



HAL
open science

Canis Major OB1 stellar group contents revealed by Gaia

T. Santos-Silva, H. D. Perottoni, F. Almeida-Fernandes, J. Gregorio-Hetem, V. Jatenco-Pereira, C. Mendes de Oliveira, T. Montmerle, E. Bica, C. Bonatto, H. Monteiro, et al.

► **To cite this version:**

T. Santos-Silva, H. D. Perottoni, F. Almeida-Fernandes, J. Gregorio-Hetem, V. Jatenco-Pereira, et al.. Canis Major OB1 stellar group contents revealed by Gaia. Monthly Notices of the Royal Astronomical Society, 2021, 508, pp.1033-1055. 10.1093/mnras/stab2409 . insu-03678877

HAL Id: insu-03678877

<https://insu.hal.science/insu-03678877>

Submitted on 2 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Canis Major OB1 stellar group contents revealed by *Gaia*

T. Santos-Silva¹,^{*} H. D. Perottoni¹, F. Almeida-Fernandes¹, J. Gregorio-Hetem¹, V. Jatenco-Pereira¹, C. Mendes de Oliveira¹, T. Montmerle², E. Bica³, C. Bonatto³, H. Monteiro⁴, W. S. Dias⁴, C. E. Barbosa¹, B. Fernandes¹, P. A. B. Galli⁵, M. Borges Fernandes⁶, A. Kanaan⁷, T. Ribeiro⁸ and W. Schoenell⁹

¹Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Departamento de Astronomia, São Paulo, SP 05508-090, Brazil

²Institut d'Astrophysique de Paris, F-75014 Paris, France

³Departamento de Astronomia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, RS 91501-970, Brazil

⁴Instituto de Física e Química, Universidade Federal de Itajubá, Av. BPS 1303 Pinheirinho, Itajubá, MG 37500-903, Brazil

⁵Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hillaire, F-33615 Pessac, France

⁶Observatório Nacional, Rua General José Cristino 77, São Cristóvão, Rio de Janeiro, RJ 20921-400, Brasil

⁷Departamento de Física, Universidade Federal de Santa Catarina, Florianópolis, SC 88040-900, Brazil

⁸NOAO, PO Box 26732, Tucson, AZ 85726, USA

⁹GMTO Corporation, 465 N. Halstead Street, Suite 250, Pasadena, CA 91107, USA

Accepted 2021 August 12. Received 2021 August 12; in original form 2020 December 29

ABSTRACT

Canis Major OB1 (CMa OB1) is a Galactic stellar association with a very intriguing star-formation scenario. There are more than two dozen known star clusters in its line of sight, but it is not clear which ones are physically associated with CMa OB1. We use a clustering code that employs five-dimensional data from the *Gaia* DR2 catalogue to identify physical groups and obtain their astrometric parameters and, in addition, we use two different isochrone-fitting methods to estimate the ages of these groups. We find 15 stellar groups with distances between 570 and 1650 pc, including 10 previously known and five new open cluster candidates. Four groups, precisely the youngest ones (< 20 Myr), CMa05, CMa06, CMa07, and CMa08, are confirmed to be part of CMa OB1. We find that CMa08, a new cluster candidate, may be the progenitor cluster of runaway stars. CMa06 coincides with the well-studied CMa R1 star-forming region. While CMa06 is still forming stars, due to the remaining material of the molecular cloud associated with the Sh 2-262 nebula, CMa05, CMa07, and CMa08 seem to be in more evolved stages of evolution, with no recent star-forming activity. The properties of these CMa OB1 physical groups fit well in a monolithic scenario of star formation, with a common formation mechanism, and having suffered multiple episodes of star formation. This suggests that the hierarchical model alone, which explains the populations of other parts of the same association, is not sufficient to explain its whole formation history.

Key words: star: early-type – stars: formation – stars: pre-main-sequence – open clusters and associations: general.

1 INTRODUCTION

Detailed studies of the stellar content of OB associations, as well as star clusters, provide insights into several important issues regarding Galactic structure and evolution. They bring crucial information that allows the identification of different stellar populations in the Galaxy and testing stellar models. Issues such as the origin of binary star populations, differences in initial mass functions, identification of the dominant processes in star formation, and stellar fragmentation may all be enlightened by the study of young populations, while processes of dissolving clusters and chemical Galaxy evolution can be investigated through the study of more evolved stellar groups (Brown 2001).

It is accepted historically that most stars are born in gravitationally bound groups, inside molecular clouds (embedded clus-

ters) with at least 35 stars with a density greater than $1 M_{\odot} \text{pc}^{-3}$ (Lada & Lada 2003). Depending on the formation scenario, most of them (~ 95 percent) should evolve into unbound groups that should dissolve themselves within 10–20 Myr (Lada & Lada 2003; Pfalzner 2009, 2011), becoming field stars or associations, currently unbound (e.g. Melnik & Dambis 2017). In the monolithic scenario of formation, OB associations are the current configurations of systems that were originally much more compact, and have subsequently expanded from a single (singularly monolithic) or several (multiply monolithic) clusters (e.g. Lada & Lada 1991; Brown, Dekker & de Zeeuw 1997; Kroupa, Aarseth & Hurley 2001). The most accepted mechanism responsible for these processes has been the expulsion of residual gas from the embedded clusters by means of stellar feedback, making the clusters supervirial (e.g. Hills 1980; Goodwin & Bastian 2006; Baumgardt & Kroupa 2007).

A number of theoretical (Girichidis et al. 2012; Kruijssen et al. 2012; Dale, Ercolano & Bonnell 2015) and observational studies

* E-mail: thaisfi@gmail.com

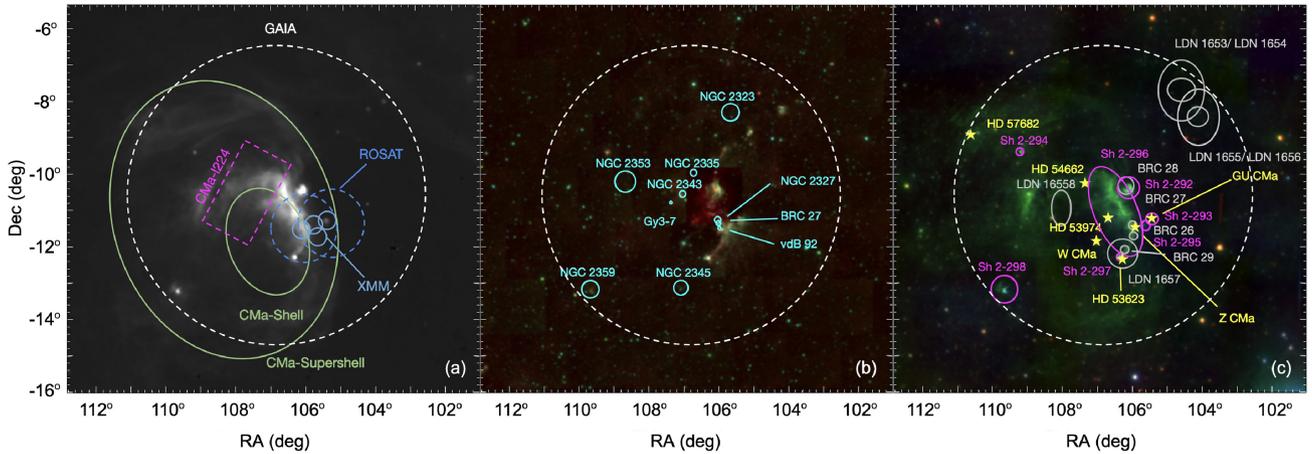


Figure 1. (a) $H\alpha$ image from The Southern $H\alpha$ Sky Survey Atlas (SHASSA) of the stellar association CMA OB1 superposed with regions studied in previous works, including *ROSAT* (dashed blue circles) and *XMM-Newton* (solid blue circles) X-ray fields studied by Gregorio-Hetem et al. (2009) and Santos-Silva et al. (2018), respectively (blue circles); shell and supershell structures identified by Fernandes et al. (2019, grey ellipses); CMA-1224 from Sewilo et al. (2019, magenta rectangle) and with *Gaia* DR2 data in this work (white circle). (b) Combined Wide-field Infrared Survey (WISE) image in W1 (blue), W2 (green), and W4 (red) bands; the cyan circles represent clusters with distances similar to the CMA OB1 association known in the literature (Clariá 1974b; Dias et al. 2002; Kharchenko et al. 2016; Cantat-Gaudin et al. 2018; Bica et al. 2019; Bossini et al. 2019; Liu & Pang 2019). (c) Combined Digital Sky Survey (DSS) image at 700 nm (red), 640 nm (green), and 468 nm (blue). H II regions (Sh 2-292, Sh 2-293, Sh 2-294, Sh 2-295, Sh 2-296, Sh 2-297, and Sh 2-298 (Sharpless 1959) are highlighted in magenta. Dark clouds LDN 1653, LDN 1654, LDN 1655, LDN 1656, LDN 1657, and LDN 1658 (Dobashi et al. 2005) and bright-rimmed clouds BRC 26, BRC 27, BRC 28, and BRC 29 (Sugitani, Fukui & Ogura 1991) are represented in grey. Stars HD 53623, W CMa, Z CMa, and GU CMa and the three runaway stars (HD 53974, HD 54662, and HD 57682) from Fernandes et al. (2019) are shown in yellow.

(Ginsburg et al. 2016; Wright & Mamajek 2018; Ward, Kruijssen & Rix 2020) have, however, disputed this general belief: they have shown that gas exhaustion has been a more efficient mechanism for the dynamical evolution of young clusters than gas expulsion (see Longmore et al. 2014). This is in agreement with a scenario in which the stars are formed through a continuous distribution of densities, following the fractal structure of the gas distribution from which they formed; thus, according to this scenario, most OB associations were never grouped (Ward et al. 2020). Moreover, using *Gaia* data, Melnik & Dambis (2017, 2020) found that most associations are not undergoing expansion. On the other hand, Cantat-Gaudin et al. (2019a) suggest that, although Vela OB2, as well as its stellar distribution, is expanding, this expansion started before the formation of the stars. Vela OB2 is, therefore, an example of an OB association formed globally unbound but nevertheless showing signs of expansion.

Our particular interest is the study of the stellar association Canis Major OB1 (henceforth CMA OB1), located at a distance of ~ 1200 pc (Zucker et al. 2019, 2020). It is composed of more than 200 B stars, a few late-type O stars (Gregorio-Hetem 2008), almost 500 young stellar objects (YSOs: Fischer et al. 2016; Sewilo et al. 2019), and about 400 $H\alpha$ emitters (Pettersson & Reipurth 2019), showing a low fraction of disc-bearing stars (Fernandes et al. 2015; Fischer et al. 2016). This population is mostly related to a reflection nebula association CMA R1, including three connected H II regions (Herbst, Racine & Warner 1978: Sh 2-292, Sh 2-296, Sh 2-297, see Sharpless 1959), four bright-rimmed clouds (BRC 26–29: see Sugitani et al. 1991), six dark clouds (LDN 1653, LDN 1654, LDN 1655, LDN 1656, LDN 1657, LDN 1658: see Dobashi et al. 2005), and more than a dozen clusters (Dias et al. 2002; Kharchenko et al. 2016; Cantat-Gaudin et al. 2018; Bica et al. 2019; Liu & Pang 2019) in its line of sight, which are shown in Fig. 1.

There are at least two open questions related to the CMA OB1 association: one refers to membership of its clusters and the other regards its star-formation history. It is well known that there are

embedded clusters such as NGC 2327 and BRC 27, with age 1.5 Myr, Gy 3-7, with age ~ 2 Myr (Soares & Bica 2002, 2003; Rebull et al. 2013), Z CMa and GU CMa, with ages < 5 Myr and > 10 Myr respectively (Gregorio-Hetem et al. 2009), and vdB 92, possibly a cluster in a dissolution stage, with age ~ 5 –7 Myr (Rebull et al. 2013; Bonatto & Bica 2010). All of these are probably related to CMA OB1, given their measured distances and youth. On the other hand, there are famous clusters like NGC 2353, initially suggested to be the nucleus of CMA OB1, as well as NGC 2343, NGC 2335, and NGC 2323 (Clariá 1974b; Clariá, Piatti & Lapasset 1998), that were discarded and the general conclusion was that these are unrelated to CMA OB1 due to their ages (> 100 Myr).

There have been, however, other contrasting views about the plausible star formation and evolutionary scenario of the CMA OB1 association. One of them, developed by Herbst & Assoua (1977), suggests star formation induced by a supernova explosion (SNE) about 0.5 Myr ago. This scenario is consistent with the Fischer et al. (2016) results on the distribution of YSOs found in the centre of CMA OB1. On the other hand, Reynolds & Ogden (1978) proposed that star formation in the region is triggered by strong stellar winds or an expanding old ‘fossil’ H II region.

Based on the spatial distribution of YSOs of the H II region Sh 2-297 to the west side of CMA OB1, Mallick et al. (2012) argue that the youngest sources in the region are distributed away from the ionizing source, indicating a possible evolutionary sequence. This scenario supports the hypothesis of triggered star formation in this region, which seems to have propagated from the massive ionizing star HD 53623 towards the cold dark cloud LDN 1657A. On the other hand, Sewilo et al. (2019) studied the other side of CMA OB1 (east side: see the rectangle in Fig. 1a), in a region now dubbed CMA-1224, and they suggest that the most likely scenario that explains the star formation observed in this region includes the spontaneous gravitational collapse of filaments.

The star-formation region CMA R1, the most prominent feature of CMA OB1, has also been studied by our group at several wavelengths

for many years. Gregorio-Hetem et al. (2009), based on *ROSAT* data, found two distinct groups: one with ages < 5 Myr, immersed in a region with a high concentration of gas and dust around the Z CMA star, and another on the opposite side, near the GU CMA star, with an older young population (> 10 Myr). Both populations were confirmed by Santos-Silva et al. (2018), using the *XMM-Newton* satellite. The authors not only increased the number of known sources in the region to about 400 X-ray sources, but also, using *2MASS* and *WISE* counterparts and CO maps, proposed a new star-formation scenario including at least two episodes. The first episode occurred more slowly, within the whole region studied, at least 10 Myr ago, dispersing almost all of the present gas, while the second, ongoing episode has been occurring for less than 5 Myr. The latter is dynamically faster, and it takes place in the region where gas is still present, suggesting that the association is going through the final stages of star formation.

Finally, in a recent work, Fernandes et al. (2019) used images at several wavelengths (optical, IR, H I, CO, etc.) to show that CMA OB1 consists of a shell with diameter $D \sim 60$ pc, where the Sh 2-296 nebula is nested in a superbubble 140 pc in diameter. They also found three runaway stars that were probably ejected from approximately the same location within the CMA shell, by at least three successive SNEs, for ~ 6 Myr, ~ 2 Myr and ~ 1 Myr. This suggests that there were more SNEs than the one predicted by Herbst & Assousa (1977). In that work, Fernandes et al. (2019) also show that the O stars in the region cannot, by themselves, be responsible for the nebula heating and they suggest, taking into account the scenario of multiple SNEs, that Sh 2-296 is being heated by X-rays.

These results are in agreement with a scenario that considers a second episode of star formation, as proposed previously by Santos-Silva et al. (2018). However, this cannot explain the older stellar population (> 10 Myr) found by them and by Gregorio-Hetem et al. (2009), which leads us to believe that the association may have an even more intriguing star-formation scenario.

This work was done in the context of the Southern Photometric Local Universe Survey¹ (S-PLUS; Mendes de Oliveira et al. 2019) collaboration. In the future, we plan to apply the method developed in this work to perform a massive search for stellar associations in the S-PLUS catalogue, which will then be characterized in detail. We then intend to compute physical parameters, such as the mass, age, distance, extinction, and metallicity of the populations of the associations, as well as of young star clusters in the Galaxy, using the five Sloan bands and seven narrow bands from the 12-band Javalambre system of S-PLUS.

Aiming at clarifying the complex star-forming history of CMA OB1 and confirming its cluster membership, we conduct a multidimensional study of stellar groups in the region, taking into account the positions, proper motions, and parallax from the *Gaia* DR2 catalogue (Gaia Collaboration et al. 2018).

This work is structured as follows. In Section 2, we present the data used in our analysis. In Section 3, we describe in detail the clustering method we apply to find groups and the comparison with previous results from the literature. In Section 4, we describe the determination of fundamental parameters: age, distance, visual extinction, and metallicity. In Section 5, we report on the stellar populations of the region and in Section 6 we discuss the content of the clusters of CMA OB1 and the star-forming history of the association. Finally, in Section 7 we present a summary of our results and conclusions.

2 DATA

The *Gaia* Mission is an ambitious survey that aims to construct the most accurate 3D map of our Galaxy. Its second data release (hereafter *Gaia* DR2; Gaia Collaboration et al. 2018) provides a five-parameter astrometric solution (α , δ , $\mu_\alpha \cos \delta$, μ_δ , ϖ) with unprecedented precision and photometry in three bands (G , G_{BP} , and G_{RP}) for more than 1.3 billion stars. In addition, there are also radial velocities, astrophysical parameters (stellar effective temperature, extinction, reddening, radius, and luminosity), and variability measurements for part of the sample.

2.1 Sample selection

We queried the *Gaia* DR2 catalogue data from the *Gaia* archive² taking into account the astrometric and photometric constraints detailed below. We selected objects within a search radius of 4.1 degrees centred on the coordinates (RA, Dec.) = (106.7°, -10.6°), covering the entire CMA OB1 association (see the dashed circle in Fig. 1).

To ensure good astrometric and photometric quality of the data, we applied constraints using the re-normalized unit weight error (RUWE). We still took into account the zero-point (0.029 mas) provided by Lindegren et al. (2018), and followed their suggested criteria and also those suggested by Arenou et al. (2018) for the photometric data. Thus we chose only the objects that follow

- (i) $\text{RUWE} \leq 1.4$,
- (ii) $|(\varpi + 0.029 \text{ mas})/\sigma_\varpi| < 5$,
- (iii) $\text{phot_g_mean_flux_over_error} > 50$,
- (iv) $\text{phot_bp_mean_flux_over_error} > 20$,
- (v) $\text{phot_rp_mean_flux_over_error} > 20$,
- (vi) $\text{phot_bp_rp_excess_factor} < 1.3 + 0.0 \cdot (\text{bp_rp})^2$,
- (vii) $\text{phot_bp_rp_excess_factor} > 1.0 + 0.015 \cdot (\text{bp_rp})^2$,
- (viii) $\text{visibility_periods_used} > 8$.

We use, as a further selection criterion, a cut in parallax values, choosing only objects with $0.4 < \varpi$ (mas) < 2.0 in order to ensure that all CMA OB1 potential members are considered and to obtain a knowledge of their close neighbourhood. We emphasize that the association has an estimated distance of around 1200 pc, using *Gaia* data (Zucker et al. 2019, 2020), and our parallax constraints correspond to distances between 500 and 2500 pc. After applying all of these criteria, our final sample contains 249 522 stars.

2.2 Astrometric distances and extinction correction

Recent studies have determined the distance of CMA OB1 quite accurately (Ward et al. 2020; Zucker et al. 2019, 2020) from *Gaia* data. Therefore, for consistency, in this work we also used astrometric distances (D_A) estimated by Bailer-Jones et al. (2018) using *Gaia* DR2 data for all stars selected in the previous section.

As is well known, the association is located in a region of high extinction, and therefore the visual extinction values (A_V) can vary significantly with distance. In order to take this into account, we use the astrometric distance of each star to obtain its A_V from the three-dimensional dust map of BAYESTAR19³ (Green et al. 2019). To consider the probabilistic nature of BAYESTAR19, we used it with mode = *mean*. Furthermore, we applied the corrections from Schlafly & Finkbeiner (2011) to the A_V values.

²<https://gea.esac.esa.int/archive/>

³<http://argonaut.skymaps.info/>

¹www.splus.iag.usp.br

3 SUBSTRUCTURE SEARCH

Especially after *Gaia* DR2 was published, several different techniques based on multidimensional parameter space analysis have been applied to the data to find new and/or confirm existing stellar populations. In particular, these kinds of analysis have been successfully applied in the search for associations (Gagné et al. 2018; Cantat-Gaudin et al. 2019a,b; Liu, Fang & Liu 2020), star-forming regions (Zari, Brown & de Zeeuw 2019; Galli et al. 2020), and open clusters (see Cantat-Gaudin et al. 2018, 2020; Dias et al. 2018; Gao 2018; Liu & Pang 2019; Lodieu et al. 2019; Cantat-Gaudin & Anders 2020; Castro-Ginard et al. 2020).

Clustering methods constitute the most commonly used technique of unsupervised learning and are a powerful tool for data analysis. There are several clustering algorithms (for instance, see Pedregosa et al. 2011, and references therein),⁴ but one shown to be powerful and efficient in different astronomy fields is Hierarchical Density-Based Spatial Clustering of Applications with Noise⁵ (HDBSCAN: Campello, Moulavi & Sander 2013; Campello et al. 2015), which has been used in a variety of contexts (Kounkel & Covey 2019; Kounkel, Covey & Stassun 2020; Limberg et al. 2020; Logan & Fotopoulou 2020; Kuhn et al. 2021). This algorithm is based on DBSCAN (Ester et al. 1996), also used to search for stellar clusters (Castro-Ginard et al. 2018, 2019, 2020). This section summarizes the methodology used to search for stellar populations in CMa OB1, as well as its validation.

3.1 Searching for groups

In this work, we opted for use of the HDBSCAN code in order to search for physical groups of stars in the line of sight of CMa OB1. The most remarkable feature of the tool in the era of large surveys is that it can identify groups with varying densities and arbitrary shapes without the need to specify the number of clusters in the sample, as opposed to, for example, the more commonly used *k*-means algorithm (MacQueen 1967).

HDBSCAN can handle multidimensional data and depends on six main parameters, including three parameters (`min_cluster_size`, `min_samples`, and `cluster_selection_method`) that have significant effects on our work, which we discuss below.

The two primary parameters are the minimum number of objects to be classified as a cluster, `min_cluster_size`, and the minimum number of samples in a neighborhood for a point to be considered as a core point, `min_samples`. In other words, for a larger value of `min_samples`, the cluster will be reduced to a more dense area; consequently, in these cases, there will be fewer clusters and a stronger connection between objects. On the other hand, the smaller `min_samples` values could lead to fragmentation into many small clusters and a decrease in noise. Determining the parameters of `min_samples` and `min_cluster_size` is data-dependent and might be difficult. The `cluster_selection_method` parameter is used to select the clusters from the cluster tree hierarchy. The standard approach is Excess of Mass (EOM) which tends to select one or two of the largest clusters and some smaller clusters. Another option is to use the method `Leaf`, which selects several small and more homogeneous clusters.

We apply HDBSCAN using a PYTHON implementation of the sample selected in Section 2.1 in the 5D space of astrometric

parameters ($\mu_\alpha \cos \delta$, μ_δ , ϖ , α , and δ). The HDBSCAN configuration we adopted is `min_cluster_size = 30`, `min_samples = 60`, and `cluster_selection_method = Leaf`. These selection parameters were used aiming at a search for more homogeneous clusters, while the value of `min_cluster_size` was chosen in order to guarantee that the groups had at least 30 stars, to allow good characterization.

In an experimental phase, we performed tests varying `min_samples` from 20 to 250, in steps of 10, to choose the best value for this parameter. In this set, the number of groups found varies significantly, from 36 to 11,⁶ and the number of groups with distances compatible with CMa OB1 falls from 13 to 4. Farther, for $40 < \text{min_samples} < 80$, the number of groups with distances consistent with the association varies from 8 to 12, thus, the choice for `min_samples = 60` ensures that we are not discarding groups that could possibly be members of CMa OB1.

Using this configuration, the code found 29 groups, the spatial and proper-motion distributions of which, as well as a violin parallax histogram, are shown in Fig. 2.

To validate the clusters identified by HDBSCAN and estimate the membership probability of each star belonging to a specific group, we have computed 400 bootstrap repetitions, taking into the account the uncertainties of the astrometric parameters of the stars. For each repetition, we have used the HDBSCAN function `approximate_predict` to evaluate the cluster to which the star belongs according to the original hierarchical cluster tree. We attribute a membership probability (P) of each star belonging to a specific cluster according to the percentage of assignment to it. We considered that a star is a cluster member if it is assigned to a specific cluster in at least 50 per cent of the realizations ($P \geq 50$ per cent). This method is similar to that one described by Limberg et al. (2020).

3.2 Validation

The method used here to search for stellar groups was evaluated and validated by us by performing two tests that aim to prove that the code can find real structures, already known in the literature, and that it is able to select the correct membership for the groups. First, we compare the spatial distribution of our groups with clusters known in the literature. For the second test, we compare membership of two groups with CMa R1 star-forming region members selected using a Bayesian method.

3.2.1 Cross-matching with groups previously known in the literature

In the first test we compared the spatial distribution of our 29 groups with the angular dimensions of clusters present in four large stellar cluster catalogues: two before the *Gaia* era (Dias et al. 2002; Kharchenko et al. 2016) and two based on *Gaia* data (Liu & Pang 2019; Cantat-Gaudin et al. 2020); see Fig. 3.

There are, respectively, 24, 14, 10, and 14 clusters with parallaxes or distances within the same range adopted here (see Section 2.1) in each one of these catalogues. Seven clusters are present in all four catalogues, another three are present in all but in one (Liu & Pang 2019), four are found only in the two older catalogues, those before

⁶For `min_samples < 30`, the number of groups increases rapidly for smaller values of this parameter, finding about 80 groups for `min_samples = 10`. On the other hand, for `min_samples > 250` the number of groups fluctuates between 9 and 11, up to `min_samples = 350`.

⁴<https://scikit-learn.org/>

⁵<https://hdbscan.readthedocs.io/en/latest/>

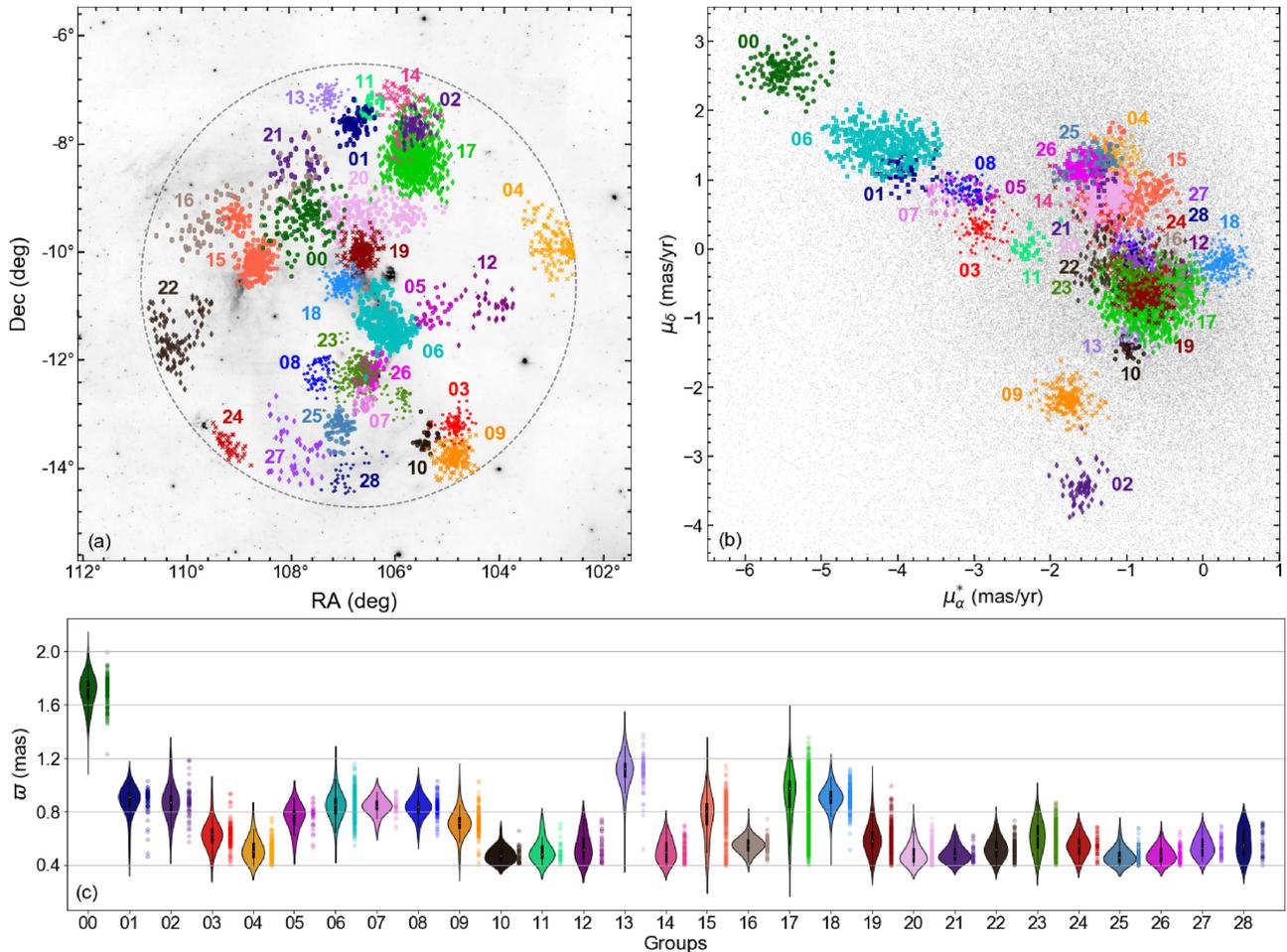


Figure 2. (a) Spatial distribution of groups found by HDBSCAN in the CMa OB1 region on a Digital Sky Survey image (640 nm) of CMa OB1. The dashed circle delimits the selection area from *Gaia* DR2. (b) Proper-motion distribution of groups, including field stars (all sources selected in our sample that were not considered in any group: grey points) $\mu_\alpha^* = \mu_\alpha \cos \delta$. (c) Violin histograms of parallaxes and their exact values (light points on the right side of the histograms) for each group.

publication of the *Gaia* catalogues,⁷ and BDSB 93 is present only in Kharchenko et al. (2016) and Cantat-Gaudin et al. (2020). Finally, another 15 clusters are present in only one of these catalogues: nine in Kharchenko et al. (2016), three in Liu & Pang (2019), two in Cantat-Gaudin et al. (2020), and NGC 2351 is in Dias et al. (2002).

Our comparisons were performed by means of a visual inspection of Fig. 3, in which we prioritize the use of more recent estimates of the angular dimensions of clusters, so that for objects present only in Dias et al. (2002) and/or Kharchenko et al. (2016) we use values available in Bica et al. (2019). For Gulliver 13 and UPK 452, present only in Cantat-Gaudin et al. (2020), we use $r50$ (the radius containing half of the members) provided by the authors, because they have no values for the total sizes of the clusters. For all other objects, we

use $rmax$ (maximum cluster member’s distance to average position) provided by Liu & Pang (2019).⁸

We found 12 clusters, the angular dimensions of which are, as a whole, projected on to at least one of our groups and, therefore, are probably from the same population. These are the clusters highlighted by solid black, blue, and magenta contours in Fig. 3. A total of 11 clusters have a few members of our groups projected within their angular dimension (dashed contours). Another seven clusters are projected on to regions with no overlap with any members of our groups (grey lines). On the other hand, 13 of our groups are located in places where there are no clusters present in the catalogues used here, indicating that they could be new cluster candidates.

Among the clusters having no groups projected in their direction, five of them (FSR 1158, FSR 1159, FSR 1160, FSR 1164, and FSR 1194) present only in Kharchenko et al. (2016) were not detected by HDBSCAN, probably because they are asterisms according to Cantat-Gaudin et al. (2018), and they were not found by catalogues

⁷FSR 1163, FSR 1172, FSR 1180, FSR 1207, FSR 1212, and Berkeley 76 from Kharchenko et al. (2016), NGC 2345 from Dias et al. (2002), and FSR 1170 and Ivanov 4 from both, even having distances smaller than 2500 pc in these catalogues, were excluded of these comparisons because they have parallaxes estimated by Cantat-Gaudin & Anders (2020) using *Gaia* data, outside the range adopted in this work (see Section 2.1).

⁸Only for Alessi 21, also present in this catalogue, do we adopt the radius provided by Bica et al. (2019) to make it clearer in Fig. 3, because the $rmax$ of this cluster is too large (1.83°) and it occupies the same area as many groups and clusters.

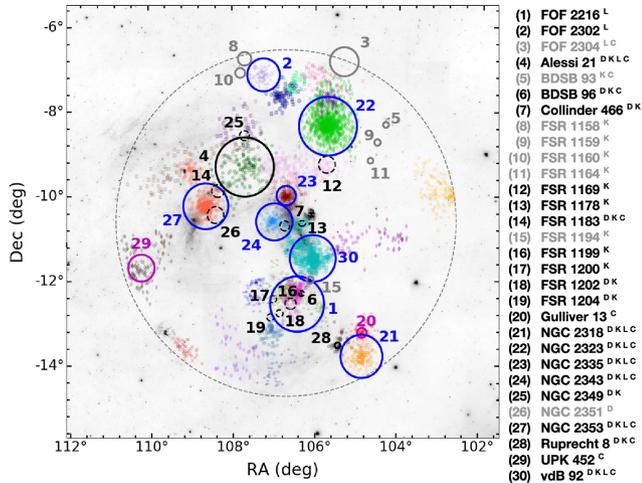


Figure 3. Spatial distribution of 29 groups found by HDBSCAN, following Fig. 2 colours, on a Digital Sky Survey image (640 nm) and 30 clusters from the literature. Solid contours present the angular dimension of clusters known in the literature that coincide with the groups detected in this work. In black are clusters from (D) Dias et al. (2002) and (K) Kharchenko et al. (2016) with angular dimensions from Bica et al. (2019), in blue are clusters from (L) Liu & Pang (2019) and the magenta circles are radii of 50 per cent of cluster members from (C) Cantat-Gaudin et al. (2020). Dashed contours present objects that can be related to a given group and grey circles are clusters from these authors that were not detected by us. The letters indicate in which catalogues each cluster is present.

based on *Gaia* data. Another two clusters, FOF 2304 and BDSB 93, have no groups detected by the code in their lines of sight, likely due to our sample selection criteria (Section 2.1) and our choice of parameters for HDBSCAN (Section 3.1). A more detailed discussion of our groups that are clusters known in the literature is presented in Section 5.1, and clusters not found by HDBSCAN are shown in Appendix A2.

3.2.2 Bayesian method

The second test to validate our results is based on comparison with a recent study of the young population related to our groups. Gregorio-Hetem et al. (2021b, hereafter GH21) used multiband photometry from the T80S telescope of the S-PLUS collaboration (Mendes de Oliveira et al. 2019), covering an area of 2 deg² in the direction of CMa R1. The sample selection was performed by combining a Bayesian model and the cross-entropy technique (see Hetem & Gregorio-Hetem 2019) using astrometric data ($\mu_\alpha \cos \delta$, μ_δ) from *Gaia* DR2.⁹ Their entire sample with *Gaia* DR2 data contains 669 stars, of which 155 are already known in the literature, 395 are new CMa R1 members (with membership probability $P \geq 50$ per cent, determined by them), and 119 are member candidates ($P < 50$ per cent).

We performed coherent cross-matching between CMa R1 members and our groups, selecting 501 stars from GH21 that comply with our selection criteria described in Section 2.1. From our sample, we selected 275 members of CMa06 and 12 from CMa05 within the T80S FOV that fulfil the authors' selection criteria.¹⁰ It is important

⁹The data query was limited to $0.8 < \varpi$ (mas) < 1.25 , $\varpi/\sigma_\varpi > 3$, and $\text{RUWE} < 1.4$.

¹⁰CMa06 has 31 members out of the T80S FOV and another 98 stars with parallax lower than the limit of 0.8 mas adopted by GH21 to select their

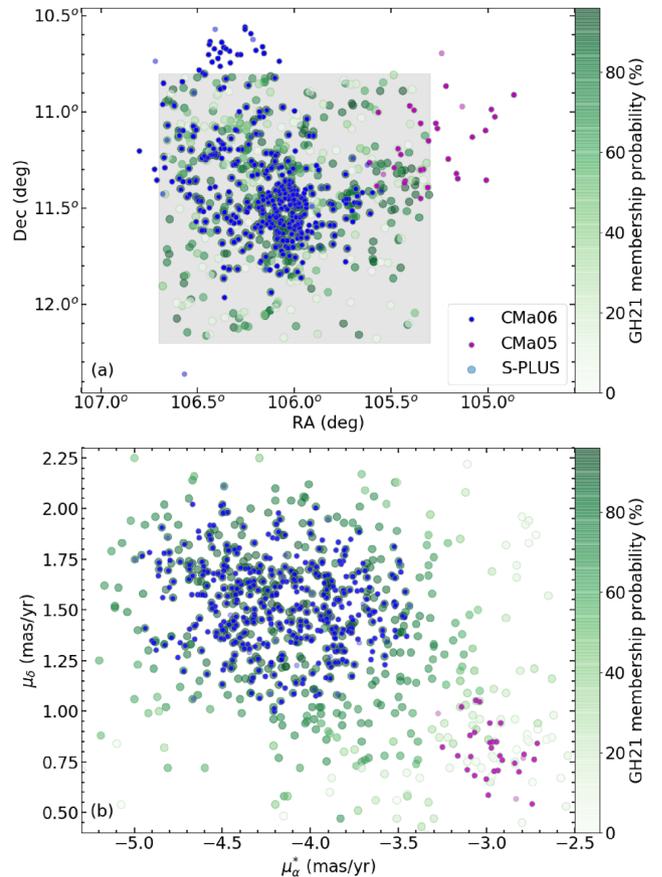


Figure 4. (a) Spatial and (b) proper-motion distributions of CMa06 and CMa05 (blue and pink points, respectively) compared with CMa R1 members from Gregorio-Hetem et al. (2021b) shown by green circles, according to their membership probability. CMa05 and CMa06 members with $P < 50$ per cent are represented by light points. The hatched area in panel (a) represents the region observed by the T80-S telescope.

to note that the T80S survey only covers, spatially, a fraction of our sources identified in CMa05 and CMa06 (see the hatched area in Fig. 4a).

In the first cross-matching, implemented in CMa06, we found 251 members of CMa R1 in our group. We noticed that all these objects were classified as $P > 75$ per cent by GH21, even considering 18 stars with membership probability $P < 50$ per cent obtained by us, showing that our method is more conservative than theirs. Actually, this is not surprising, since, in addition to the proper motion used by the authors to calculate membership probabilities, HDBSCAN also takes into account the parallax and spatial distribution for the selection of groups. On the other hand, HDBSCAN selected, in our group CMa06, about 72 per cent (251/347) of the CMa R1 members with $P > 75$ per cent that follow our selection criteria, or, more conservatively, 85 per cent (201/238) of $P \geq 90$ per cent CMa R1 members, proving that our method was effective in finding high-probability CMa R1 members. Moreover, GH21 estimated, using the cross-entropy technique, the mean proper motion of CMa R1: $\mu_\alpha^* = -4.1 \pm 0.6$ mas yr⁻¹ and $\mu_\delta = 1.5 \pm 0.4$ mas yr⁻¹, which is compatible with our group CMa06, with $\mu_\alpha^* = -4.18 \pm 0.36$ mas yr⁻¹ and $\mu_\delta = 1.52 \pm 0.21$ mas yr⁻¹ (see Table 1).

sample and CMa05 has 14 objects out of the T80S FOV and five with lower parallax.

Table 1. Astrometric parameters of 15 groups found in the CMa OB1 region.

| Group | $N^{(a)}$ (stars) | $N_{50}^{(b)}$ (stars) | $P_{50}^{(c)}$ (%) | RA ^(d) (deg) | Dec. ^(e) (deg) | $\mu_{\alpha}\cos\delta^{(f)}$ (mas yr ⁻¹) | $\mu_{\delta}^{(g)}$ (mas yr ⁻¹) | $\varpi^{(h)}$ (mas) | $D_A^{(i)}$ (pc) | $A_V^{(j)}$ (mag) | $V_t^{(k)}$ (km s ⁻¹) | $\sigma_{V2D}^{(l)}$ (km s ⁻¹) |
|-------|----------------------|---------------------------|-----------------------|--|--|---|---|--|--------------------------------------|--|--------------------------------------|---|
| CMa00 | 152 | 148 | 97 | 107.70 ^{+0.25} _{-0.35} | -9.36 ^{+0.37} _{-0.27} | -5.52 ^{+0.19} _{-0.26} | 2.59 ^{+0.22} _{-0.19} | 1.73 ^{+0.14} _{-0.07} | 571 ⁺²³ ₋₅₁ | 0.27 ^{+0.12} _{-0.13} | 16.6 | 1.2 |
| CMa01 | 66 | 64 | 97 | 106.80 ^{+0.18} _{-0.14} | -7.66 ^{+0.16} _{-0.11} | -3.94 ^{+0.14} _{-0.18} | 1.13 ^{+0.10} _{-0.15} | 0.91 ^{+0.09} _{-0.03} | 1066 ⁺³² ₋₁₃₄ | 0.47 ^{+0.10} _{-0.33} | 20.8 | 4.2 |
| CMa02 | 71 | 65 | 92 | 105.68 ^{+0.13} _{-0.16} | -7.74 ^{+0.13} _{-0.13} | -1.59 ^{+0.13} _{-0.09} | -3.45 ^{+0.21} _{-0.14} | 0.87 ^{+0.08} _{-0.08} | 1120 ⁺⁸⁷ ₋₁₂₉ | 0.53 ^{+0.11} _{-0.09} | 20.1 | 2.6 |
| CMa03 | 103 | 101 | 98 | 104.86 ^{+0.09} _{-0.13} | -13.25 ^{+0.10} _{-0.14} | -2.92 ^{+0.13} _{-0.16} | 0.31 ^{+0.15} _{-0.18} | 0.63 ^{+0.08} _{-0.06} | 1526 ⁺¹²⁰ ₋₂₁₁ | 0.94 ^{+0.31} _{-0.31} | 21.5 | 3.3 |
| CMa05 | 31 | 28 | 90 | 105.32 ^{+0.29} _{-0.14} | -11.20 ^{+0.15} _{-0.21} | -2.98 ^{+0.11} _{-0.15} | 0.79 ^{+0.08} _{-0.14} | 0.78 ^{+0.10} _{-0.05} | 1224 ⁺⁵⁵ ₋₁₃₉ | 1.08 ^{+0.38} _{-0.35} | 18.5 | 1.9 |
| CMa06 | 404 | 377 | 93 | 106.11 ^{+0.14} _{-0.29} | -11.45 ^{+0.17} _{-0.29} | -4.20 ^{+0.35} _{-0.38} | 1.52 ^{+0.19} _{-0.21} | 0.85 ^{+0.09} _{-0.07} | 1147 ⁺⁷⁷ ₋₁₃₃ | 1.18 ^{+0.46} _{-1.09} | 24.3 | 3.4 |
| CMa07 | 34 | 26 | 76 | 106.60 ^{+0.11} _{-0.15} | -12.79 ^{+0.11} _{-0.18} | -3.36 ^{+0.20} _{-0.08} | 0.79 ^{+0.13} _{-0.13} | 0.84 ^{+0.04} _{-0.06} | 1159 ⁺⁵⁵ ₋₆₉ | 0.70 ^{+0.13} _{-0.19} | 19.2 | 1.6 |
| CMa08 | 76 | 64 | 84 | 107.43 ^{+0.11} _{-0.14} | -12.30 ^{+0.20} _{-0.23} | -3.13 ^{+0.19} _{-0.17} | 0.87 ^{+0.13} _{-0.13} | 0.84 ^{+0.05} _{-0.06} | 1162 ⁺⁶² ₋₆₂ | 0.97 ^{+0.22} _{-0.26} | 18.2 | 1.6 |
| CMa09 | 180 | 173 | 96 | 104.87 ^{+0.16} _{-0.17} | -13.81 ^{+0.16} _{-0.17} | -1.78 ^{+0.14} _{-0.13} | -2.18 ^{+0.14} _{-0.14} | 0.71 ^{+0.05} _{-0.08} | 1356 ⁺¹⁰⁰ ₋₁₁₇ | 0.74 ^{+0.19} _{-0.14} | 18.1 | 2.6 |
| CMa13 | 80 | 68 | 85 | 107.27 ^{+0.11} _{-0.17} | -7.19 ^{+0.15} _{-0.12} | -0.91 ^{+0.11} _{-0.11} | -1.22 ^{+0.12} _{-0.11} | 1.13 ^{+0.11} _{-0.07} | 869 ⁺³¹ ₋₅₁ | 0.40 ^{+0.10} _{-0.12} | 6.3 | 0.9 |
| CMa15 | 384 | 365 | 95 | 108.66 ^{+0.13} _{-0.20} | -10.23 ^{+0.13} _{-0.39} | -1.09 ^{+0.18} _{-0.29} | 0.77 ^{+0.18} _{-0.19} | 0.80 ^{+0.22} _{-0.08} | 1204 ⁺⁹⁶ ₋₄₄₀ | 0.46 ^{+0.17} _{-0.31} | 7.6 | 3.6 |
| CMa17 | 1096 | 1039 | 95 | 105.70 ^{+0.21} _{-0.19} | -8.33 ^{+0.20} _{-0.30} | -0.71 ^{+0.26} _{-0.33} | -0.66 ^{+0.28} _{-0.21} | 0.97 ^{+0.25} _{-0.08} | 1003 ⁺⁶⁸ ₋₃₁₇ | 0.65 ^{+0.14} _{-0.21} | 5.0 | 2.1 |
| CMa18 | 227 | 218 | 96 | 107.01 ^{+0.15} _{-0.10} | -10.62 ^{+0.10} _{-0.10} | 0.26 ^{+0.18} _{-0.13} | -0.18 ^{+0.15} _{-0.12} | 0.91 ^{+0.06} _{-0.06} | 1076 ⁺⁵⁶ ₋₇₄ | 0.59 ^{+0.16} _{-0.27} | 1.7 | 0.7 |
| CMa19 | 245 | 233 | 95 | 106.67 ^{+0.16} _{-0.11} | -10.03 ^{+0.16} _{-0.13} | -0.77 ^{+0.14} _{-0.16} | -0.66 ^{+0.13} _{-0.16} | 0.58 ^{+0.08} _{-0.10} | 1654 ⁺²¹³ ₋₂₆₅ | 1.09 ^{+0.32} _{-0.46} | 7.9 | 1.9 |
| CMa23 | 239 | 217 | 91 | 106.66 ^{+0.33} _{-0.23} | -12.35 ^{+0.31} _{-0.24} | -0.96 ^{+0.17} _{-0.19} | -0.28 ^{+0.21} _{-0.16} | 0.62 ^{+0.12} _{-0.11} | 1583 ⁺²²⁰ ₋₃₈₄ | 0.81 ^{+0.12} _{-0.26} | 7.7 | 2.0 |

Note. ^(a) Number of stars in the group; ^(b) number of stars with membership probability $P \geq 50\%$; ^(c) percentage of stars with membership probability $P \geq 50\%$; ^(d) right ascension (ICRS) at Ep. = 2015.5; ^(e) declination (ICRS) at Ep. = 2015.5; ^(f) right ascension proper motion; ^(g) declination proper motion; ^(h) parallax; ⁽ⁱ⁾ astrometric distance from Bailer-Jones et al. (2018); ^(j) visual extinction from three-dimensional maps of dust (Green et al. 2019); ^(k) tangential velocity; ^(l) velocity dispersion in two dimensions (RA and Dec.).

By means of the proper-motion distribution of the CMa R1 members (see Fig. 4b), we also note that some of the CMa R1 candidates, selected within a 3σ distribution around the main cluster, coincide with our group CMa05. For this reason, we performed another cross-matching and found six CMa R1 member candidates ($P < 50$ per cent) in CMa05.

In short, our methodology proved to be quite efficient to find clusters known in the literature, as well as detecting new groups that could potentially be new identified clusters. In addition, comparing stars selected by HDBSCAN in our groups (especially CMa05 and CMa06 with CMa R1 members from GH21 and objects well known in the literature, including bright stars, H α emitters, X-ray sources and YSOs; see the discussion in Appendix B), we verify that our method is also efficient to find the fiducial members. This is particularly true for objects with a high membership probability, since we found 98 per cent of CMa06 members in the T80S FOV fulfilling the criteria selection of GH21, classified by the authors as $P > 75$ per cent. On the other hand, due to our conservative method, we were unable to find 28 per cent of members of CMa R1 with $P > 75$ per cent according to GH21.

4 GROUP CHARACTERIZATION

Focusing on groups with a high probability of being associated with CMa OB1, with a distance of ~ 1200 pc (Zucker et al. 2019, 2020), and avoiding groups that may not have been well determined by our choice of HDBSCAN parameter (see Appendix A), we select a sample of 15 physical groups. These 15 groups are those in which more than 75 per cent of their members have membership probability $P \geq 50$ per cent and $D_A < 1700$ pc (see Table 1). In order to understand these groups better, we derived their astrometric parameters using a statistical approach and applied two algorithms, which use different isochrone-fitting methods, combined with the *Gaia* DR2 photometric data to determine fundamental parameters:

the age, distance, visual extinction, and metallicity of our groups. All parameters are derived considering only members with $P \geq 50$ per cent.

A more detailed discussion about unreliable groups is presented in Appendix A1. We also show a list of all possible physical groups that were not well determined in this work, but deserve to be studied in a future work, considering different ranges of parameters in the sample selection.

4.1 Astrometric parameters

After determining all groups using HDBSCAN, we calculate their astrometric parameters from the distribution of the individual parameters for objects with membership probabilities greater than 50 per cent in each group. The position, proper motion, and parallax attributed to each group are given by the 50th percentile (median), and for their respective errors we assume the 16th–84th percentile ranges. The astrometric distances and the visual extinction of our groups are derived in the same way using values described in Section 2.2. Furthermore, we calculate the tangential velocity of each star from its individual proper motions and astrometric distances. We assigned the tangential velocity of each group from the 50th percentile also, and the two-dimensional velocity dispersion from the standard deviation of the tangential velocity distribution. All these parameters for each group are presented in Table 1.

4.2 Fundamental parameters

The first code that we adopted to obtain the fundamental parameters, developed by Monteiro et al. (2020) and henceforth referred to as the M20 code, is based on the cross-entropy continuous multi-extremal optimization method (CE), which takes into account the astrometric membership of the stars obtained in Section 3.1 as well as the nominal errors of the data. The M20 code uses theoretical isochrones from

Table 2. Fundamental parameters obtained from isochrone fitting.

| Group | Age ^M (Myr) | D _P ^M (pc) | A _{VP} ^M (mag) | [Fe/H] ^M | Age ^f (Myr) | D _P ^f (pc) | A _{VP} ^f (mag) | [Fe/H] ^{f*} | Cluster name |
|-------|---------------------------|-------------------------------------|---------------------------------------|---------------------|--------------------------------------|---------------------------------------|--|---|---------------------------|
| CMA00 | 85 ± 21 | 570.0 ± 1.0 | 0.34 ± 0.05 | 0.00 ± 0.07 | 70 ⁺⁵ ₋₂₀ | 549.3 ^{+1.3} _{-6.3} | 0.31 ^{+0.01} _{-0.05} | 0.07 ^{+0.05} _{-0.05} | Alessi 21 ^{DKLC} |
| CMA01 | 172 ± 36 | 1059 ± 19 | 0.42 ± 0.08 | 0.07 ± 0.09 | 200 ⁺¹³ ₋₁₃ | 1015 ⁺¹⁶ ₋₄₃ | 0.43 ^{+0.10} _{-0.02} | 0.05 ^{+0.03} _{-0.01} | |
| CMA02 | 137 ± 67 | 1071 ± 35 | 0.99 ± 0.12 | -0.15 ± 0.12 | 125 ⁺¹³ ₋₁₃ | 1020 ⁺⁷⁶ ₋₁₁₃ | 1.25 ^{+0.09} _{-0.05} | -0.48 ^{+0.32} _{-0.04} | |
| CMA03 | 536 ± 419 | 1531 ± 35 | 1.71 ± 0.17 | -0.12 ± 0.11 | 650 ⁺²⁵ ₋₂₅ | 1542 ⁺⁵⁸ ₋₇₆ | 1.18 ^{+0.19} _{-0.05} | 0.30 ^{+0.01} _{-0.01} | Gulliver 13 ^C |
| CMA05 | 17 ± 27 | 1229 ± 28 | 1.27 ± 0.16 | -0.14 ± 0.06 | 18.0 ^{+7.0} _{-2.0} | 1130 ⁺⁹ ₋₃₃ | 1.11 ^{+0.05} _{-0.03} | -0.10 ^{+0.04} _{-0.02} | |
| CMA06 | 10.1 ± 1.0 | 1099 ± 29 | 0.96 ± 0.09 | -0.15 ± 0.05 | 9.0 ^{+0.5} _{-0.5} | 1069 ⁺²⁵ ₋₈ | 1.18 ^{+0.04} _{-0.03} | 0.07 ^{+0.02} _{-0.01} | VdB 92 ^{DKLC} |
| CMA07 | 13.1 ± 1.5 | 1149 ± 45 | 0.67 ± 0.08 | -0.03 ± 0.11 | 14.0 ^{+1.0} _{-1.0} | 1092 ⁺⁴ ₋₅ | 0.74 ^{+0.02} _{-0.03} | 0.07 ^{+0.02} _{-0.02} | FOF 2216 ^L |
| CMA08 | 18 ± 5 | 1138 ± 46 | 1.01 ± 0.07 | 0.05 ± 0.09 | 18.0 ^{+1.0} _{-1.0} | 1005 ⁺⁵⁵ ₋₁₂ | 0.99 ^{+0.01} _{-0.05} | 0.05 ^{+0.03} _{-0.01} | |
| CMA09 | 329 ± 186 | 1346 ± 9 | 1.08 ± 0.29 | 0.07 ± 0.21 | 325 ⁺¹³ ₋₁₃ | 1219 ⁺¹⁰⁹ ₋₂₁ | 1.08 ^{+0.13} _{-0.11} | 0.14 ^{+0.02} _{-0.01} | NGC 2318 ^{DKLC} |
| CMA13 | 208 ± 63 | 873 ± 4 | 0.27 ± 0.07 | 0.07 ± 0.04 | 225 ⁺¹³ ₋₁₃ | 855 ⁺¹⁹ ₋₆₁ | 0.37 ^{+0.05} _{-0.03} | 0.02 ^{+0.03} _{-0.05} | FOF 2302 ^L |
| CMA15 | 114 ± 36 | 1074 ± 63 | 0.53 ± 0.06 | -0.13 ± 0.11 | 150 ⁺¹³ ₋₁₃ | 1251 ⁺⁵³ ₋₆₂ | 0.45 ^{+0.05} _{-0.02} | 0.28 ^{+0.01} _{-0.01} | NGC 2353 ^{DKLC} |
| CMA17 | 163 ± 81 | 986 ± 34 | 0.91 ± 0.11 | -0.15 ± 0.06 | 175 ⁺¹³ ₋₁₃ | 933 ⁺¹⁹ ₋₇₄ | 0.86 ^{+0.10} _{-0.05} | 0.07 ^{+0.07} _{-0.02} | NGC 2323 ^{DKLC} |
| CMA18 | 181 ± 43 | 1024 ± 25 | 0.63 ± 0.10 | -0.03 ± 0.06 | 150 ⁺¹³ ₋₁₃ | 1050 ⁺²⁴ ₋₇₇ | 0.59 ^{+0.09} _{-0.05} | 0.14 ^{+0.02} _{-0.01} | NGC 2343 ^{DKLC} |
| CMA19 | 42 ± 30 | 1446 ± 99 | 1.54 ± 0.11 | -0.09 ± 0.11 | 60 ⁺⁵ ₋₁₅ | 1413 ⁺⁶⁶ ₋₄₃₆ | 1.24 ^{+0.01} _{-0.10} | 0.30 ^{+0.01} _{-0.20} | NGC 2335 ^{DKLC} |
| CMA23 | 131 ± 65 | 1374 ± 99 | 1.20 ± 0.10 | -0.16 ± 0.13 | 150 ⁺²⁵ ₋₁₃ | 1552 ⁺³⁹ ₋₁₈₂ | 1.20 ^{+0.09} _{-0.05} | 0.14 ^{+0.02} _{-0.01} | |

Notes. ^M Parameters: age, photometric distance (D_P), visual extinction (A_{VP}) and metallicity, obtained using M20 code, see also Appendix C1; ^f same parameters obtained using FITCMD, see Appendix C2; (*) metallicity: $[Fe/H] = \log_{10}(Z/Z_{\odot})$. Clusters found in ^(D) Dias et al. (2002); ^(K) Kharchenko et al. (2016); ^(L) Liu & Pang (2019); ^(C) Cantat-Gaudin et al. (2020), Cantat-Gaudin & Anders (2020), and Dias et al. (2021).

the Padova (PARSEC version 1.2S) database of stellar evolutionary tracks and isochrones (Bressan et al. 2012) fitted to the *Gaia* DR2 G_{BP} and G_{RP} photometric data. A more detailed description of the code is provided in Appendix C1.

For the M20 code, we applied the priors for distance and visual extinction to the mean values present in Table 1. For $[Fe/H]$ we use the default prior of the code, estimated from the Galactic metallicity gradient published by Donor et al. (2020). The age, distance, A_V , and $[Fe/H]$ for the 15 groups are presented in the first four columns in Table 2. Additional parameters estimated by the M20 code are given in Table C1.

The second code we use here, developed by Bonatto (2019) and known as FITCMD, is a statistical approach to extract fundamental parameters of star clusters from the photometric information contained in observed colour–magnitude diagrams (CMDs). FITCMD searches for physical parameters able to build a synthetic CMD, based on the properties of the initial mass function (IMF) obtained from isochrones, also from the Padova database, that best reproduce the observed one. The detailed description of FITCMD is presented in Appendix C2.

The FITCMD input parameters used by us here were the following: ages from 1 Myr–1 Gyr; apparent distance modulus (DM) from 8.5–12 mag; colour excess (CEX) from 0–9 mag; metallicities from 0.1 Z_{\odot} –2.5 Z_{\odot} with $[\alpha/Fe] = 0.0$; the cluster mass (M_{cl}) was default (from 1 M_{\odot} – $10^5 M_{\odot}$). The DM range corresponds to the distances adopted here (500–2500 pc) and metallicities are the values expected for open clusters in the Milky Way disc. The partial results of FITCMD are also given in Table 2 (last four columns) and Table C2 presents additional parameters estimated by this code.

4.3 Overall comparison among groups

The results in Table 2, evaluated by us with the M20 code and FITCMD, show that the groups found in the neighbourhood of

CMA OB1 have ages ranging from 9 Myr to ~ 650 Myr and photometric distances (D_P) varying from ~ 550 pc to ~ 1600 pc. The interstellar visual extinction in the direction of these groups is less than 2 mag, with photometric values (A_{VP}) ranging from 0.27–1.71 mag. Metallicity obtained from the M20 code varies in range $-0.16 < [Fe/H] < 0.07$, while with FITCMD we obtain values in the range $-0.48 < [Fe/H] < 0.28$.

Most of the parameters obtained by the two methods are in good agreement and are consistent with each other (see Fig. 5). The ages of all groups differ by less than 30 per cent from one method to another. Distances have values determined by both methods with a percentage difference smaller than the other three parameters; only for the groups CMA08, CMA15, and CMA23 is this difference greater than 10 per cent. However, CMA00 and CMA05 have incompatible distances between the two methods, mostly because both codes determined small errors for these measurements. The visual extinction of all groups is also compatible between the methods, except for CMA03, which has A_V determined by the M20 code almost 45 per cent higher than the value estimated by FITCMD.

Metallicity is the parameter that has the largest differences between values determined by the two methods. This can be explained in part by the use of a prior in metallicity in the M20 code, which is based on the metallicity gradient of the Galaxy (Appendix C1 and Donor et al. 2020; Monteiro et al. 2020), while for FITCMD no prior is used. The range used to fit this parameter is $-0.9 < [Fe/H] < 0.7$ for the M20 code and $-1.0 < [Fe/H] < 0.4$ for FITCMD (typical for open clusters, see Sections 4.2 and C2). It is important to note that, due to the low sensitivity of the data to metallicity, the M20 code fits values distant from the gradient used as prior only if the weight of the evidence is significant. However, the values found by both methods are compatible for most groups, since the errors are significantly large in either case. Moreover, when looking at the other panels in Fig. 5, we notice that the different metallicity values do not affect age and distance determinations significantly and may

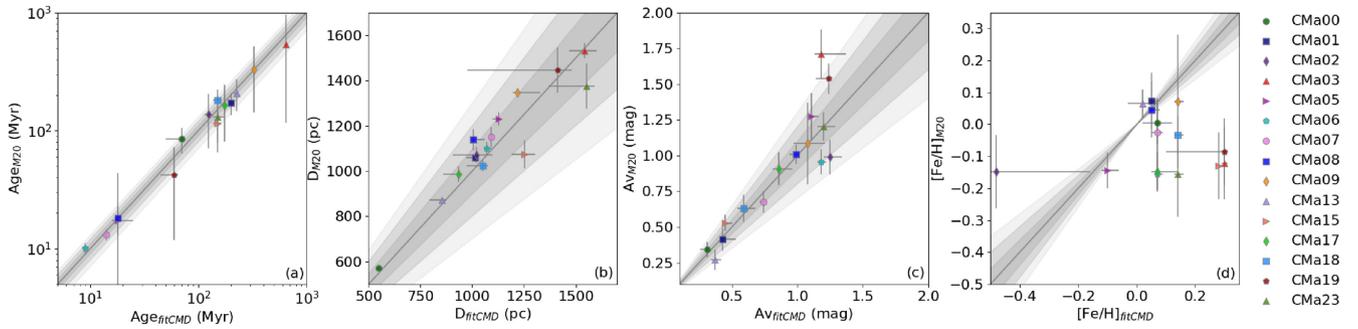


Figure 5. Comparison between fundamental parameters obtained using both isochrone-fitting methods, M20 code and FITCMD: (a) ages, (b) distances, (c) visual extinctions, and (d) metallicities for 15 groups. The dark grey line represents equality between the two methods while the hatched areas show differences of 10 per cent (grey), 20 per cent (light grey) and 30 per cent (very light grey) from this line.

only have a small effect on the determination of visual extinction of some groups such as CMA02, CMA03, and CMA019, which are still within a 30 per cent difference when taking uncertainties into account.

To validate the ages, distances, and visual extinction determined by the isochrone fittings, in Fig. 6 we compared them with parameters available in the literature (Dias et al. 2002; Kharchenko et al. 2016; Liu & Pang 2019; Cantat-Gaudin et al. 2020) for 10 open clusters compatible with our groups (see Section 5.1), including parameters of eight clusters determined by Dias et al. (2021) using the M20 code in objects with membership probabilities determined by Cantat-Gaudin et al. (2020). Astrometric distances and visual extinction from 3D maps of dust extinction for each group are also compared in the middle and bottom panels, respectively.

Four of our groups, CMA05, CMA06, CMA07, and CMA08, have young populations with ages under 20 Myr. Most groups have intermediate ages, older than 100 Myr and younger than 500 Myr. CMA03 is the oldest cluster with age ~ 600 Myr. This parameter is the one with the greatest divergence from the values in the literature catalogues for most of the clusters compared with our groups. However, almost all groups have some ages similar to those found in the literature, except CMA07 and CMA13, which have discrepant values provided for their corresponding clusters that are present only in Liu & Pang (2019).

The isochronal distances of our groups have similar values determined by the two methods and are also within the range of astrometric distances (Table 1), while only the A_V determined for CMA02 by both methods and the A_V from M20 code are outside the range of visual extinction obtained from the 3D maps. Both parameters of our groups are also in good agreement with the literature, although VdB 92, comparable with CMA06, has distance overestimated in the two catalogues before *Gaia*. The cluster corresponding to CMA09, NGC 2318, also has distance overestimated by Kharchenko et al. (2016) and is also the only one with the highest visual extinctions provided by the two catalogues before *Gaia*.

It is important to emphasize that all recent parameters estimated by Dias et al. (2021) using one of the same codes as us (M20 code) are similar to those found in our groups, except for the ages of VdB 92 (~ 6 Myr) and NGC 2323 (~ 470 Myr), reinforcing that some groups were detected well by HDBSCAN, selecting members that preserve the same characteristics as clusters known in the literature with membership determined by another method.

These comparisons helped us understand better the relationship between our groups and some literature clusters, as discussed in the following section.

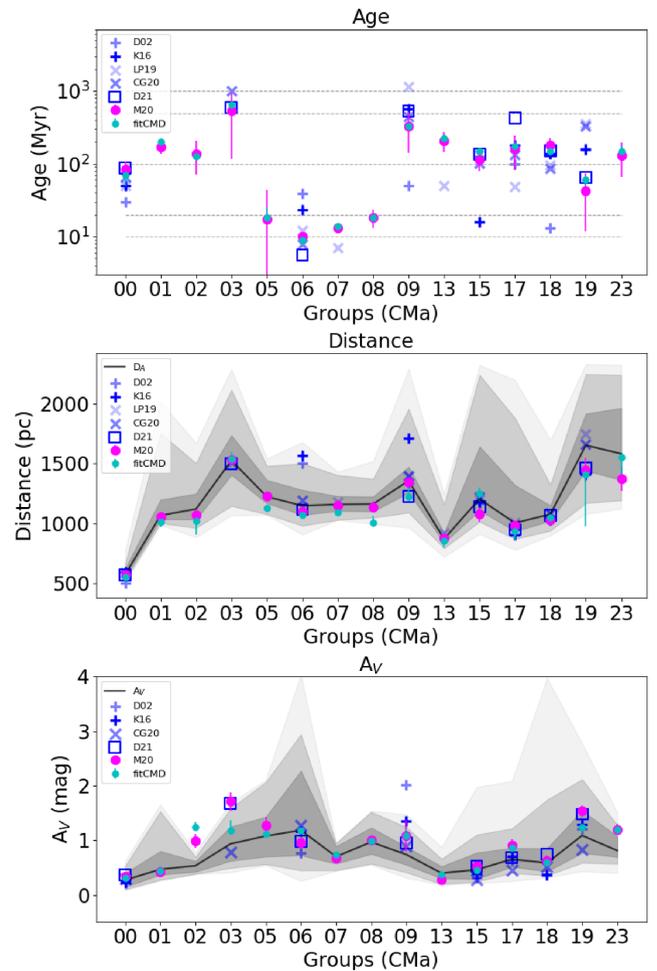


Figure 6. Comparison of ages (top panel), distances (middle panel), and visual extinction (bottom panel) for 15 groups in the line of sight to CMa OB1 ($D_A < 1700$ pc) derived using both M20 code (magenta points: Section C1) and FITCMD (cyan points: Section C2). Parameters from the literature: Dias et al. (2002, D02), dark, and Kharchenko et al. (2016, K16), light blue crosses; Liu & Pang (2019, LP19), light, and Cantat-Gaudin et al. (2020, C20) blue ‘X’ markers, and Dias et al. (2021, D21), blue squares. Grey dashed lines indicate ages of 10, 20, 100, and 500 Myr and 1 Gyr. Mean values of astrometric distances (D_A) and mean visual extinction from 3D maps (A_V) are presented by black lines and their errors (1σ , 2σ , and 3σ , see Section 3.1) are presented by grey shaded areas.

5 STELLAR CONTENT

We employ parameters evaluated using astrometric and photometric data (see Tables 1 and 2) to recognize and confirm that some groups found in this work are, in fact, clusters already known in the literature or new candidates.

5.1 Known open clusters

We consider that our groups are previously known clusters if all astrometric information for them is compatible with objects present in at least one stellar cluster catalogue discussed in Section 3.2.1, preferably in the catalogues with parameters determined from *Gaia* DR2 data (Liu & Pang 2019; Cantat-Gaudin et al. 2020). Among the 15 groups discussed in Section 4.3, we confirm that 10 of them are associated with already known open clusters.

The CMa00, CMa03, CMa09, CMa17, CMa18, and CMa19 groups were recognized to be the clusters Alessi 21, Gulliver 13, NGC 2318, NGC 2323, NGC 2343, and NGC 2335, respectively, having all five parameters similar to those provided in both catalogues based on *Gaia* DR2 data. In addition, similar ages were found in at least two catalogues, considering also Dias et al. (2002, 2021) and Kharchenko et al. (2016), and helped us to confirm them, except for Gulliver 13 and NGC 2335, which had ages estimated only by Dias et al. (2021) similar to CMa03 (~ 600 Myr) and CMa19 (~ 50 Myr), respectively. Gulliver 13 has an age of 1 Gyr according to Cantat-Gaudin et al. (2020), and NGC 2335 has an age of about 160 Myr provided by both catalogues before *Gaia* and ~ 350 Myr by Liu & Pang (2019) and Cantat-Gaudin et al. (2020).

The CMa06 and CMa15 groups also have all five parameters compatible with the vdB 92 and NGC 2353 clusters, respectively. However, their number of objects and spatial distribution suggests that these clusters are only part of their corresponding group (see Fig. 3). CMa06 has 377 members with $P \geq 50$ per cent distributed across the molecular cloud at CMa R1, in which at least almost 270 of them have $P > 75$ per cent membership in the star-forming region determined by the Bayesian method (see Section 3.2.2), while vdB 92 has less than 200 members in all catalogues. Thus, we consider CMa06 as a CMa R1 population containing vdB 92. On the other hand, CMa15 appears to have two different spatially distributed subgroups, which we call CMa15-A and CMa15-B, with the same proper motion and parallax, showing a slightly elongated distribution (see Fig. 2). The most populated subgroup (CMa15-A) is clearly NGC 2353, following the same steps to recognize other known clusters and including ages found in two catalogues that are the same as we found for CMa15 (~ 125 Myr). However, a more detailed study of this group is necessary to state whether CMa15-B is, in fact, part of NGC 2353.

Finally, CMa07 and CMa13 have astrometric parameters very similar to FOF 2216 and FOF 2302 from Liu & Pang (2019). Although this catalogue does not have photometric parameters for comparison, CMa13 has almost the same amount of members as FOF 2302, so it was enough for us to consider that both are the same object. On the other hand, CMa07 has fewer members (26 stars with $P \geq 50$ per cent) than the cluster (54 stars) and is spatially smaller (see Fig. 3). Furthermore, although the age of FOF 2216 derived by Liu & Pang (2019) is lower than our estimates for CMa07, the ages assigned to both are less than 20 Myr, encouraging us to consider that CMa07 should be part of FOF 2216.

5.2 New cluster candidates

In addition to the known clusters, another five groups with $D_A < 1700$ pc (CMa01, CMa02,¹¹ CMa05, CMa08, and CMa23) are not reported in any catalogue discussed here. Thus, we classified them as new cluster candidates. We suggest that CMa 08 and CMa 05 (< 20 Myr) are young stellar clusters. CMa01, CMa02, and CMa23 have ages varying between 125 Myr and 200 Myr and have been recognized by us as in the intermediate evolutionary stage. Therefore, in our next work, we propose a more detailed characterization of these groups, based on multiwavelength data from the S-PLUS collaboration, to confirm that these are new open clusters.

Moreover, there are four distant groups ($D_A > 1700$ pc) not related to clusters known in the literature: CMa04, CMa11, CMa20, and CMa26 (see Appendix A1). Most of them have a very asymmetric distribution of parallaxes, indicating an absence of objects with $\varpi < 0.4$ mas (see Fig. 2c). Although these groups cannot be connected to the CMa OB1 association, some of them may be new open clusters. However, it is necessary to apply our tool considering other parallax and spatial ranges to confirm them. Astrometric parameters of these clusters are present in Table A1.

6 DISCUSSION

Zucker et al. (2019) estimated an astrometric distance, based on *Gaia* DR2 data, of about 1200 pc for CMa OB1 and Fernandes et al. (2019) showed that this association is formed by a shell with a diameter of ~ 60 pc, containing the CMa R1 star-forming region, nested in a supershell with 140 pc diameter (see left panel in Fig. 1). Moreover, Fernandes et al. (2019) suggested that three supernova explosions occurred in the region: ~ 1 , ~ 2 , and ~ 6 Myr ago. Santos-Silva et al. (2018) and Gregorio-Hetem et al. (2009) also noticed a mixture of populations younger than 5 Myr, with one older than 10 Myr in CMa R1, on the west side of CMa OB1.

Starting from the assumption that the groups composing the association must have similarities between them and following the example of both families found by Yalyalieva et al. (2020) in the Sco OB1 association, we classified the 15 groups in the CMa OB1 neighbourhood ($D_A < 1700$ pc) into three sets of groups, according to their proper motion (see Fig. 7). We then evaluated each set taking into account the distances, ages, spatial distribution, and tangential velocity of each group (see Fig. 8).

Fig. 8 shows these sets separately. Set A, Set B, and Set C are shown in the panels on the left, middle, and right, respectively. The proper-motion distribution of each one is presented considering the astrometric distances of the groups in the first row and the ages in the second row. The spatial distribution and the tangential velocity vectors of each group are presented in the third and fourth rows, respectively. Each row follows a specific colour map.

Bearing in mind that the CMa supershell can also be at least 140 pc deep, we consider as possible CMa OB1 ($D \sim 1200$ pc) contents those groups with astrometric distances in the range 1000–1400 pc, in order to ensure that all groups having distances compatible with the association were not previously discarded. They are presented in the first row of Fig. 8 as blue groups. The foreground groups having $D_A < 1000$ pc (cyan) and the background groups with $D_A > 1400$ pc (magenta) were also analysed. Taking into account different evolutionary stages among our groups, we also highlight, in different colours, four main ranges of ages in the second row of

¹¹Before this work was accepted for publication, CMa01 and CMa02 were discovered by He et al. (2021).

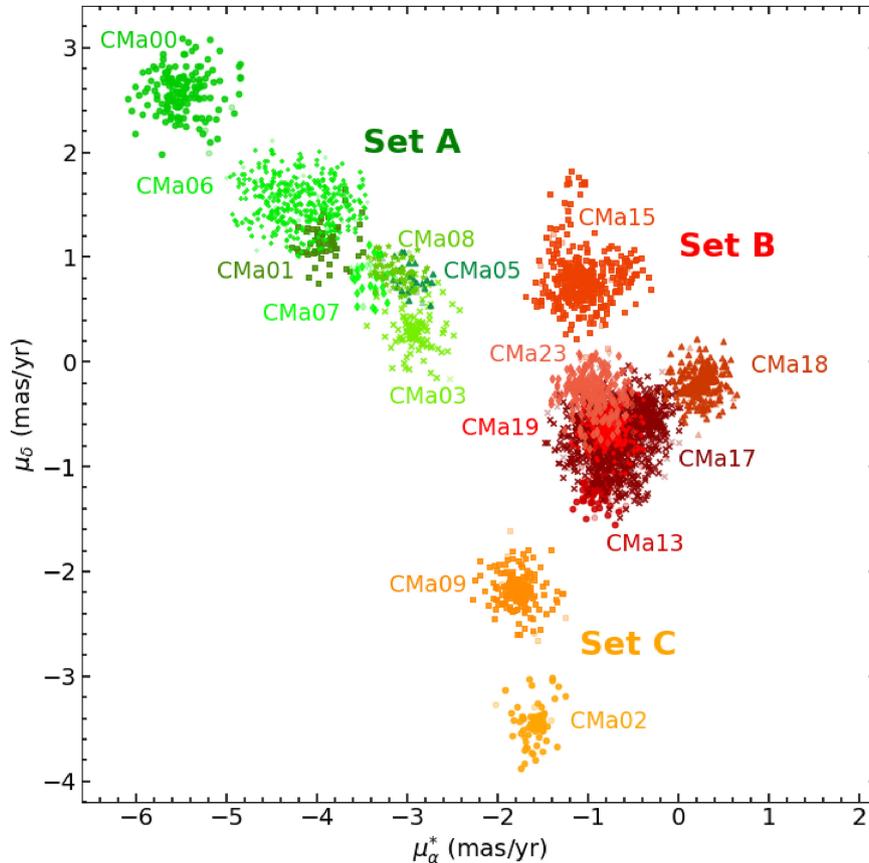


Figure 7. Proper-motion distributions of 15 groups with $D_A < 1700$ pc in the CMa OB1 association region. Green points represent groups in Set A, red points are Set B groups and orange points are Set C. Light symbols present members with $P < 50$ per cent.

Fig. 8: very young groups, with less than 20 Myr (T0-20: dark blue symbols); young groups, from 20–100 Myr (T20-100: blue); groups in an intermediate evolutionary stage, from 100–500 Myr (T100-500: cyan); and older groups, over 500 Myr (T500-1000: magenta). Note that there are no objects older than 1 Gyr. In the third row, the groups were coloured according to their location in relation to the CMa shell and CMa supershell (Fernandes et al. 2019). Although we have no radial velocities for most stars (see Table C1), to get some insight into the internal structure and internal kinematics of the groups we have highlighted the tangential velocity vectors, in the last row, according to the 2-D velocity dispersion (see Table 1).

Hereafter, we evaluate each set from Figs 7 and 8 until we find the CMa OB1 group contents and discuss the formation history of the association based on these groups and previous works.

6.1 Set of groups

Considering the proper-motion distribution in Fig. 7, one notices that Set A forms an elongated structure composed of seven groups with CMa00 at the top, CMa03 at the bottom, and the other groups, CMa01, CMa06, CMa05, CMa07, and CMa08, between the two (green symbols). Set B has six groups, in which CMa13, CMa17, CMa19, and CMa23 are mixed, constituting the main structure with CMa15 above them and CMa18 on the right side (red symbols). The other two groups, CMa02 and CMa09, having lower values of proper motion in declination ($\mu_\delta < -1.5$ mas yr $^{-1}$), are included in Set C (orange symbols).

Set A has five groups (CMa01, CMa05, CMa06, CMa07, and CMa08) with distances comparable to the CMa OB1 association ($1000 < D_A$ pc < 1400), and there is one (CMa00) in the foreground and another (CMa03) in the background. Set A is younger than the other two sets, having four groups (CMa05, CMa06, CMa07, and CMa08) younger than 20 Myr, although it also has an older group, CMa03, the only T500-1000. Another two groups, CMa00 and CMa01, are in distinct evolutionary stages T20-100 and T100-500, respectively. In the projected spatial distribution, the CMa06 and CMa05 groups appear to be connected to each other at the edge of the CMa shell (see Figs 8 and 9). The CMa07 and CMa08 groups are on the east side of them within this structure, with CMa08 almost in the centre. CMa00 and CMa01 are to the north of them, outside the CMa shell. CMa00, as well as CMa05, is inside the CMa supershell, while CMa01 and CMa03 (in the south-west) are outside both structures. As expected from our own set of group selection criteria, based on their proper motion, we can see by the projection of the tangential velocity vectors (last panel on the left side of Fig. 8) that almost all groups of this set are moving in the same direction and with a similar tangential velocity (~ 20 km s $^{-1}$); see Table 1. However, CMa00 has lower 2-D velocity dispersion ($\sigma_{v2D} \sim 1$ km s $^{-1}$). CMa05, CMa07, and CMa08 have intermediate σ_{v2D} (between 1.5 and 2.0 km s $^{-1}$), while CMa03 and CMa06 have $3.0 < \sigma_{v2D}$ (km s $^{-1}$) < 3.5 . CMa01 is the only group with $\sigma_{v2D} > 4.0$ km s $^{-1}$. Half of Set B (CMa15, CMa17, and CMa18) have similar distances to CMa OB1. CMa19 and CMa23 are background groups, while CMa13 is in the foreground. This set of groups is mainly composed of groups with intermediate ages between 100 and 500 Myr. Only CMa19 is T20-100. Concerning

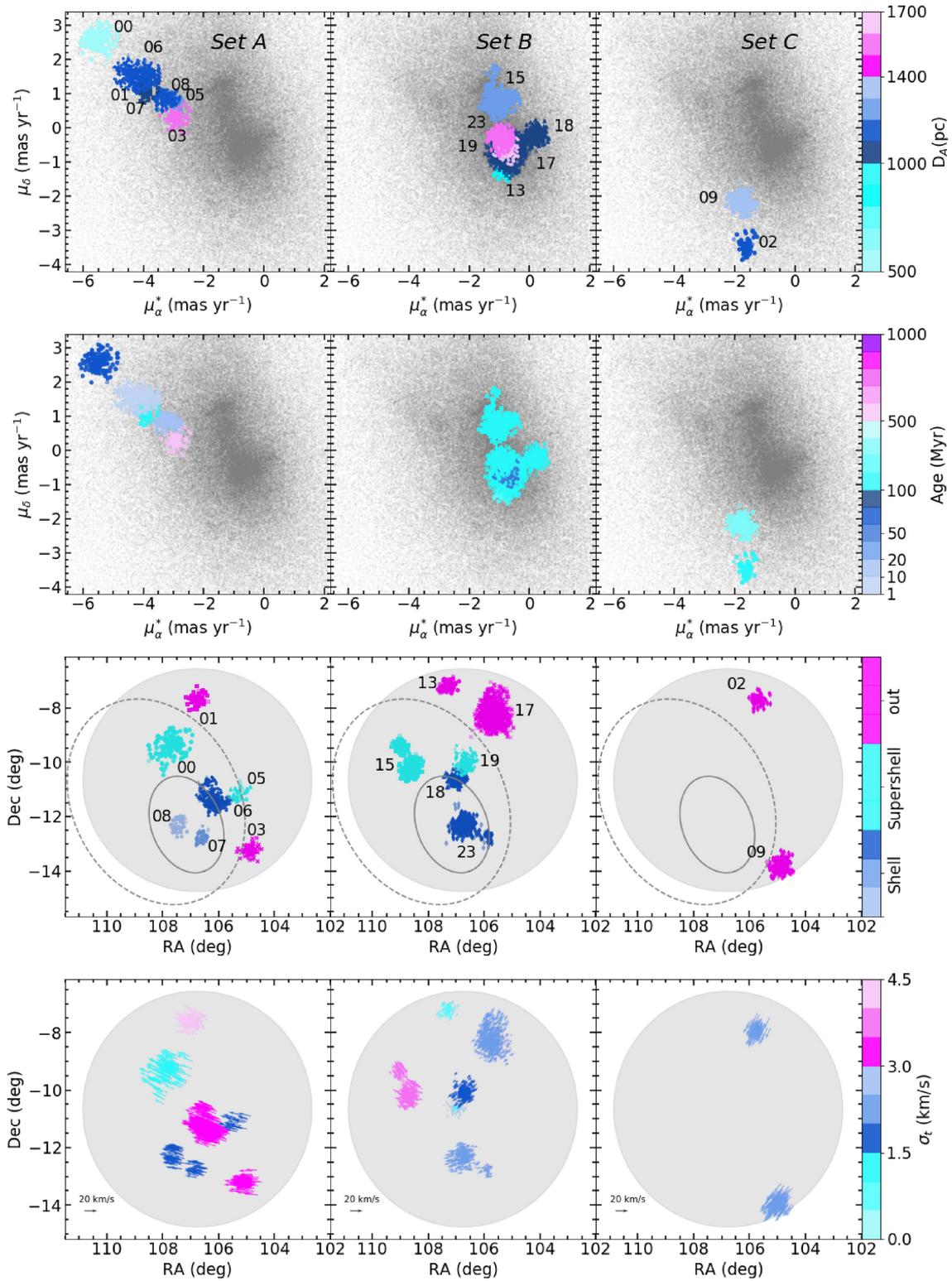


Figure 8. Proper motion and spatial distributions of 15 groups with $D_A < 1700$ pc in each set of groups: Set A (left column), Set B (centre column), and Set C (right column). First row: proper motion highlighted according to the distances of each group. Cyan groups are foreground stars with $500 < D_A$ (pc) < 1000, blue points are at the same distance as CMA OB1, $1000 \leq D_A$ (pc) ≤ 1400, and magenta points are background stars ($1400 < D_A$ (pc) < 1700). Second row: proper-motion distribution of these groups highlighting their ages. Blue groups have ages less than 100 Myr, cyan groups have $100 < \text{Age}$ (Myr) < 500, and magenta groups $500 < \text{Age}$ (Myr) < 1000. Third row: spatial distribution are coloured according to the CMA shell (grey line) and CMA supershell (dashed grey line) position (Fernandes et al. 2019). Groups in the CMA shell are blue, highlighting CMA08, almost in the centre, light-blue, and groups at the edge, dark-blue. Groups inside the CMA supershell are cyan and those outside both structures are magenta. Fourth row: projection of tangential velocities indicating their direction and magnitude according to the tangential velocity dispersion in cyan ($0.0 < \sigma_{V2D}$ (km s⁻¹) < 1.5), blue ($1.5 < \sigma_{V2D}$ (km s⁻¹) < 3.0), and magenta ($3.0 < \sigma_{V2D}$ (km s⁻¹) < 4.5).

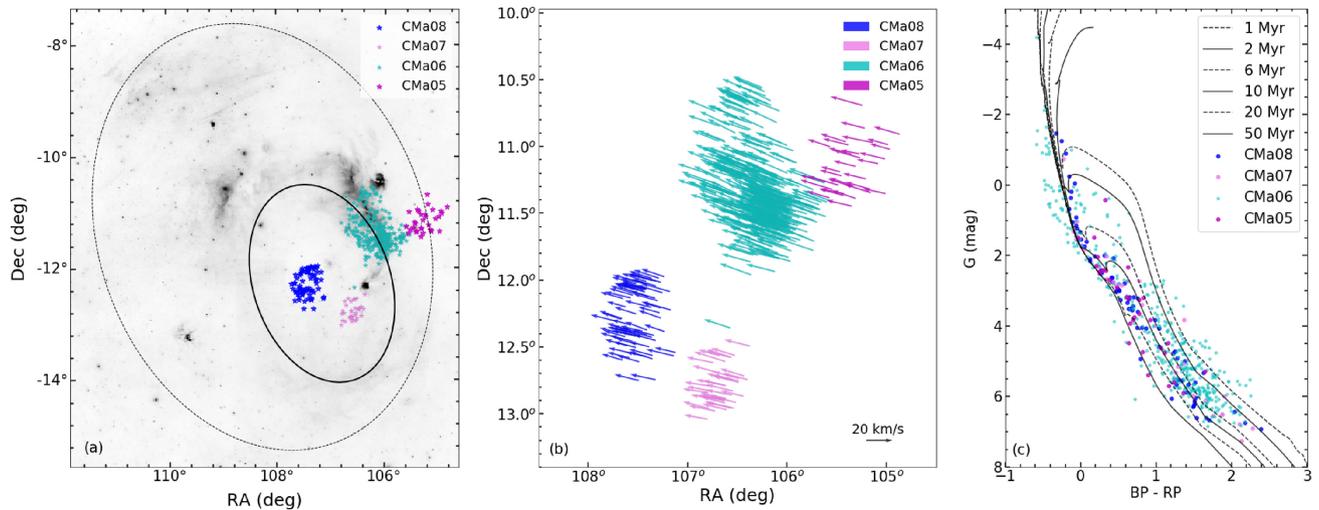


Figure 9. (a) Spatial distribution on Digital Sky Survey image (640 nm); (b) tangential velocity vectors; (c) colour–magnitude diagram of the groups that compose CMa OB1: CMa05 (magenta), CMa06 (cyan), CMa07 (light pink), and CMa08 (blue).

location, CMa23 is the only group for which most stars are inside the CMa shell, while CMa18 is on the edge, following CMa06, with some overlapping stars from CMa19. This group is inside the CMa supershell with CMa15 on the opposite side, while CMa13 and CMa17 are outside, to the north of the CMa OB1 gas structures. The groups in this set have smaller tangential velocities than the other two sets, about 6 km s^{-1} , except for CMa18 which has $V_t < 2 \text{ km s}^{-1}$, which makes it difficult to identify the preferred direction in which it is moving. On the other hand, while CMa15 seems to move in the north-east direction, like groups from Set A, another four groups are moving in the south-east direction. CMa18 and CMa13 are the only ones with 2-D velocity dispersion less than 1 km s^{-1} . CMa17, CMa19, and CMa23 have $\sigma_{V2D} \sim 2 \text{ km s}^{-1}$ and CMa15 has higher σ_{V2D} , about 3.5 km s^{-1} .

The two groups of Set C have similar characteristics. Both are at the same range of distance from the Association and have intermediate age, CMa02 being about 200 Myr younger than CMa09. Moreover, both are outside the CMa supershell, moving in the same direction with tangential velocity about 20 km s^{-1} , as well as groups in Set A, and they have $2.0 < \sigma_{V2D} (\text{km s}^{-1}) < 3$. However, both are quite far from each other: while CMa02 is on the north-west side of the CMa supershell, CMa09 is in the south-west.

6.2 CMa OB1 stellar groups

From Figs 7 and 8, we identified four groups (CMa05, CMa06, CMa07, and CMa08) in Set A with similar characteristics. A summary of the main parameters of these groups is presented in Table 3. These groups are spatially close to each other, at the same distance as the CMa OB1 association, and they seem to be moving together regardless of the other groups. Moreover, they are the youngest groups among all in the line of sight of CMa OB1, with ages less than 20 Myr.

CMa06 is the youngest, with ages 9 and 10 Myr estimated by us using both methods. It is associated with the molecular dense gas present in the star-formation region CMa R1, since this group contains many stars recognized in CMa R1, according to the discussion in Section 3.2.2 and Appendix B. The CMa07, CMa08, and CMa05 groups are older than 10 Myr and appear to be unrelated to any of the three small ($< 10^3 M_{\odot}$) ^{13}CO clouds surveyed by Kim

et al. (2004). Moreover, although CMa08 and CMa07 are spatially on the opposite side to CMa05, in relation to CMa06 they are practically overlapping in proper-motion space (see Fig. 7), having very similar tangential velocities and 2-D velocity dispersion, followed by CMa06 with both parameters slightly larger (see Table 3 and Fig. 9b), suggesting that the internal structure and kinematics of this group are a little different from the others.

On the other hand, in the panel (c) of Fig. 9, it is possible to notice a large spread in the CMa06 CMD, mainly for low-mass stars. Most of the objects can be seen between isochrones from 1–6 Myr, with some of them ranging up to about 20 Myr. Among the stars with $G > 2 \text{ mag}$ there is also a spreading, indicating that some bright stars may have an overestimated visual extinction.

Finally, CMa08 shows a very interesting projected spatial distribution with an almost circular area empty of stars in its inner region. Combined with its proximity to the crossing point predicted for runaway stars (see fig. 5 from Fernandes et al. 2019), this makes us suggest that this group may have been the progenitor cluster that expelled the runaway stars, and also any other O-type star once there were no other massive stars in the region, as well as clearing low-mass stars from its central region. It is possible that the absence of gas and dust there and in the nearby group CMa07 also occurred due to the exhaustion of the parental cloud caused by an older ($> 10 \text{ Myr}$) generation of massive stars, which are no longer seen in this group now.

6.3 CMa OB1 star forming-history

Although the monolithic, or multi-monolithic, scenario (e.g. Lada & Lada 1991; Brown et al. 1997; Kroupa et al. 2001) is not the most acceptable for the formation of OB associations today, all characteristics discussed in Section 6.2 suggest that the four physical groups found within the CMa shell have formed in a small region and are expanding, as are the Per OB1 and Car OB1 associations (Melnik & Dambis 2017). During its expansion process, the most massive stars appear to have been ejected from CMa08 at the same time that its gas was being exhausted, as well as the gas from another two groups (CMa07 and CMa05) around CMa06. The three supernova explosions reported by Fernandes et al. (2019) are feeding back the star formation in CMa06, where gas and dust are

Table 3. Parameters of CMa OB1 contents.

| Group | $N_{50}^{(a)}$ (stars) | RA ^(b) (deg) | Dec. ^(c) (deg) | Age ^M ^(d) (Myr) | Age ^f ^(e) (Myr) | D_A ^(f) (pc) | A_V ^(g) (mag) | $\mu_\alpha \cos \delta$ ^(h) (mas yr ⁻¹) | μ_δ ⁽ⁱ⁾ (mas yr ⁻¹) | ϖ ^(j) (mas) | V_2 ^(k) (km s ⁻¹) | $\sigma_{V2D}^{(l)}$ (km s ⁻¹) |
|-------|---------------------------|----------------------------|------------------------------|--|--|-------------------------------------|--|--|--|--|---|---|
| CMa05 | 28 | 105.32 | -11.20 | 17 ± 27 | 16.0 ^{+0.5} _{-0.5} | 1224 ⁺⁵⁵ ₋₁₃₉ | 1.08 ^{+0.38} _{-0.35} | -2.98 ^{+0.11} _{-0.15} | 0.79 ^{+0.08} _{-0.14} | 0.78 ^{+0.10} _{-0.05} | 18.5 | 1.9 |
| CMa06 | 377 | 106.11 | -11.45 | 10.1 ± 1.0 | 9.0 ^{+0.5} _{-0.5} | 1147 ⁺⁷⁷ ₋₁₃₃ | 1.18 ^{+0.46} _{-1.09} | -4.20 ^{+0.35} _{-0.38} | 1.52 ^{+0.19} _{-0.21} | 0.85 ^{+0.09} _{-0.07} | 24.3 | 3.4 |
| CMa07 | 26 | 106.60 | -12.79 | 13.1 ± 1.5 | 14.0 ^{+0.5} _{-0.5} | 1159 ⁺⁵⁵ ₋₆₉ | 0.70 ^{+0.13} _{-0.19} | -3.36 ^{+0.20} _{-0.08} | 0.79 ^{+0.13} _{-0.13} | 0.84 ^{+0.04} _{-0.06} | 19.2 | 1.6 |
| CMa08 | 64 | 107.43 | -12.30 | 18 ± 5 | 18.0 ^{+1.0} _{-4.0} | 1162 ⁺⁶² ₋₆₂ | 0.97 ^{+0.22} _{-0.26} | -3.13 ^{+0.19} _{-0.17} | 0.87 ^{+0.13} _{-0.13} | 0.84 ^{+0.05} _{-0.06} | 18.2 | 1.6 |

Note. ^(a) Number of stars with $P \geq 50\%$ in the group; ^(b) right ascension (ICRS) at Ep. = 2015.5; ^(c) declination (ICRS) at Ep. = 2015.5; ^(d) age obtained using M20 code (M), see Section C1; ^(e) age obtained using FITCMD (f), see Section C2; ^(f) astrometric distance from Bailer-Jones et al. (2018); ^(g) visual extinction from three-dimensional maps of dust (Green et al. 2019); ^(h) right ascension proper motion; ⁽ⁱ⁾ declination proper motion; ^(j) parallax; ^(k) tangential velocity; ^(l) Velocity dispersion in two dimensions (RA and Dec.).

still abundant, reinforcing the scenario of multiple star formation announced by Santos-Silva et al. (2018). However the widely spaced stars of CMa05 can be interpreted as older low-mass stars (from the first episode of star formation) that are dispersing from CMa06. The CMa07 and CMa08 groups may also go through this process in a few Myr, leaving a bound system and dissolving into field stars, since only 5 per cent of clusters survive their first 20 Myr (Lada & Lada 2003; Pfalzner 2009, 2011) while remaining bound. Furthermore, a small group composed of stars known in the literature, which coincides with the BDSB 96 cluster (not found by HDBSCAN, see Appendix A2) and Sh 2-297 nebula (Cantat-Gaudin & Anders 2020; Mallick et al. 2012), is found between CMa07 and CMa06 and has astrometric characteristics and probably similar age to these two groups, indicating that it is also part of the Association and is perhaps a remnant population from the distancing of these two groups. On the other hand, CMa06 must have a longer survival time, as long as there is star-forming material, but once this is finished, the association reaches dissolution.

This process may have been essential for the maintenance of these physical groups until today. However, a hierarchical scenario in the star-formation history of association CMa OB1 cannot be totally discarded, since several small groups of bright stars, YSOs, H α emitters, and X-ray sources are found, mainly associated with the edges of the CMa shell and CMa supershell from Fernandes et al. (2019), see Appendix B3.

In particular, Sewilo et al. (2019) recently searched YSOs using infrared data, *Herschel* (HI-GAL survey), *Spitzer*, and 2MASS, where they found nearly 300 YSOs in a region dubbed CMa-I224 (see Fig. 1a), at the opposite side of CMa06, following the borders of the CMa shell. They report that the vast majority of these objects are associated with high H $_2$ column density regions and they argue that these stars have formed in filaments and become more dispersed over time, reinforcing a hierarchical star-formation scenario.

These structures could not be found in this work for two main reasons: the first is the detection limit of *Gaia*, for which it is challenging to find embedded stars, and the second is the configuration of HDBSCAN chosen in this work (Section 2.1), which allows only detection of groups with more than 30 stars. It is important to note that most of the objects reported by Sewilo et al. (2019) were detected by infrared surveys and suffer from high extinction, being immersed in gas and dust, making their detection by optical surveys such as *Gaia* difficult. For example, Pettersson & Reipurth (2019) found only 98 stars with useful parallaxes determined by *Gaia*, among all 334 H α star emitters found by them. On the other hand, most YSOs are found in small groups, such as those in Fischer et al. (2016), who found no groups containing more than 25 objects. A brief analysis of the objects studied by these authors, in addition to bright sources from Gregorio-Hetem (2008), YSOs from Fernandes

et al. (2015), and X-rays (Gregorio-Hetem et al. 2009; Santos-Silva et al. 2018), shows that most of these objects are clustered on the edges of the CMa shell but only some of them have good-quality optical data from *Gaia* DR2. However, we found four subgroups containing between six and 14 known stars that have parameters similar to those of our CMa OB1 groups (see Appendix B1), giving us a strong indication that, although these structures are not large bound groups like clusters, they are also part of the Association.

7 SUMMARY AND CONCLUSIONS

In order to find physical groups or star clusters in the CMa OB1 association, we use a clustering algorithm – HDBSCAN (McInnes, Healy & Astels 2017) – in five-dimensional parameter space: $\mu_\alpha \cos \delta$, μ_δ , ϖ , RA, and Dec., to search for different stellar groups in regions as large as stellar associations and to provide their astrometric parameters and characterize their properties. We applied this code to a sample of about 250 000 stars with good astrometric and photometric quality from *Gaia* DR2, with parallaxes between 0.4 and 2.0 mas, covering an area of $R = 4.1$ centred on RA = 106.7° and Dec. = -10.6°, in which we found 29 groups containing from 31 to 1096 stars. We also used the astrometric distances derived by Bailer-Jones et al. (2018) and visual extinction computed from 3D maps to help us characterize the stellar groups identified in our study.

15 groups were found by our method, with distances between 570 and 1650 pc around CMa OB1 (1200 pc). Seven of them have been recognized in the literature as open clusters. CMa06 and CMa15 contain the known clusters VdB 92 and NGC 2353, respectively, and CMa07 is probably part of a new cluster, FOF 2216, identified by Liu & Pang (2019). The other five groups are unpublished in the literature and they were considered by us as open cluster candidates. In particular, CMa06 also contains a vast population of young objects associated with the CMa R1 star-formation region, including 166 objects known in the literature, of which 76 per cent are X-ray sources studied by Santos-Silva et al. (2018), and many of them (~67 per cent) appear to be very likely members as discussed by GH21 estimation of membership probabilities. Moreover, by comparing the members of CMa05 and CMa06 with data from Santos-Silva et al. (2018), we were able to corroborate that 55 X-ray sources confirmed by the authors are in fact members of CMa R1 (CMa06). In addition, 57 objects classified by them as possible members and two of unknown origin can be reclassified as CMa R1 members.

We used two different algorithms developed by Monteiro et al. (2020) and Bonatto (2019), both based on PARSEC isochrone fitting of *Gaia* magnitudes, to determine the ages of the physical groups found here. These fittings also provided distances, visual extinction, and metallicities. The ages determined by the two algorithms were

compatible, between them, for all 15 groups around CMA OB1. The photometric distances and visual extinction also agree well between both estimations and with astrometric distances and mean visual extinction from 3D maps, respectively. Only CMA02 and CMA03 have overestimated photometric A_V obtained by at least one method.

In order to identify group membership of the association, we segregate the 15 groups into three distinct sets of groups, according to their proper-motion distribution. Set A has seven groups (CMA00, CMA01, CMA03, CMA05, CMA06, CMA07, and CMA08). Set B is composed of CMA13, CMA15, CMA17, CMA18, CMA19, and CMA23, and Set C has two groups, CMA02 and CMA09.

The younger groups in Set A (CMA05, CMA06, CMA07, and CMA08) are considered by us to be associated with CMA OB1, mainly because they are located within the CMA shell reported by Fernandes et al. (2019). Moreover, CMA06 (~ 10 Myr) contains the youngest stellar population associated with CMA R1 (Santos-Silva et al. 2018, GH21). CMA07 has a smaller (26 stars) and more intermediate-aged population (~ 14 Myr) in the south of CMA OB1. CMA05 (28 stars) and CMA08 (64 stars) are aged about 18 Myr and are on the west and east sides of CMA R1, respectively.

Astrometric and photometric analysis of these four physical stellar groups in the association CMA OB1 helped us to clarify their star-formation history better. These relatively close physical groups within the 60-pc shell are consistent with the monolithic model of association formation. However, it does not explain the embedded stellar content associated with gas and dust filaments found by other authors (see Fernandes et al. 2015; Fischer et al. 2016; Pettersson & Reipurth 2019; Sewilo et al. 2019), mainly at the edges of the CMA shell on the opposite side of the CMA R1 star formation, where there are no larger stellar physical groups, probably due to our selection criteria, which are constrained to visible stars.

We suggest that these groups were all born together from a smaller space in the centre of the CMA OB1 association, like the Per OB1 and Car OB1 associations (Melnik & Dambis 2017), and are expanding. A first generation of stars older than 10 Myr gave rise to CMA06, CMA07, and CMA08. The morphology of CMA08 suggests that it is probably the progenitor cluster of massive runaway stars expelled during the following episodes of star formation (~ 6 , ~ 2 , and ~ 1 Myr) in CMA OB1 (Fernandes et al. 2019), phenomena also responsible for expelling low-mass stars from the CMA08 centre. CMA06 is immersed in a dense molecular cloud in which both first-generation stars (> 10 Myr) and the young population (< 5 Myr) born from these recent episodes coexist, according to the findings of Santos-Silva et al. (2018) and Gregorio-Hetem et al. (2009) based on X-ray data. On the other hand, both CMA07 and CMA08 are not having their populations renewed, as most of their gas and dust have already been exhausted during the first star formation, and they may lose their members in the coming Myr, since 95 percent of clusters are expected not to survive their first 20 Myr (Lada & Lada 2003; Pfalzner 2009, 2011), as this is happening on the west edge of CMA06, giving rise to CMA05, which is probably composed of older low-mass stars leaving the larger group. CMA R1 must still form stars for as long as its interstellar material lasts.

In the near future, we intend to find different characteristics in the existing population in the CMA OB1 association, such as mass segregation or multiple ages, from characterization of new cluster candidates and those already studied, using multiband photometry from the S-PLUS collaboration. On the other hand, astrometric data from the next *Gaia* data releases combined with radial velocities will be essential to confirm the expansion of younger groups in CMA OB1, as well as to unravel the mystery of the older population (> 100 Myr) found in the same region as CMA OB1.

ACKNOWLEDGEMENTS

We thank our anonymous referee for useful and constructive comments and suggestions that improved our work. TSS acknowledges FAPESP proc. 2018/06822-6, HDP acknowledges FAPESP proc. 2018/21250-9, FAF acknowledges FAPESP proc. 2018/20977-2, JGH acknowledges FAPESP Proc. 2014/18100-4, and VJP acknowledges FAPESP Proc. 2015/24946-6. We thank Friedrich Anders, R. Herpich, Luis A. G. Soto, and Ángeles Pérez-Villegasábio for their suggestions and comments on the manuscript. This research was performed in part using the facilities of the Laboratório de Astrofísica Computacional da Universidade Federal de Itajubá (LAC-UNIFED). This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC: <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the *Gaia* Multilateral Agreement.

DATA AVAILABILITY

Most of the underlying data in this article are available in the main part of the work. The tables containing individual star information (IDs and astrometric parameters from *Gaia* and probability membership estimated by us) of the 29 groups that we found using HDBSCAN are available in machine-readable form at the CDS and in the online supplementary material, as well as their analysis figures, in PDF format, as presented in Appendix D.

REFERENCES

- Arenou F. et al., 2018, *A&A*, 616, A17
 Bailer-Jones C. A. L., Rybizki J., Foesneau M., Mantelet G., Andrae R., 2018, *AJ*, 156, 58
 Baumgardt H., Kroupa P., 2007, *MNRAS*, 380, 1589
 Bica E., Pavani D. B., Bonatto C. J., Lima E. F., 2019, *AJ*, 157, 12
 Bonatto C., 2019, *MNRAS*, 483, 2758
 Bonatto C., Bica E., 2010, *A&A*, 516, A81
 Bonatto C., Chies-Santos A. L., 2020, *MNRAS*, 493, 2688
 Bossini D. et al., 2019, *A&A*, 623, A108
 Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127
 Brown A. G. A., 2001, in Aguilar A., Carramiñana A., eds. *Rev. Mex. Astron. Astrofis. Conf. Ser.*, Mexico, 11, p. 89
 Brown A. G. A., Dekker G., de Zeeuw P. T., 1997, *MNRAS*, 285, 479
 Campello R. J. G. B., Moulavi D., Sander J., 2013, in Pei J., Tseng V. S., Cao L., Motoda H., Xu G., eds. *Advances in Knowledge Discovery and Data Mining*. Springer, Berlin, Heidelberg, p. 160
 Campello R. J. G. B., Moulavi D., Zimek A., Sander J., 2015, *ACM Trans. Knowl. Discov. Data*, 10, 1
 Cantat-Gaudin T., Anders F., 2020, *A&A*, 633, A99
 Cantat-Gaudin T. et al., 2018, *A&A*, 618, A93
 Cantat-Gaudin T., Mapelli M., Balaguer-Núñez L., Jordi C., Sacco G., Vallenari A., 2019a, *A&A*, 621, A115
 Cantat-Gaudin T. et al., 2019b, *A&A*, 626, A17
 Cantat-Gaudin T. et al., 2020, *A&A*, 640, A1
 Capitanio L., Lallement R., Vergely J. L., Elyajouri M., Monreal-Ibero A., 2017, *A&A*, 606, A65
 Castro-Ginard A., Jordi C., Luri X., Julbe F., Morvan M., Balaguer-Núñez L., Cantat-Gaudin T., 2018, *A&A*, 618, A59
 Castro-Ginard A., Jordi C., Luri X., Cantat-Gaudin T., Balaguer-Núñez L., 2019, *A&A*, 627, A35
 Castro-Ginard A. et al., 2020, *A&A*, 635, A45
 Claria J. J., 1974a, *AJ*, 79, 1022
 Clariá J. J., 1974b, *A&A*, 37, 229

- Claria J. J., Piatti A. E., Lapasset E., 1998, *A&AS*, 128, 131
- Dale J. E., Ercolano B., Bonnell I. A., 2015, *MNRAS*, 451, 987
- Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, *A&A*, 389, 871
- Dias W. S., Monteiro H., Lépine J. R. D., Prates R., Gneiding C. D., Sacchi M., 2018, *MNRAS*, 481, 3887
- Dias W. S., Monteiro H., Moitinho A., Lépine J. R. D., Carraro G., Paunzen E., Alessi B., Vilella L., 2021, *MNRAS*, 504, 356
- Dobashi K., Uehara H., Kandori R., Sakurai T., Kaiden M., Umemoto T., Sato F., 2005, *PASJ*, 57, S1
- Donor J. et al., 2020, *AJ*, 159, 199
- Ester M., Kriegel H.-P., Sander J., Xu X., 1996, in Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, California, p. 226
- Fernandes B., Gregorio-Hetem J., Montmerle T., Rojas G., 2015, *MNRAS*, 448, 119
- Fernandes B., Montmerle T., Santos-Silva T., Gregorio-Hetem J., 2019, *A&A*, 628, A44
- Fischer W. J., Padgett D. L., Stapelfeldt K. L., Sewilo M., 2016, *ApJ*, 827, 96
- Gagné J. et al., 2018, *ApJ*, 856, 23
- Gaia Collaboration, 2018, *A&A*, 616, A1
- Galli P. A. B., Bouy H., Olivares J., Miret-Roig N., Sarro L. M., Barrado D., Berihuete A., Brandner W., 2020, *A&A*, 634, A98
- Gao X., 2018, *AJ*, 156, 121
- Ginsburg A. et al., 2016, *A&A*, 595, A27
- Girichidis P., Federrath C., Banerjee R., Klessen R. S., 2012, *MNRAS*, 420, 613
- Goffe W. L., Ferrier G. D., Rogers J., 1994, *J. Econometrics*, 60, 65
- Goodwin S. P., Bastian N., 2006, *MNRAS*, 373, 752
- Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, *ApJ*, 887, 93
- Gregorio-Hetem J., 2008, in Reipurth B., ed., ASP Conf. Ser. Vol. 5, Handbook of Star Forming Regions, Volume II: The Southern Sky. Astron. Soc. Pac., San Francisco, p. 1
- Gregorio-Hetem J., Montmerle T., Rodrigues C. V., Marciotto E., Preibisch T., Zinnecker H., 2009, *A&A*, 506, 711
- Gregorio-Hetem J., Lefloch B., Hetem A., Montmerle T., Fernandes B., Mendoza E. F., De Simone M., 2021a, preprint ([arXiv:2108.08192](https://arxiv.org/abs/2108.08192))
- Gregorio-Hetem J. et al., 2021b, *AJ*, 161, 133 (GH21)
- He Z.-H., Xu Y., Hao C.-J., Wu Z.-Y., Li J.-J., 2021, *Res. Astron. Astrophys.*, 21, 093
- Herbst W., Assousa G. E., 1977, *ApJ*, 217, 473
- Herbst W., Racine R., Warner J. W., 1978, *ApJ*, 223, 471
- Hetem A., Gregorio-Hetem J., 2019, *MNRAS*, 490, 2521
- Hills J. G., 1980, *ApJ*, 235, 986
- Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R. D., 2016, *A&A*, 585, A101
- Kim B. G., Kawamura A., Yonekura Y., Fukui Y., 2004, *PASJ*, 56, 313
- Kounkel M., Covey K., 2019, *AJ*, 158, 122
- Kounkel M., Covey K., Stassun K. G., 2020, *AJ*, 160, 279
- Kroupa P., Aarseth S., Hurley J., 2001, *MNRAS*, 321, 699
- Kruijssen J. M. D., Maschberger T., Moeckel N., Clarke C. J., Bastian N., Bonnell I. A., 2012, *MNRAS*, 419, 841
- Kuhn M. A., de Souza R. S., Krone-Martins A., Castro-Ginard A., Ishida E. O., Povich M. S., Hillenbrand L. A., 2021, *ApJS*, 254, 33
- Lada C. J., Lada E. A., 1991, in Janes K., ed., ASP Conf. Ser. Vol. 13, The Formation and Evolution of Star Clusters. Astron. Soc. Pac., San Francisco, California, p. 3
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57
- Limberg G. et al., 2021, *ApJ*, 907, 10
- Lindegren L. et al., 2018, *A&A*, 616, A2
- Liu L., Pang X., 2019, *ApJS*, 245, 32
- Liu J., Fang M., Liu C., 2020, *AJ*, 159, 105
- Lodieu N., Pérez-Garrido A., Smart R. L., Silvotti R., 2019, *A&A*, 628, A66
- Logan C. H. A., Fotopoulou S., 2020, *A&A*, 633, A154
- Longmore S. N. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. University of Arizona Press, Tucson, Arizona, p. 291
- MacQueen R. M., 1967, *AJ*, 72, 814
- Maíz Apellániz J., Weiler M., 2018, *A&A*, 619, A180
- Mallick K. K., Ojha D. K., Samal M. R., Pandey A. K., Bhatt B. C., Ghosh S. K., Dewangan L. K., Tamura M., 2012, *ApJ*, 759, 48
- Marigo P. et al., 2017, *ApJ*, 835, 77
- McInnes L., Healy J., Astels S., 2017, *JOSS*, 2, 205
- Melnik A. M., Dambis A. K., 2017, *MNRAS*, 472, 3887
- Melnik A. M., Dambis A. K., 2020, *MNRAS*, 493, 2339
- Mendes de Oliveira C. et al., 2019, *MNRAS*, 489, 241
- Monteiro H., Dias W. S., Moitinho A., Cantat-Gaudin T., Lépine J. R. D., Carraro G., Paunzen E., 2020, *MNRAS*, 499, 1874
- Pedregosa F. et al., 2011, *J. Mach. Learn. Research*, 12, 2825
- Pettersson B., Reipurth B., 2019, *A&A*, 630, A90
- Pfalzner S., 2009, *A&A*, 498, L37
- Pfalzner S., 2011, *A&A*, 536, A90
- Piatti A. E., Bonatto C., 2019, *MNRAS*, 490, 2414
- Rebull L. M. et al., 2013, *AJ*, 145, 15
- Reynolds R. J., Ogden P. M., 1978, *ApJ*, 224, 94
- Santos-Silva T., Gregorio-Hetem J., Montmerle T., Fernandes B., Stelzer B., 2018, *A&A*, 609, A127
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Sewilo M. et al., 2019, *ApJS*, 240, 26
- Sharpless S., 1959, *ApJS*, 4, 257
- Shevchenko V. S., Ezhkova O. V., Ibrahimov M. A., van den Ancker M. E., Tjin A. Djie H. R. E., 1999, *MNRAS*, 310, 210
- Soares J. B., Bica E., 2002, *A&A*, 388, 172
- Soares J. B., Bica E., 2003, *Bulletin of the Astronomical Society of Brazil*, 23, 192
- Sugitani K., Fukui Y., Ogura K., 1991, *ApJS*, 77, 59
- Ward J. L., Kruijssen J. M. D., Rix H.-W., 2020, *MNRAS*, 495, 663
- Wright N. J., Mamajek E. E., 2018, *MNRAS*, 476, 381
- Yalyalieva L., Carraro G., Vazquez R., Rizzo L., Glushkova E., Costa E., 2020, *MNRAS*, 495, 1349
- Zari E., Brown A. G. A., de Zeeuw P. T., 2019, *A&A*, 628, A123
- Zucker C., Speagle J. S., Schlafly E. F., Green G. M., Finkbeiner D. P., Goodman A. A., Alves J., 2019, *ApJ*, 879, 125
- Zucker C., Speagle J. S., Schlafly E. F., Green G. M., Finkbeiner D. P., Goodman A., Alves J., 2020, *A&A*, 633, A51

SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://academic.oup.com/mnras/article/508/1/1033/6379568) online.

supplementary_material_online.CMa_paper.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: HDBSCAN CAVEATS

It can be noted in Fig. 2(a) that groups CMa04, CMa22, CMa24, CMa27, and CMa28 are close to the edge of the sample on the celestial sphere, and in Fig. 2(c) several groups seem to be at the lower limit of parallax adopted in this work ($\varpi > 0.4$ mas). In order to avoid bias introduced by these constraints in the inferred characteristics of 14 unreliable groups, they were excluded from our analysis (see Section 4).

This Appendix is dedicated to summarizing the characteristics of some of the discarded groups that deserve to be studied in a forthcoming work (Section A1). In Section A2 we address the issue of some previously known stellar clusters that were missed, probably due to the conservative criteria adopted in our methodology.

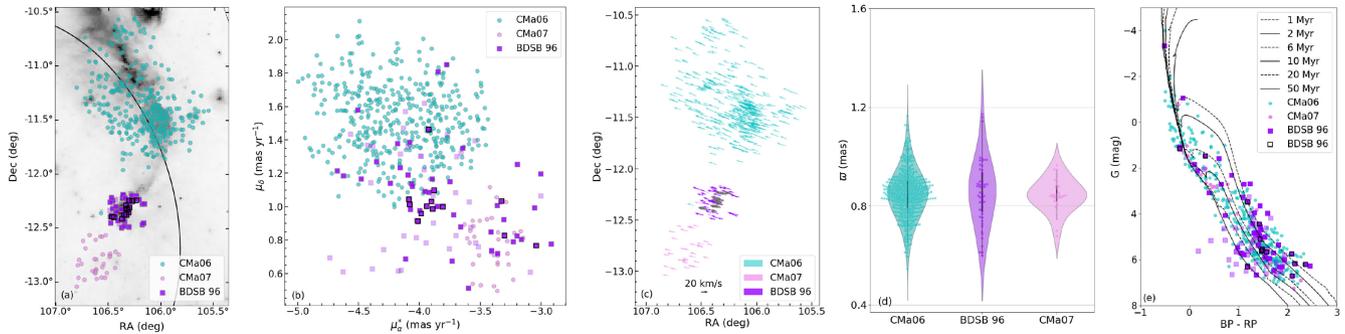


Figure A1. (a) Spatial distribution on Digital Sky Survey image (640 nm) of CMA06 (cyan) and CMA07 (light pink) compared with BDSB 96 cluster (purple squares: Cantat-Gaudin & Anders 2020). Its light points present members with $P < 50$ per cent according to these authors and black contours highlights CMA06 (see Appendix B3). A black line shows the CMA shell edge from Fernandes et al. (2019). (b) Proper-motion distribution; (c) tangential velocity vectors; (d) violin histograms; and (e) colour–magnitude diagram of both groups and BDSB 96.

Table A1. Astrometric median values of *Gaia* five-dimensional parameters, distance and visual extinction of seven distant possible groups not well determined by HDBSCAN in the CMa OB1 region.

| Group | N ^(a) (stars) | N_{50} ^(b) (stars) | P_{N50} ^(c) (%) | RA ^(d) (deg) | Dec. ^(e) (deg) | $\mu_{\alpha} \cos \delta$ ^(f) (mas yr ⁻¹) | μ_{δ} ^(g) (mas yr ⁻¹) | ϖ ^(h) (mas) | D_A ⁽ⁱ⁾ (pc) | A_V ^(j) (mag) |
|---------|-------------------------------|------------------------------------|---------------------------------|----------------------------------|----------------------------------|--|--|----------------------------------|------------------------------|--------------------------------|
| CMA04 | 132 | 123 | 93 | 102.98 ^{+0.25} -0.29 | -9.96 ^{+0.17} -0.41 | -1.25 ^{+0.12} -0.21 | 1.26 ^{+0.17} -0.16 | 0.50 ^{+0.06} -0.09 | 1921 ⁺²⁵² -233 | 0.96 ^{+0.14} -0.23 |
| CMA10* | 37 | 35 | 95 | 105.40 ^{+0.06} -0.12 | -13.56 ^{+0.08} -0.18 | -0.99 ^{+0.06} -0.09 | -1.43 ^{+0.08} -0.08 | 0.47 ^{+0.05} -0.05 | 2050 ⁺²⁰⁰ -185 | 0.84 ^{+0.06} -0.18 |
| CMA11 | 31 | 28 | 90 | 106.53 ^{+0.07} -0.11 | -7.41 ^{+0.11} -0.17 | -2.31 ^{+0.12} -0.08 | 0.02 ^{+0.09} -0.14 | 0.48 ^{+0.04} -0.08 | 1976 ⁺²¹² -164 | 0.91 ^{+0.18} -0.30 |
| CMA20 | 258 | 229 | 89 | 106.56 ^{+0.70} -0.51 | -9.37 ^{+0.26} -0.23 | -1.26 ^{+0.19} -0.15 | 0.78 ^{+0.19} -0.16 | 0.46 ^{+0.04} -0.09 | 2063 ⁺²⁵⁸ -183 | 1.15 ^{+0.27} -0.48 |
| CMA25** | 96 | 88 | 92 | 107.06 ^{+0.12} -0.12 | -13.22 ^{+0.17} -0.12 | -1.42 ^{+0.30} -0.17 | 1.26 ^{+0.18} -0.14 | 0.46 ^{+0.04} -0.06 | 2091 ⁺¹⁹⁴ -173 | 1.18 ^{+0.20} -0.32 |
| CMA26 | 70 | 61 | 87 | 106.45 ^{+0.08} -0.16 | -12.27 ^{+0.17} -0.16 | -1.59 ^{+0.12} -0.13 | 1.17 ^{+0.12} -0.10 | 0.47 ^{+0.05} -0.07 | 2030 ⁺²⁴¹ -206 | 0.85 ^{+0.08} -0.16 |

Note. ^(a) Number of stars in the group; ^(b) number of stars with membership probability $P \geq 50\%$; ^(c) percentage of stars with membership probability $P \geq 50\%$; ^(d) right ascension (ICRS) at Ep. = 2015.5; ^(e) declination (ICRS) at Ep. = 2015.5; ^(f) right ascension proper motion; ^(g) declination proper motion; ^(h) parallax; ⁽ⁱ⁾ astrometric distance from Bailer-Jones et al. (2018); ^(j) visual extinction from three-dimensional maps of dust (Green et al. 2019). Probable clusters: *Ruprecht 8 and **NGC 2345.

A1 Unreliable groups detected by HDBSCAN

Checking the diagnostic figures (See Appendix D and online supplementary material) we realized that HDBSCAN may fail to characterize groups that have parameters near to the edges of our selection criteria (see Section 2.1). For instance, CMA04 has more than 100 stars and seems to be part of a population that could not be fully identified, since some members may be outside the spatial distribution of the sample. Despite the lack of detected objects, this group has more than 85 per cent of the population showing $P \geq 50$ per cent. This high membership probability for a large fraction of the objects in CMA04, as well as for CMA10, CMA11, CMA20, CMA25, and CMA26, indicates that they are probably physical groups that deserve to be better studied, considering a different range of parameters.

The median values of astrometric parameters from *Gaia* DR2, distance from Bailer-Jones et al. (2018), visual extinction from 3D maps, and number of members (total and $P \geq 50$ per cent) of these six possible groups are presented in Table A1. However, they cannot be considered the actual parameters of these groups, but only a first guess to help search for the entire group.

Comparing these parameters with previously known clusters (see Section 3.2.1), we found that Ruprecht 8 ($\varpi = 0.44$ mas, Cantat-Gaudin et al. 2020) and NGC 2345 ($\varpi = 0.35$ mas) can be related to CMA10 and CMA25, respectively, once our groups have positions and proper motions similar to these clusters. These coinciding results reinforce our hypothesis that these are actual groups. The other candidates, CMA04, CMA11, and CMA26 are probably new stellar

clusters, while CMA20 is suggested to be a large moving group, due to its extensive spatial dispersion.

A2 Clusters not found by HDBSCAN

As discussed in Section 3.2.1, there are 30 clusters in the literature suggested to be related to the CMa OB1 region. However, HDBSCAN did not find physical groups associated with 19 of these previously known clusters. In fact, most of them (74 per cent)¹² are not confirmed as clusters by any of the recent catalogues based on *Gaia* data (e.g. Liu & Pang 2019; Cantat-Gaudin & Anders 2020).

Our conclusion is that only five previously known clusters were missed by HDBSCAN: FOF 2304 (Liu & Pang 2019) and BDSB 93, BDSB 96, FSR 1183, and UPK 452 (Cantat-Gaudin & Anders 2020). It is possible the some of these clusters were not detected due to our choice of the minimum number of objects constraining identification as a group (`min_cluster` = 30). Cantat-Gaudin & Anders (2020) list 28 and 36 members for UPK 452 and BDSB 93, respectively. However, only 25 objects in each cluster would be selected according to our criteria (Section 2.1). On the other hand, FSR 1183 has 42 members following these criteria, but several of them have low

¹²Collinder 466, FSR 1158, FSR 1159, FSR 1160, FSR 1164, FSR 1169, FSR 1178, FSR 1194, FSR 1199, FSR 1200, FSR 1202, FSR 1204, NGC 2349, and NGC 2351.

membership probability,¹³ leaving fewer than 30 candidates to be selected and identified as a group by the code. Still, its parallax ($\varpi = 0.428$ mas) is within the limit adopted by us, which is an additional reason that HDBSCAN has missed this cluster. Finally, despite FOF 2304 having 50 members, no data are available for comparison. It is also located on the edge of our sample spatial distribution, so that seemingly fewer than 30 of its stars coincide with the area studied by us.

Another way to ensure the detection of these groups by the code is to decrease the values of `min_cluster`; however, groups with fewer than 30 objects are difficult to characterize because they are scarce and less cohesive. These kinds of groups deserve more detailed studies that are not the purpose of this work. Moreover, we are looking for stellar groups at ~ 1200 pc, related to the CMa OB1 association, which does not seem to be the case for FOF 2304, BDSB 93, FSR 1183, and UPK 452.

BDSB 96, present in four catalogues discussed in Section 3.2.1, is the only important cluster for our analyses, that should be detected by HDBSCAN but was not. According to Cantat-Gaudin & Anders (2020) and Dias et al. (2021), this cluster contains 91 objects, but only 52 of them follow our selection criteria. An intriguing issue is that BDSB 96 ($d \sim 1145$ pc) shows a spatial distribution between our groups CMa06 and CMa07 and has characteristics similar to both. This led us to question why this cluster was not found by the code. Could it be an important part of the CMa OB1 star-formation scenario that is not being taken into account?

For this reason, we performed a comparison between data of BDSB 96 members with $P \geq 50$ per cent (Cantat-Gaudin & Anders 2020) and our groups CMa06 and CMa07. Although spatially BDSB 96 looks like a physical group, the members have an elongated spread distribution of proper motion, partially covering the parameters of both CMa06 and CMa07. Fewer than 20 objects are concentrated around $\mu_\alpha \cos \delta \sim -4$ mas yr⁻¹ and $\mu_\delta \sim 1$ mas yr⁻¹, meaning that it is not as cohesive as the groups found by HDBSCAN (see Fig. A1b). The tangential velocity vectors (see Fig. A1c) show that the members of BDSB 96 are moving in the same direction and have similar tangential velocity to the groups CMa06 and CMa07. The CMD shown in Fig. A1(e) was obtained using the same visual extinction adopted for our groups. The ages distribution indicates that the BDSB 96 population is also similar to CMa06 and CMa07. Moreover, about 13 members of BDSB 96 are known in the literature as bright stars, H α emitters, or YSOs (see discussion at Section B1). Six of these objects are also related to the H II region Sh2-297 (Mallick et al. 2012).

Even relaxing the minimal number of members (e.g. `min_cluster` = 10) to be searched with HDBSCAN, which would allow identification of less cohesive physical groups, BDSB 96 is not detected. This fact leads us to suggest it is not formed by a single physical group but may be the remaining population from the rupture of two groups during the expansion process of the entire CMa OB1 association. Besides being associated with this complex region, BDSB 96 may have a particular and located star-formation history that could be supported by interaction between the ionizing star HD 53623 (spectral type B0 V, see Fig. 1(c)), and the H II region Sh2-297, for which a dynamical age of 1.07 Myr was estimated by Mallick et al. (2012). However, to confirm this hypothesis, a more

detailed multiwavelength study is needed, including the embedded objects reported by Mallick et al. (2012) and data from *Gaia* EDR3 and S-PLUS obtained for the region between CMa06 and CMa07.

APPENDIX B: PREVIOUSLY KNOWN STELLAR CONTENT

Still trying to understand the stellar population of CMa OB1 better, we performed a search in the literature for objects identified in association with this region (see Fig. B1). As mentioned in Section 1 and Section 6.3, part of CMa OB1, mainly its famous star-formation region CMa R1, has been widely studied at different wavelengths. In this Appendix we describe the comparison of members of our groups with published catalogues related to CMa OB1.

B1 Known stellar group membership

As compiled by Gregorio-Hetem (2008), there are 114 bright stars in the CMa region, most of them having spectral type B, including objects from catalogues published by Claria (1974a) and Shevchenko et al. (1999). About 334 H α emitters were detected by Pettersson & Reipurth (2019) using the ESO 1-m Schmidt telescope at the La Silla Observatory, covering a field of $\sim 5^\circ \times 5^\circ$ on the sky in CMa OB1. Gregorio-Hetem et al. (2009) found 98 X-ray sources using *ROSAT* observations and Santos-Silva et al. (2018) detected most of these sources and almost 300 new ones using sensitive observations from the *XMM-Newton* satellite, which gives a total of 387 X-ray sources found in CMa R1. Additionally, Fernandes et al. (2015) characterized 56 young stars in CMa R1, using multi-object spectroscopy performed by the Gemini South telescope, based on optical spectral features, mainly H α and lithium lines. The authors confirmed 41 T Tauri stars: seven classical T Tauri (CTT), 34 weak T Tauri (WTT), and 15 very likely pre-main-sequence stars. Fischer et al. (2016) also searched for young stellar objects (YSOs) using the *Wide-field Infrared Survey Explorer* (*WISE*). They found 144 Class I and 335 Class II candidates in a FOV of about $10^\circ \times 10^\circ$, covering the entire region of CMa OB1. Finally, Sewilo et al. (2019) identified 293 YSO candidates on the opposite side of CMa06 ($\sim l = 224^\circ$) in relation to the CMa shell (magenta rectangle in Fig. 1a), using data from far-, mid- and near-infrared surveys from the *Herschel Space Observatory*, *Spitzer Space Telescope*, 2MASS and *WISE*. The authors characterized 210 of these YSOs and classified them into three groups: 16 objects with ‘envelope only’; 21 with ‘envelope and disc’, and 173 ‘disc only’.

In Fig. B1 we present all these objects, highlighting with filled grey symbols the stars with *Gaia* counterparts that satisfy our selection criteria (Section 2.1), and we use black symbols to represent objects found in our groups. We consider four sets of objects based on their characteristics in each survey: bright stars (Gregorio-Hetem 2008), H α emitters (Pettersson & Reipurth 2019), X-ray sources (Gregorio-Hetem et al. 2009; Santos-Silva et al. 2018), and YSOs (Fernandes et al. 2015; Fischer et al. 2016; Sewilo et al. 2019).

The first step in identifying known stars in association with our physical groups was to perform a cross-matching with objects detected in the surveys mentioned above. We found 26 bright stars, 34 H α emitters, 126 X-ray sources, and 41 YSOs in our physical groups. Taking into account the overlap among the catalogues that occurs for some of the objects, there are in our groups a total of 166 objects already known in the literature; 29 per cent (48/166) of them appear in two or more of the inspected catalogues. Among the YSOs that were previously classified, we found one Class I, 17 Class II (CTT), 18 Class III (WTT), and five very likely pre-main-sequence

¹³ Among the 42 stars in the FSR 1183 cluster following our selection criteria, 10 have probability membership $P \leq 10$ per cent, 9 have 10 per cent $< P < 50$ per cent and only 23 have $P \geq 50$ per cent according to Cantat-Gaudin & Anders (2020).

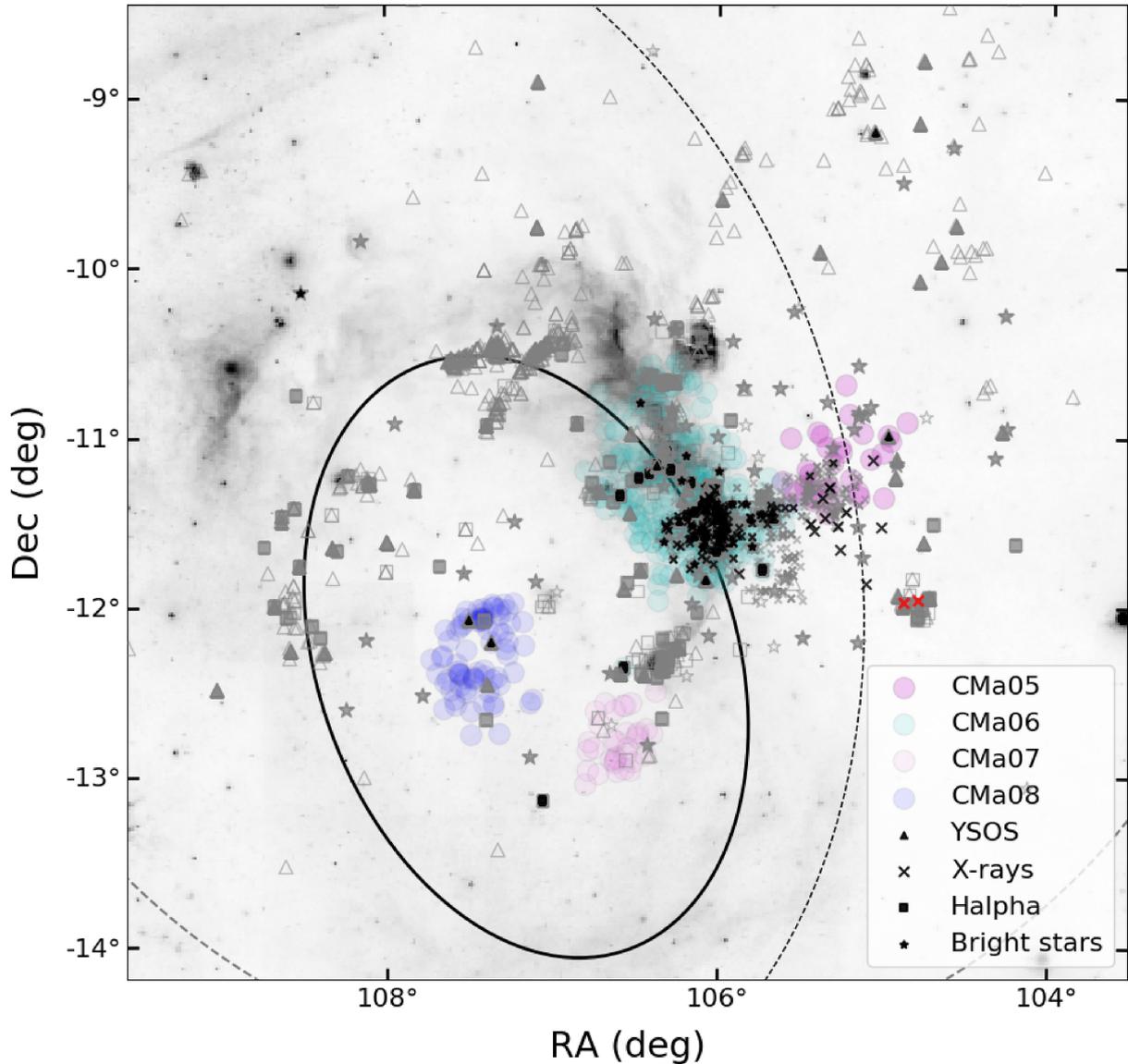


Figure B1. Spatial distribution of four groups of CMa OB1 and stars known in the literature on a Digital Sky Survey image (640 nm) of CMa OB1: bright stars from Gregorio-Hetem (2008) (stars); $H\alpha$ emitters from Pettersson & Reipurth (2019) (squares); X-ray sources from Gregorio-Hetem et al. (2009) and Santos-Silva et al. (2018) (X); and YSOs from Fernandes et al. (2015) and (Fischer et al. 2016) (triangles). Grey open and filled symbols represent objects without and with *Gaia* counterparts, respectively, while black symbols highlight objects known in the literature to be present in our groups. Red ‘X’ symbols are X-ray sources CMaX06 and CMaX07 from Gregorio-Hetem et al. (2009).

stars. Most of the known population is associated with CMa06. Only 10 objects are found in the other groups: CMa05 contains four known stars, CMa08 has two stars, while CMa 15, CMa25, CMa 14, and CMa20 have only one known object each.

As can be seen in Fig. B1, there are a large number of objects that were previously identified in the direction of the CMa OB1 region, mainly YSOs and $H\alpha$ emitters. However, most of them do not have *Gaia* data or they are outside the astrometric and kinematic criteria adopted by us to identify reliable members of the Association.

B2 X-ray sources

X-ray sources correspond to 76 per cent (126/166) of the objects from the literature that coincide with our groups; for this reason, we dedicate this subsection to confirming the efficiency of our method

in finding objects previously known and to discussing the nature of some of these objects.

Our analysis started with a sample of 387 X-ray sources studied by Santos-Silva et al. (2018) in CMa R1, of which 118 were suggested to be very likely members, 185 were possible members, and 84 remained without classification. The spatial distribution of these sources is constrained to the same region as two groups found by us: CMa05 and CMa06 (see Fig. B2).

We performed a cross-match of the X-ray sources with *Gaia* objects that fulfil our selection criteria (see Section 2.1).¹⁴ We apply

¹⁴Among 387 X-ray sources from Santos-Silva et al. (2018), 222 have reliable optical counterparts, i.e. stars from *Gaia* DR2 following our selection criteria (see Section 2.1); 157 have counterparts outside these criteria; and eight have no *Gaia* counterpart.

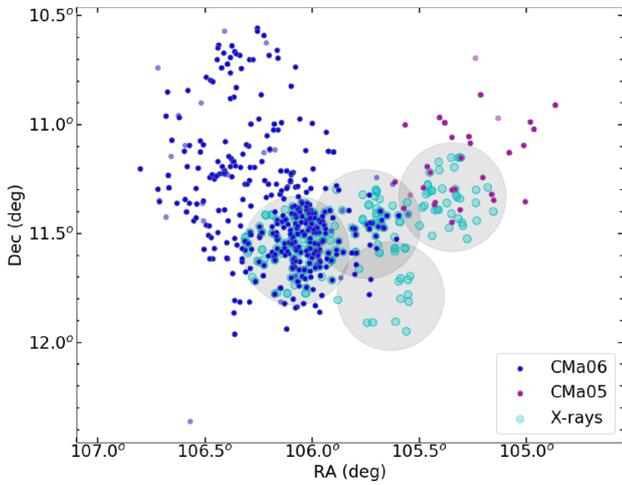


Figure B2. Spatial distributions of CMA06 and CMA05 (blue and magenta circles, respectively) compared with X-ray sources (cyan circles) from Santos-Silva et al. (2018). Group members with $P < 50$ per cent are light points. The hatched area represents the fields observed by the *XMM-Newton* satellite.

the same methodology used by Santos-Silva et al. (2018) to search for infrared counterparts and we find 270 stars from *Gaia* DR2 associated with 222 X-ray sources. Note that some X-ray sources have more than one optical counterpart.

Three *Gaia* counterparts are in CMA05 and 123 are in CMA06 (nine X-ray sources from this group have two optical counterparts); see the grey ‘X’ in Fig. B1 and cyan circles in Fig. B2. Another 144 *Gaia* counterparts related to 105 X-ray sources were not found to be associated with any of our groups. There are different reasons that these sources have not been selected with HDBSCAN: 42 have parallax incompatible with CMA05 or CMA06; 77 do not have proper motion compatible with these groups, 12 have parallax and proper motion similar to one group, but the spatial distribution does not agree with that group. Finally, there are 13 objects having all the *Gaia* parameters compatible with CMA05 or CMA06, but not detected by our HDBSCAN methodology. This means that, among the analysed sources, we missed ~ 3 per cent (13/387) that could be associated with CMA OB1. We also conclude that a significant fraction of the sources that were previously considered possibly related to CMA OB1 are in fact foreground or background objects.

It is interesting to note that, among 117 X-ray sources identified in association with our groups, only 57 were previously classified as very likely members of CMA R1, while the other sources were suggested to be candidates or had an unknown origin (Santos-Silva et al. 2018). In other words, thanks to the present study, it was possible to confirm that the other 60 X-ray sources are indeed members of CMA OB1 association.

B3 Subgroups in CMA OB1

In order to investigate whether there are other known objects that should be detected as members of our groups, we inspected the spatial distribution shown in Fig. B1 searching for possible additional subgroups. By selecting objects with distances < 1700 pc that form subgroups with at least six objects, we detected four small clusters, which are called CMA S1, CMA S2, CMA S3, and CMA S4 in Fig. B3.

The subgroup CMA S1 coincides with the BDSB 96 cluster (Sh 2-297 nebula), discussed in Section A2, while CMA S3 and CMA S4 are scattered around the borders of the CMA shell (Fernandes et al. 2019)

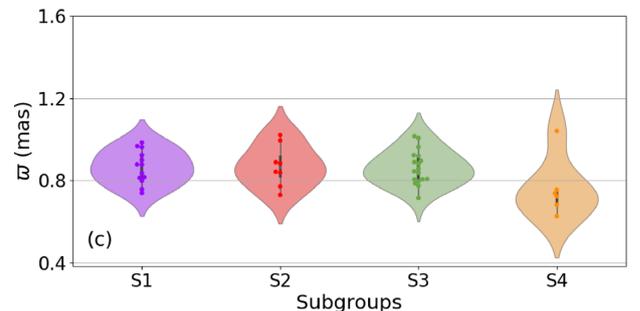
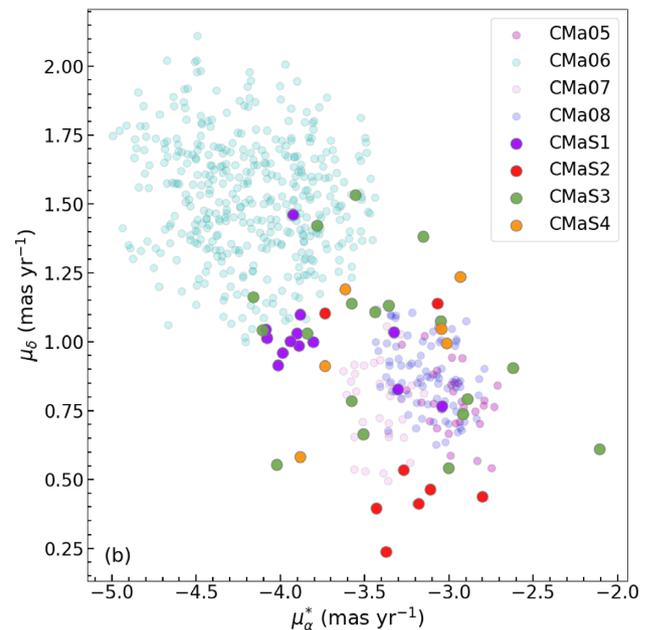
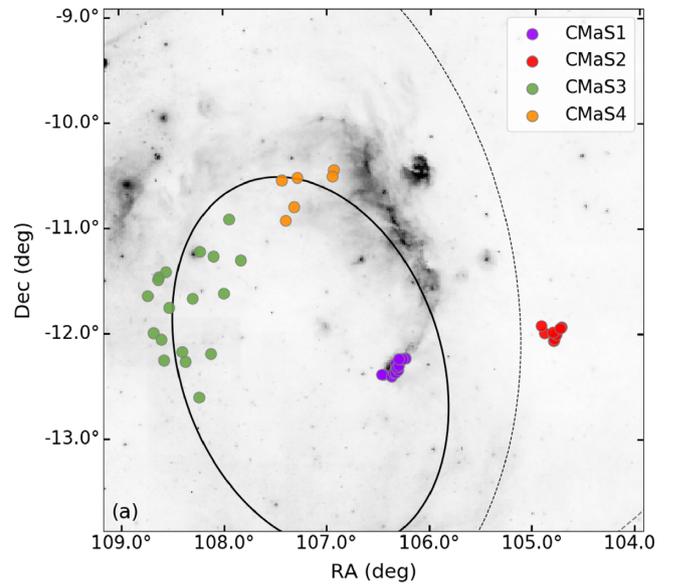


Figure B3. (a) Spatial distribution of four subgroups of stars known in the literature, potentially with CMA OB1 membership on a Digital Sky Survey image (640 nm) of CMA OB1. The black and dashed lines show the CMA shell and CMA supershell from Fernandes et al. (2019). (b) Proper-motion distribution of subgroups, including four groups of CMA OB1 for comparison. (c) Violin histograms of parallaxes for each subgroup.

and CMaS2 appears a few arcmin to the west side of the supershell structure. It is interesting to note that, despite CMaS2 seeming to be outside the structures considered, this subgroup coincides with two unresolved (positional error ~ 30 arcsec) X-ray sources detected by *ROSAT*, called CMaX06 and CMaX07. For both objects, Gregorio-Hetem et al. (2009) identified infrared counterparts with ages in the range 2–5 Myr and mass 2–3 M_{\odot} .

By analysing the proper-motion distribution of these subgroups (see middle panel of Fig. B3), we note that all the subgroups show similarities with results found for our groups CMa05, CMa07, CMa08 and marginally coinciding with CMa06. CMaS2, CMaS3 and CMaS4 are scattered in the same space of proper motion ($\mu_{\alpha} \cos \delta \sim -3.3$, $\mu_{\delta} \sim +0.75$) mas yr^{-1} , while CMaS1 is mainly concentrated around ($\mu_{\alpha} \cos \delta \sim -4$, $\mu_{\delta} \sim +1$) mas yr^{-1} (excepting four dispersed objects). Considering the parallax values (see Fig B2, bottom panel), we note that CMaS1, CMaS2, and CMaS3 have $\varpi \sim 0.9$ mas, corresponding to distances compatible with our groups, while CMaS4 seems to be more distant ($\varpi \sim 0.7$ mas), but still within the range of distances expected for the overall structure of the Association.

In a recent study based on CO observations of a molecular cloud found on the east side of the CMa shell, Gregorio-Hetem (2021a) investigated the stellar population associated with the dense cores revealed by the emission peaks in the CO map. Nine dense cores were found, distributed in filamentary structures that follow the same spatial distribution of dust filaments observed in the infrared. They found 44 cloud members in the region $\alpha = 108.51 \pm 0.08$, $\delta = -12.1 \pm 0.14$, showing *Gaia* parameters ($\mu_{\alpha} \cos \delta = -3.3 \pm 0.3$, $\mu_{\delta} = 0.97 \pm 0.2$) mas yr^{-1} and $\varpi \sim 0.86 \pm 0.3$ mas. These results are similar to those found here for CMaS3, which can be considered an independent confirmation of the existence of this subgroup belonging to the Association. In addition, several of the objects in CMaS3 studied by us were detected in the YSO survey by Sewilo et al. (2019).

All these previous results confirming the young nature of objects appearing in subgroups, as well as the astrometric and kinematic similarities with our physical groups, lead us to suggest that they are actually associated with CMa OB1. As discussed in Section 5.1, suggesting that CMa15 has two subgroups, it is possible that some other large groups have complex distributions because they are composed of several subgroups more or less scattered in the five-parameter space, e.g. also CMa23. This seems to be the case for CMa06, which shows possible small groups in the spatial distribution (see Fig. D6), which could not be resolved by the HDBSCAN method because the subgroups are too scarce. These small groups need to be analysed in more detail, using additional observational information. An example is CMaS1, which coincides with BDSB 96, which we suggest to be related to CMa06 and CMa07 (see Section A2).

Despite the evidence being less strong for the other two subgroups, there are similarities indicating that they may remain as possible members of this complex Association. CMaS2 is located at the same distance as the other groups and coincides in position with X-ray (*ROSAT*) sources. Its proper motion is comparable with group CMa03, and at the lower limit of CMa05, but the declination has an offset of $\sim 1^{\circ}$ from each of these main groups. Finally, CMaS4 has parallax compatible to CM05, as well as proper motion, but the position is more than 1° separate in the sky.

APPENDIX C: ISOCHRONE FITTING CODES

Here we present the two codes, the M20 code and FITCMD, used by us to estimate fundamental parameters of the groups found in this work.

C1 Cross-entropy method

The cross-entropy method we called the M20 code was developed by Monteiro et al. (2020). The code uses theoretical isochrones that are fitted to the *Gaia* DR2 G_{BP} and G_{RP} photometric data and is based on the cross-entropy continuous multi-extremal optimization method (CE), which takes into account the astrometric membership of the star, as well the nominal errors of the data. The method has been explained in detail in Monteiro et al. (2020) and references therein. In summary, the CE method involves an iterative statistical procedure, where the following operations are performed in each iteration:

- (i) random generation of the initial sample of fitting parameters, respecting predefined criteria based on an assumed distribution;
- (ii) selection of the 10 per cent best candidates based on weighted likelihood values;
- (iii) generation of a new random sample of fitting parameters derived from the distribution obtained based on the 10 per cent best candidates of the previous step;
- (iv) repeat until convergence or stopping criteria are reached.

The code interpolates from the Padova (PARSEC version 1.2S) database of stellar evolutionary tracks and isochrones (Bressan et al. 2012), generated for the *Gaia* filter passbands of Maíz Apellániz & Weiler (2018), scaled to solar metal content with $Z_{\odot} = 0.0152$. The grid is constructed from isochrones with steps of 0.05 in $\log(\text{age})$ and 0.002 in metallicity. The search for the best solution can be performed in the following parameter space:

- (i) age: from $\log(\text{age}) = 6.60$ to $\log(\text{age}) = 10.15$;
- (ii) distance: from 1–25 000 pc;
- (iii) A_V : from 0.0–5.0 mag;
- (iv) [Fe/H]: from -0.90 to $+0.70$ dex.

Given the large passbands of the *Gaia* filters, the extinction coefficients depend on the colour and the extinction itself. To account for this effect, we use an updated extinction polynomial for the *Gaia* DR2 photometric bandpasses, as presented in detail by Monteiro et al. (2020).

The code also allows the user to specify priors in the parameter distance, A_V , and [Fe/H]. In general, when not specified in the text, we adopt distance $\mathcal{N}(\mu, \sigma^2)$ obtained with Bayesian inference from the parallax (ϖ) and its uncertainty (σ_{ϖ}). The variance (σ^2) is obtained from the distance interval calculated from the inference using the uncertainty as $1\sigma_{\varpi}$. The prior in A_V is also adopted as a normal distribution with μ and variance (σ^2) for each cluster taken from the 3D extinction map produced by Capitanio et al. (2017).¹⁵ The prior for [Fe/H] was estimated from the Galactic metallicity gradient published by Donor et al. (2020). For the age we adopt a flat prior, so that $P(X_i) = 1$.

To estimate uncertainties in the fundamental parameters, we used a Monte Carlo technique, resampling in each run with a replacement in the original dataset, to perform a bootstrap procedure. The isochrones are also re-generated in each run from the adopted initial mass function (IMF). The final fundamental parameters and respective errors were estimated by the mean and one standard deviation of ten runs. The code also provides radial velocities, the radius containing 50 per cent of members (r_{50}) and maximum radius (r_{max}), containing all the stars in each group. The radial velocities (RV) are estimated from *Gaia* DR2 data of the individual stars. However, there are no

¹⁵The 3D extinction map is available online at <https://stilism.obspm.fr/>

Table C1. Group parameters obtained by the M20 code.

| Group | $r_{50}^{(a)}$ (pc) | $r_{50}^{(a)}$ (arcmin) | $r_{\max}^{(b)}$ (arcmin) | $\log(\text{Age})^{(c)}$ (dex) | $A_G^{(d)}$ (mag) | $\text{RV}^{(e)}$ (km s ⁻¹) | $\text{NRV}^{(f)}$ |
|-------|------------------------|----------------------------|------------------------------|-----------------------------------|----------------------|--|--------------------|
| CMa00 | 0.363 | 22 | 67 | 7.93 ± 0.11 | 0.38 ± 0.36 | 39 ± 7 | 9 |
| CMa01 | 0.163 | 10 | 28 | 8.24 ± 0.09 | 0.52 ± 0.41 | – | 0 |
| CMa02 | 0.162 | 10 | 27 | 8.14 ± 0.21 | 0.65 ± 0.23 | – | 0 |
| CMa03 | 0.136 | 8 | 30 | 8.73 ± 0.34 | 1.07 ± 0.53 | 56.9 ± 0.7 | 4 |
| CMa05 | 0.217 | 13 | 29 | 7.24 ± 0.67 | 1.17 ± 0.59 | – | 0 |
| CMa06 | 0.278 | 17 | 50 | 7.00 ± 0.05 | 1.54 ± 0.90 | 30 ± 4 | 2 |
| CMa07 | 0.178 | 11 | 19 | 7.12 ± 0.05 | 0.87 ± 0.52 | – | 0 |
| CMa08 | 0.219 | 13 | 29 | 7.26 ± 0.12 | 1.05 ± 0.55 | – | 0 |
| CMa09 | 0.185 | 11 | 33 | 8.52 ± 0.25 | 0.96 ± 0.43 | 8.7 ± 0.3 | 2 |
| CMa13 | 0.148 | 9 | 25 | 8.32 ± 0.13 | 0.31 ± 0.22 | 3 ± 10 | 4 |
| CMa15 | 0.219 | 13 | 63 | 8.06 ± 0.14 | 0.48 ± 0.37 | 67.1 ± 0.3 | 2 |
| CMa17 | 0.245 | 15 | 68 | 8.21 ± 0.22 | 0.67 ± 0.30 | 15.7 ± 0.7 | 13 |
| CMa18 | 0.132 | 8 | 27 | 8.26 ± 0.10 | 0.57 ± 0.30 | 15.7 ± 0.3 | 2 |
| CMa19 | 0.146 | 9 | 42 | 7.62 ± 0.31 | 1.27 ± 0.55 | – | 0 |
| CMa23 | 0.315 | 19 | 63 | 8.12 ± 0.22 | 0.97 ± 0.32 | – | 0 |

Note. ^(a) Radius containing 50% of group members; ^(b) maximum radius; ^(c) age logarithm; ^(d) photometric extinction in *G* band; ^(e) radial velocity; ^(f) number of stars used to calculate the group radial velocity.

Table C2. Group fundamental parameters evaluated by FITCMD.

| Group | $M_g^{(a)}$ (M_\odot) | $(m - M)_0^{(b)}$ (mag) | $E(B - V)^{(c)}$ (mag) | $Z^{(d)}$ | $Z/Z_\odot^{(e)}$ |
|-------|--|---|---|--|--|
| CMa00 | 317 ⁺¹⁴⁶ ₋₁₁₆ | 8.699 ^{+0.004} _{-0.025} | 0.100 ^{+0.003} _{-0.020} | 0.0180 ^{+0.0020} _{-0.0020} | 1.18 ^{+0.13} _{-0.13} |
| CMa01 | 213 ⁺⁷⁰ ₋₆₂ | 10.03 ^{+0.03} _{-0.09} | 0.14 ^{+0.03} _{-0.01} | 0.0170 ^{+0.0010} _{-0.0005} | 1.12 ^{+0.06} _{-0.03} |
| CMa02 | 310 ⁺¹⁴⁷ ₋₁₁₇ | 10.04 ^{+0.16} _{-0.24} | 0.40 ^{+0.03} _{-0.02} | 0.0050 ^{+0.0037} _{-0.0005} | 0.33 ^{+0.24} _{-0.03} |
| CMa03 | 507 ⁺¹⁹⁹ ₋₁₆₉ | 10.94 ^{+0.08} _{-0.11} | 0.38 ^{+0.06} _{-0.02} | 0.0300 ^{+0.0005} _{-0.0005} | 1.97 ^{+0.03} _{-0.03} |
| CMa05 | 265 ⁺¹⁷³ ₋₁₂₀ | 10.26 ^{+0.02} _{-0.06} | 0.37 ^{+0.02} _{-0.01} | 0.0120 ^{+0.0010} _{-0.0005} | 0.79 ^{+0.07} _{-0.03} |
| CMa06 | 1990 ⁺⁶⁴³ ₋₅₇₃ | 10.15 ^{+0.05} _{-0.02} | 0.38 ^{+0.01} _{-0.01} | 0.0180 ^{+0.0010} _{-0.0005} | 1.18 ^{+0.07} _{-0.03} |
| CMa07 | 75 ⁺⁴⁵ ₋₃₂ | 10.19 ^{+0.01} _{-0.01} | 0.210 ^{+0.003} _{-0.010} | 0.0180 ^{+0.0010} _{-0.0010} | 1.18 ^{+0.07} _{-0.07} |
| CMa08 | 195 ⁺⁹² ₋₇₄ | 10.01 ^{+0.12} _{-0.03} | 0.41 ^{+0.03} _{-0.02} | 0.0170 ^{+0.0010} _{-0.0005} | 1.12 ^{+0.07} _{-0.03} |
| CMa09 | 646 ⁺²¹⁷ ₋₁₉₂ | 10.43 ^{+0.19} _{-0.04} | 0.35 ^{+0.04} _{-0.03} | 0.0210 ^{+0.0010} _{-0.0005} | 1.38 ^{+0.06} _{-0.03} |
| CMa13 | 202 ⁺⁶⁸ ₋₅₉ | 9.66 ^{+0.05} _{-0.16} | 0.12 ^{+0.02} _{-0.01} | 0.0160 ^{+0.0010} _{-0.0020} | 1.05 ^{+0.06} _{-0.13} |
| CMa15 | 1510 ⁺³⁹⁰ ₋₃₅₈ | 10.49 ^{+0.09} _{-0.11} | 0.15 ^{+0.02} _{-0.01} | 0.0290 ^{+0.0010} _{-0.0005} | 1.91 ^{+0.06} _{-0.03} |
| CMa17 | 3130 ⁺¹³⁶⁰ ₋₁₁₁₀ | 10.57 ^{+0.03} _{-0.17} | 0.28 ^{+0.03} _{-0.02} | 0.0180 ^{+0.0030} _{-0.0010} | 1.18 ^{+0.20} _{-0.07} |
| CMa18 | 699 ⁺²⁶⁵ ₋₂₂₄ | 10.11 ^{+0.05} _{-0.16} | 0.19 ^{+0.03} _{-0.02} | 0.0210 ^{+0.0010} _{-0.0005} | 1.38 ^{+0.06} _{-0.03} |
| CMa19 | 1600 ⁺¹³⁴⁰ ₋₈₀₈ | 10.75 ^{+0.10} _{-0.67} | 0.400 ^{+0.002} _{-0.030} | 0.0300 ^{+0.0005} _{-0.0140} | 1.97 ^{+0.03} _{-0.92} |
| CMa23 | 1400 ⁺⁵⁸³ ₋₄₈₀ | 10.95 ^{+0.05} _{-0.26} | 0.39 ^{+0.03} _{-0.02} | 0.0210 ^{+0.0010} _{-0.0005} | 1.38 ^{+0.06} _{-0.03} |

Note. ^(a) Total mass cluster; ^(b) apparent distance modulus; ^(c) colour excess; ^(d) metallicity; ^(e) metallicity in terms of Solar values, assuming $Z_\odot = 0.0152$.

groups with more than 13 stars with RV values; most of them have only one or two objects with this parameter, so that the estimates of RV for our groups are not reliable and they were not used in this work. The additional parameters estimated using the M20 code are presented in Table C1.

C2 FITCMD

The FITCMD code was developed by Bonatto (2019). In general terms, the code is a statistical approach intended to extract fundamental parameters of star clusters by means of the photometric information contained in the observed colour–magnitude diagram (CMD). Based on properties of the IMF obtained from isochrones,

FITCMD searches for values of the total mass stored in stars (or cluster mass, M_{cl}), age (t_{age}), global metallicity (Z), foreground extinction (or colour excess, CEX), and apparent distance modulus (DM) in order to build the synthetic CMD that reproduces the observed one best. Magnitude-dependent photometric completeness and photometric scatter are also taken into account. Finally, the best-fitting parameters are found by minimizing the residual differences between the synthetic and observed CMDs by means of the global optimization algorithm Simulated Annealing (Goffe, Ferrier & Rogers 1994). Other applications of FITCMD can be found in Piatti & Bonatto (2019) and Bonatto & Chies-Santos (2020).

The FITCMD version used here (2020 April 14) works on several photometric systems. In particular, for this work we use *Gaia* [Vega-

Mags] (G , G_{BP} , G_{RP}) with PARSEC v1.2S + COLIBRI PR16/NBC models (Bressan et al. 2012; Marigo et al. 2017). The parameter space allowed by the code is as follows:

- (i) masses: starting at $0.1 M_{\odot}$;
- (ii) DM: free;
- (iii) CEx: free;
- (iv) ages: from 1 Myr and 13.5 Gyr;
- (v) Z : from 0.0001–0.03, assuming solar metallicity $Z_{\odot} = 0.0152$.

For adequate coverage, ages vary in steps of $\Delta t = 1$ Myr for 1–10 Myr, $\Delta t = 2$ Myr for 10–20 Myr, $\Delta t = 5$ Myr for 20–50 Myr, $\Delta t = 10$ Myr for 50–100 Myr, $\Delta t = 25$ Myr for 100–500 Myr, $\Delta t = 50$ Myr for 500–1000 Myr, and $\Delta t = 250$ Myr for 1000–13 500 Myr, and the metallicities vary in steps of $\Delta Z = 10^{-4}$ from 10^{-4} – 10^{-3} and $\Delta Z = 10^{-3}$ from 10^{-3} – 3×10^{-2} .

To derive the visual extinction, FITCMD uses the relation $A_V = 3.1 \times E(B-V)$ and for the metallicities it uses $[\text{Fe}/\text{H}] = \log_{10}(Z/Z_{\odot}) - 0.8[\alpha/\text{Fe}] - 0.05[\alpha/\text{Fe}]^2$. For consistence with the M20 code we adopt $[\alpha/\text{Fe}] = 0.0$.

Additional parameters computed with FITCMD for the 15 clusters studied in this work are presented in Table C2.

APPENDIX D: TABLES AND ANALYSIS FIGURES

The table containing the properties of the members of the 29 groups, Table D1, is available in its entirety in machine-readable form at the CDS.

Also, when running HDBSCAN our code also makes analysis images containing histograms and distributions of the parameter

members for each group: these are presented in Figs D1–D15. In the first row RA and Dec. are presented; in the second row, proper motion. In the last panels of these rows, the group members are contrasted with the *Gaia* total sample (grey points); the first two panels in the third row are proper motion (pm) and tangential velocity (V_t) vectors, respectively, while the third and fourth panels are histograms of the visual extinction from 3D maps (Green et al. 2019) and the membership probability of the group contents; in the fourth row we present histograms of parallax, distances obtained from it ($1/\varpi$), and distances from Bailer-Jones et al. (2018). In the last panel a comparison between both distances is shown. The members with $P \geq 50$ per cent are presented in bright colours, while those with $P < 50$ per cent are in light colours. In the histograms, the median of each distribution, considering only stars with $P \geq 50$ per cent, which were used to determine the parameters of each group presented in Table 1 (see Section 4.1), is indicated by a solid line, while the 16th and 84th percentiles (used for the uncertainties) are dashed lines. Moreover, the 2.5th and 97.5th percentiles are represented by dot–dashed lines and 0.15th and 99.85th percentiles by dotted lines. In this Appendix we present the diagnostic figures for the 15 groups discussed in this work; the figures of another 14 groups are presented in the online supplementary material.

In addition to the analysis figures, we also present, for the 15 groups, in the last line the colour-magnitude diagrams provided by each fitting code: on the left is the CMD from the M20 code and on the right is the CMD from FITCMD. In both panels, the line represents the best-fitting isochrone from the codes. The M20 code CMD shows the distribution of stars according to membership probability, while the FITCMD diagram also presents the synthetic Hess diagram generated by this code (coloured hatched area).

Table D1. Parameters of the members of group CMa00 selected by HDBSCAN.

| ID_{Gaia} | P (%) | RA (deg) | Dec. (deg) | $\mu_{\alpha} \cos \delta$ (mas yr $^{-1}$) | μ_{δ} (mas yr $^{-1}$) | ϖ (mas) | D_{Λ} (pc) | A_V (mag) | V_t (km s $^{-1}$) |
|---------------------|------------|---------------|---------------|---|-------------------------------------|-------------------|-----------------------|----------------|--------------------------|
| 3051226788077089664 | 100 | 107.36 ± 0.04 | −9.15 ± 0.05 | −5.24 ± 0.08 | 2.29 ± 0.08 | 1.74 ± 0.06 | 565 | 0.37 | 11.74 |
| 3051233312124183168 | 100 | 107.46 ± 0.07 | −9.09 ± 0.07 | −5.56 ± 0.14 | 2.65 ± 0.15 | 1.64 ± 0.10 | 604 | 0.45 | 14.66 |
| 3051226375760232960 | 99 | 107.34 ± 0.06 | −9.17 ± 0.07 | −5.88 ± 0.12 | 2.62 ± 0.10 | 1.55 ± 0.09 | 637 | 0.38 | 17.80 |
| 3051254894342334592 | 90 | 107.17 ± 0.05 | −8.88 ± 0.05 | −5.61 ± 0.11 | 2.34 ± 0.09 | 1.81 ± 0.07 | 545 | 0.50 | 12.40 |
| 3051255787695465344 | 72 | 107.19 ± 0.04 | −8.81 ± 0.05 | −5.15 ± 0.09 | 2.67 ± 0.08 | 1.75 ± 0.06 | 562 | 0.38 | 11.29 |
| 3051247163401380992 | 66 | 106.94 ± 0.04 | −9.00 ± 0.04 | −5.66 ± 0.08 | 2.44 ± 0.09 | 1.79 ± 0.05 | 550 | 0.32 | 12.79 |
| 3051203728391103232 | 56 | 106.98 ± 0.12 | −9.26 ± 0.13 | −5.45 ± 0.27 | 2.24 ± 0.28 | 1.52 ± 0.15 | 655 | 0.41 | 16.31 |
| 3051241356605265408 | 99 | 107.38 ± 0.04 | −8.86 ± 0.05 | −5.34 ± 0.09 | 2.32 ± 0.09 | 1.81 ± 0.05 | 543 | 0.30 | 11.32 |
| 3051338349846802560 | 80 | 107.74 ± 0.11 | −8.82 ± 0.12 | −5.57 ± 0.24 | 2.26 ± 0.27 | 1.73 ± 0.15 | 574 | 0.35 | 13.41 |
| 3051330932445906560 | 100 | 107.50 ± 0.03 | −8.91 ± 0.04 | −5.58 ± 0.07 | 2.45 ± 0.07 | 1.86 ± 0.04 | 530 | 0.24 | 11.69 |

Notes. We provide for each star the *Gaia* DR2 identifier, membership probability (P), position, proper motion, and parallax (without zero-point offset), distance obtained by Bailer-Jones et al. (2018), visual extinction from 3D maps (Green et al. 2019), and tangential velocities in the equatorial system. This entire table containing 152 CMa00 members is available in machine-readable form at the CDS.