lines represent the Na and O yields (and their possible uncertainty) for AGBs as a function of the AGB mass, decreasing from left to right. It is clear that the anticorrelation shown by observations (dots) cannot be explained by the occurrence of star formation in pure matter expelled from AGBs of decreasing mass. The AGB matter must be diluted, at some level, with pristine matter, providing values of Na and O intermediate between the AGB starting yield and the pristine value. This is exemplified by the Na-O area within the cone drawn in the figure. The Na-O anticorrelation thus requires two events: (1) the minimum mass contributing to the SG should not have sodium abundance too high with respect to the observed values; (2) the AGB matter must be diluted with pristine gas.

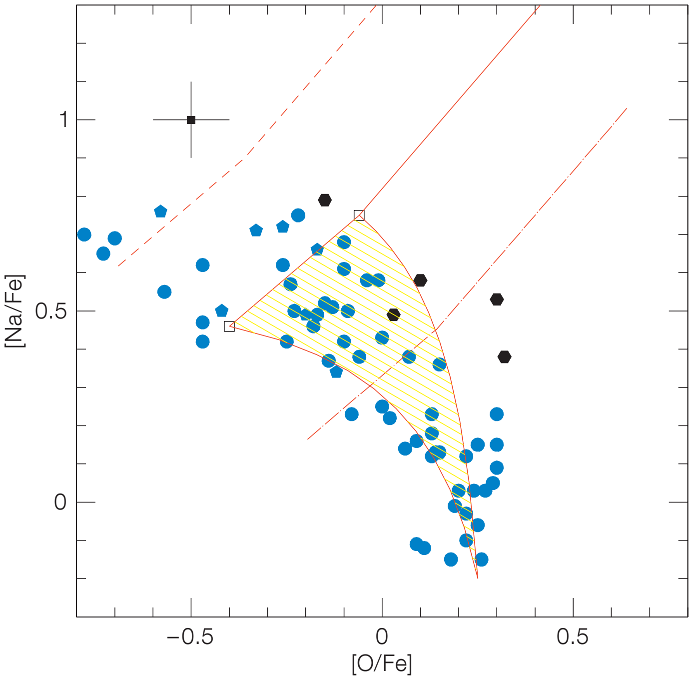

Figure 2. Schematic representation of the observed Na-O anticorrelation (dots). The diagonal lines represent schematic yields of AGB models, having decreasing mass from left to right. The middle line represents yields that can be consistent with observations; the lower line represents the yields obtained by adopting smaller cross-sections for the ^{22}Ne(p,γ)^{23}Na reaction and the highest line represents yields that can be obtained with a less efficient convection model. The dashed cone region includes abundances obtained by diluting the yields given by the two open squares, along the chosen yield line, with different amounts of pristine material having the composition of the cone vertex. It is evident that a satisfactory interpretation of the data requires such a dilution model, and cannot simply rely on the 'pure' abundances of the ejecta. In this scheme, the abundance points outside the cone (smaller O abundances, found only in red giants) require some extra-mixing process in order to be explained.

A related observational and modelling problem is that of the CNO total abundances. If they are constant among normal and anomalous stars, this implies that the AGB evolution must have suffered only a few episodes of third dredge-up. Or it might indeed imply that the matter is only CNO cycled, as can happen preferentially in the evolution of massive stars. The CNO data of Carretta (2005) seem to indicate a small, but unequivocal, overabundance of total CNO in the anomalous stars. This may favour the AGB scenario, indicating that a limited number of third dredge-up episodes actually play a (small) role. The most interesting indications in favour of the model came recently from the spectroscopic and photometric analysis of the cluster NGC 1851: this is the only cluster for which the total CNO variation may reach a factor ~ 4, as shown by Yong and collaborators in this meeting, and the giants also show an s-abundance spread, unlike all other GCs (Yong & Grundahl, 2008). Are we witnessing a GC in which the formation of the SG was slightly delayed with respect to the other clusters, so that it also involved the ejecta of AGBs suffering some more third dredge-up processing? The CNO total abundance variation in NGC 1851 is probably also related to the splitting of the subgiant branch and to its bimodal horizontal branch (HB) morphology, as presented in other reviews.

We finally touch on the problem of the lithium abundance of unevolved stars in GCs. In spite of the difficulties of observing this element in low luminosity turn-off stars, it has been known for many years that ^{7}Li in GCs varies from star to star, much more than observed in halo stars. In NGC 6752, Pasquini et al. (2005) have shown that lithium is anticorrelated with sodium. Na-normal, Li-normal stars do exist, as in the halo, and these should in fact be the first generation (FG) stars. Other stars then formed from mixing of the pristine matter with Na-rich, Li-poor material, thus providing the anticorrelation. But Pasquini et al. (2008) have recently shown that two stars in NGC 6397, both very similar and 'normal' in lithium, have oxygen abundances differing by ~ 0.6 dex. This may indicate that the O-poor matter is Li-rich too, and this is possible only if the SG forms from matter including AGB ejecta, and not massive star ejecta. The lithium yield of AGBs is extremely dependent on the mass-loss rates, but models providing yields close to the standard population II abundance are very reasonable.

Helium variations: super-AGBs as the site of production of extreme helium populations

All models of self-enrichment show that the ejected material must be enriched in helium with respect to the pristine one. The higher helium content has been recognised to have a strong effect on the HB morphology, possibly helping to explain some features (gaps, hot blue tails, second parameter) that have defied alternative explanation (D'Antona et al., 2002). A variety of problems along these lines have been examined in recent years: the extreme peculiarity of the HB in the massive cluster NGC 2808 (D'Antona & Caloi, 2004); the second parameter effect in M13 and M3 (Caloi & D'Antona, 2005); the peculiar features in the RR Lyrae variables; and the HB of the two metal-rich clusters NGC 6441 and NGC 6388 (Caloi & D'Antona, 2007; Busso et al., 2007). The presence of strongly enhanced helium in peculiar HB stars has been confirmed in NGC 2808 and NGC 6441 by spectroscopic observations (see, e.g., Moehler et al., 2004; 2007). In addition, an unexpected feature has recently appeared from photometric data, confirming the interpretation of helium differences amongst GC stars, *viz.* the splitting of the main sequence in NGC 2808. After first indications from a wider-than-expected colour distribution found by D'Antona et al. (2005) from archival Hubble Space Telescope (HST) data, recent HST observations by Piotto et al. (2007) leave no doubt that there are at least three different populations in this cluster. This finding came after the first discovery of a peculiar blue main sequence in ω Centauri by Bedin et al. (2004), also interpreted in terms of a very high helium content (Norris, 2004; Piotto et al., 2005). The above-mentioned cases can be considered as extreme, in the sense that no explanation had been attempted for them before the hypothesis of helium-enriched populations.

The helium yield of AGBs is in part due to the second dredge-up, in part to the effect of the third dredge-up episodes. As the number of these third dredge-ups must be small in order to preserve the quasi-constancy of CNO, the main effect must be due to the second dredge-up. While standard AGB models do not reach helium yields larger than $Y \sim 0.35$, and thus seem unable to explain the larger Y values of ω Centauri and NGC 2808, the super-AGB models by Siess (2007a) show yields $Y \sim 0.36$–0.38, and suggest these stars as candidates for the progeny of the extreme GC population.

Between the stars that evolve into core-collapse supernovae and the minimum mass for carbon ignition (below which stars evolve into CO white dwarfs), the super-AGB stars ignite carbon off-centre in semi-degenerate conditions, but are not able to ignite hydrostatic neon-burning in the resulting ONe core. Consequently, degeneracy increases in the core, and these stars may undergo thermal pulses, as first shown in models by Iben's group in the 1990s (e.g. Ritossa et al., 1999), and lose mass as 'normal' (but quite massive and luminous) AGB stars, but different in core composition (ONe versus CO) and core mass ($> 1.05\ M_{\odot}$). Although full models through the super-AGB thermal pulse phase are not yet available, we can foresee that the sum of the CNO abundances can remain close to the initial value as a result of the efficiency of third dredge-up being limited, because the helium luminosity during the thermal pulses is weak, as shown again by Siess (2007b). These are the premises that make super-AGBs good candidates for the formation of the extreme helium population harboured in the most massive GCs (Pumo et al., 2008).

The fate of super-AGBs depends on the competition between mass-loss rate and core growth: if mass-loss wins, they evolve into massive ONe white dwarfs; if the core grows until it reaches the Chandrasekhar mass, they evolve into electron capture supernovae (ecSNe), electrons being captured on the Ne nuclei. Thus a fraction of the super-AGBs may explode as supernovae, but these events are at least a factor ten less energetic than the SNe II ($\leq 10^{50}$ erg) and also a factor ten less frequent than SNe Ia. Consequently, the epoch during which super-AGBs evolve is probably the quietest period in the cluster lifetime, perturbed at most by ecSN explosions. This stage is not energetic enough to alter either the gas evolution or its chemistry, as the whole core remains locked by the remnant neutron stars. It has recently been proposed that practically all the neutron stars (NS) present today in GCs are born from ecSNe. Due to their lower energy output, it is also probable that the newborn NS receives a proportionally smaller natal kick, which allows it to remain bound to the cluster (Ivanova et al., 2008). If there are no energy sources (SN II or SN Ia) able to expel the gas ejected at low velocity by stars, this gas can collapse into the cluster core and form new stars with the chemistry of the ejecta.

What is the percentage of second generation stars?

We point out that helium variations produce appreciable differences in the location of the main sequence of clusters only if they are very large and *uniform* (D'Antona & Caloi, 2004; Salaris et al., 2006). Small helium spreads can be revealed from the HB morphology, which amplifies any small total mass decrease, by increasing the stellar T_{eff} location on the HB, but helium spread in the turn-off and main sequence stars remains hidden in the observational errors. Therefore, we should not regard the clusters with split main sequences (ω Centauri and NGC 2808) as typical examples of clusters with multiple stellar populations: they are examples of clusters also harbouring an extreme population identified by its blue main sequence, corresponding to $Y \sim 0.38–0.40$.

D'Antona & Caloi (2008) have examined the HB features of about 15 clusters and have shown in most cases that the higher helium abundances remain confined below $Y \sim 0.32$. Nevertheless, the percentage of SG stars – defined now as all stars with Y larger than the 'standard' Big Bang abundance $Y \sim 0.24$ – is generally larger than 50 %! D'Antona & Caloi (2008) also pose the question of whether GCs with only a blue HB (the classic second parameter effect) should be explained by assuming that these are clusters formed only from second generation stars. One of the interesting cases is NGC 6397, the small cluster with an HR diagram that has always been regarded as a perfect example of a simple stellar population, especially following the HST proper motion observations by King et al. (1998) and most recently by Richer et al. (2006). Nevertheless, only three scarcely evolved stars out of 14 are nitrogen normal (Carretta et al., 2005), leading us to suspect that the material from which most stars formed is CNO processed and thus belonging to the SG. This occurrence had already been noticed by Bonifacio et al. (2002), with reference to the paradox that nitrogen-rich stars had almost normal lithium content. The question remains whether NGC 6397 is an SG-only cluster.

We are finally confronted with the real question: how does a GC form? Is it possible to form a cluster with FG and SG stars in equal proportions? Is it possible to have a cluster made up only of SG stars? All GCs so far examined appear to contain an SG!

Is the second generation necessary for the survival of globular clusters?

A back-of-the-envelope computation is enough to realise that the matter forming the SG stars far exceeds the wind matter contained in massive AGBs, if the initial mass function (IMF) of the system is more or less standard and we assume that the FG low mass stars we see today represent the low mass end of the IMF. The initial population from which we need to collect AGB winds, massive enough to produce a populous SG, must have been at least a factor ten more massive than today's cluster mass. This requirement lends support to the idea that GCs are either the compact nuclei of dwarf galaxies (Bekki & Norris, 2006) or are formed within dwarf galaxies that are afterwards dispersed.

A different point of view is assumed in the recent work by D'Ercole et al. (2008): they start with a massive FG cluster, and follow the hydrodynamic formation of the SG. After the Type II supernova epoch, which cleared the cluster of its pristine gas, the low velocity winds of super-AGBs, and then of massive AGBs, collect in a cooling flow in the innermost regions of the cluster, where they form SG stars. The cluster emerging from the hydrodynamical simulations is one with an SG strongly concentrated at the inner core of a more extended FG population. The initial mass of the FG stars needs to be large enough to provide enough stellar mass return and to form a substantial number of SG stars. Consequently, the FG stars are initially the dominant stellar population with a total mass that must be about ten times larger than the total mass of the SG stars. The SG formation ends when SNe Ia begin to explode regularly in the cluster. As SN Ia explosions occur when CO WDs reach the Chandrasekhar mass by accretion in binary systems, this epoch certainly begins some time after the birth of massive CO WDs, so that both super-AGBs and the most massive AGBs can contribute to the cooling flow, consistent with the scenario outlined above. In 100 Myr (at most) a cluster with two dynamically separated components has been formed.

The dynamical evolution of the composite FG plus SG cluster is followed by means of N-body simulations, starting with a highly concentrated SG, and an FG extended to the tidal radius. The FG is also expanding, due to the SN II explosions and consequent mass loss. If the initial FG was already mass-segregated (as massive clusters seem to be observationally), the heating and the expansion due to the loss of SN ejecta are augmented by the preferential removal of the mass from the inner regions of the cluster. D'Ercole et al. (2008) find that early cluster evolution and mass loss can lead to a significant loss of FG stars. In Figure 3 we show that the number ratio of SG to FG stars (f_{MS}) may not only reach values consistent with observations (0.5–1.5), but that there may also be evolutionary routes leading to the loss of most of the FG population and leaving an SG-dominated cluster. Thus this model shows that the clusters that survive might preferentially be those in which a substantial SG has had time to form.

Back to the helium inhomogeneity

We have seen that, from the chemical point of view, the less extreme anomalies require mixing of the AGB ejecta with pristine gas in order to be explained. On the contrary, the dynamical model described above is based on the fact that the SNe II have fully cleared the cluster of its pristine gas. A way to solve this problem has been approached in the model by D'Ercole et al. (2008): if the SNe have a preferential direction of ejection, the ejected matter clears out a cone, and leaves a torus of pristine gas in the outskirts of the cluster within the tidal radius. This pristine gas may be re-accreted (in fact the hydro simulation shows that it is re-accreted) and mixes with the AGB ejecta, providing the desired solution. The model provides us with a very simple solution to the problem of the three separate helium populations in NGC 2808. In the most massive clusters, the super-AGB winds are the first to be collected in the cluster core, and the first SG stars are formed by 'pure' super-AGB ejecta, and with a homogeneous, very high, helium content as the observations require. After a while, not only have the ejecta a smaller helium content, since they come from less massive AGBs, but they also become diluted by the pristine matter at standard helium Y ~ 0.24. A result of such a simulation, obtained without any tuning of parameters, is shown in Figure 4, compared with the helium distribution inferred for the stars in NGC 2808 (D'Antona & Caloi, 2008).

Much more will no doubt be learnt about GC formation in the near future, but at least one of the most difficult problems – the mass budget and the loss of the FG stars – might be on the verge of being fully understood.

Acknowledgements

Francesca D'Antona is grateful to Vittoria Caloi, Annibale D'Ercole and Enrico Vesperini for the long-lasting exchange of ideas and work that made this review possible. Francesca D'Antona is deeply indebted to Achim Weiss for presenting the talk on her behalf.

References

Angulo, C. et al. 1999, Nucl. Phys. A, 656, 3
Bedin, L. R. et al. 2004, ApJ, 605, L125
Bekki, K. & Norris, J. E. 2006, ApJ, 637, L109
Bonifacio, P. et al. 2002, A&A, 390, 91
Busso, G. et al. 2007, A&A, 474, 105
Caloi, V. & D'Antona, F. 2005, A&A, 121, 95
Caloi, V. & D'Antona, F. 2007, A&A, 463, 949
Canuto, V. M., Goldman, I. & Mazzitelli, I. 1996, ApJ, 473, 550
Carretta, E. et al. 2005, A&A, 433, 597
Cohen, J. G., Briley, M. M. & Stetson, P. B. 2002, AJ, 123, 2525
Cohen, J. G. & Meléndez, J. 2005, AJ, 129, 303
Cottrell, P. L. & Da Costa, G. S. 1981, ApJ, 245, L79
D'Antona, F. et al. 2002, A&A, 395, 69
D'Antona, F. & Caloi, V. 2004, ApJ, 611, 871
D'Antona, F. & Caloi, V. 2008, MNRAS, 390, 693
D'Antona, F. et al. 2005, ApJ, 631, 868
Decressin, T. et al. 2007a, A&A, 464, 1029
D'Ercole, A. et al. 2008, MNRAS, in press
Gratton, R. G., Sneden, C. & Carretta, E. 2004, ARA&A, 42, 385
Hale, S. E. et al. 2002, Phys. Rev. C, 65, 5801
Iben, I. & Renzini, A. 1983, ARA&A, 21, 271
Ivanova, N. et al. 2008, MNRAS, 386, 553
Karakas, A. & Lattanzio, J. C. 2007, PASA, 24, 103
King, I. R. et al. 1998, ApJ, 492, L37
Moehler, S. et al. 2004, A&A, 415, 313
Moehler, S. et al. 2007, A&A, 475, L5
Norris, J. E. 2004, ApJL, 612, L25
Pasquini, L. et al. 2005, A&A, 441, 549
Pasquini, L. et al. 2008, A&A, 489, 315
Piotto, G. et al. 2005, ApJ, 621, 777
Piotto, G. et al. 2007, ApJ, 661, L53
Pumo, M. L., D'Antona, F. & Ventura, P. 2008, ApJ, 672, L25
Richer, H. B. et al. 2006, Science, 313, 936
Ritossa, C., Garca-Berro, E. & Iben, I. J. 1999, ApJ, 515, 381
Salaris, M. et al. 2006, ApJ, 645, 1131
Siess, L. 2007a, A&A, 476, 893
Siess, L. 2007b, *Why Galaxies Care About AGB Stars: Their Importance as Actors and Probes*, in ASP Conf. Series, 378, 9
Ventura, P. & D'Antona, F. 2005, A&A, 431, 279
Yong, D. & Grundahl, F. 2008, ApJ, 672, L29

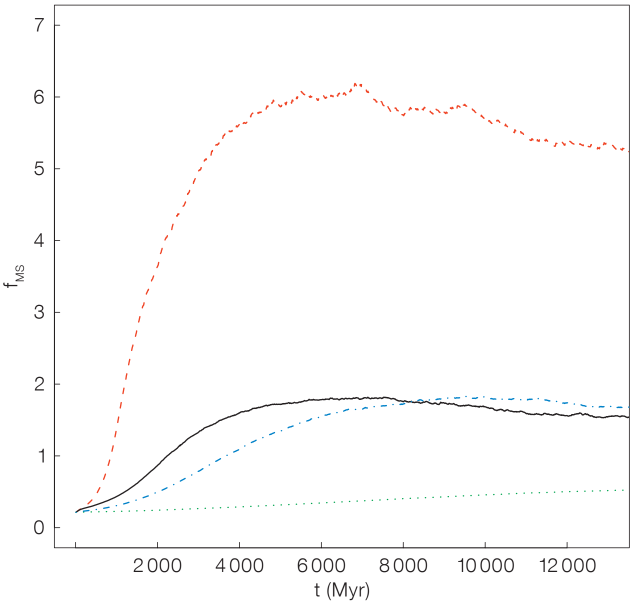

Figure 3 (above). Time evolution of the ratio of the number of second generation (SG) to first generation (FG) main sequence stars with $0.1 < M/M_{\odot} < 0.8$, f_{MS}, for different N-body simulations (D'Ercole et al., 2008). Depending on the initial expansion velocity assumed for the FG (due to the mass loss of SNe II that are more or less concentrated in the cluster core), the SG could even become the dominant cluster population.

Figure 4 (below). The empty histogram represents the number versus helium content distribution (N(Y)) for stars in NGC 2808, derived by D'Antona & Caloi (2008) on the basis of the features of the horizontal branch and main sequence. Three distinct populations are present. The hatched red histogram represents the second generation (SG) formation in a dynamical model in which it is assumed that some pristine gas of the first generation (FG) is in a torus at the periphery of the cluster following the SN II epoch. In this model there is a first phase of SG formation in the core of the cluster when only the super-AGB winds are present, followed by a phase during which the winds are diluted by pristine matter being re-accreted again. This two-phase pattern produces a gap in the N(Y) distribution. No attempt has been made here to fit the two SG populations to the data (adapted from D'Ercole et al., 2008).