144. How many open clusters?

In essay #74, in May 2022, I looked at some of the work on open clusters that had been enabled with Gaia. At that time, only Early Data Release 3 (EDR3) had been made available. Even so, by then, some two hundred or more papers dealing with a range of cluster-related studies had been published. In his review of the literature at the time, 'Milky Way Star Clusters and Gaia', Cantat-Gaudin (2022) wrote that: *'The Gaia data have unlocked a deluge of new results related to many astronomical topics and transformed our ability to study star clusters and stellar structures in the Milky Way.'*

With Data Release 3 now having been available for more than a year, I want to look here at one specific impact that Gaia is having on on the field: how many open clusters are now known in our Galaxy?

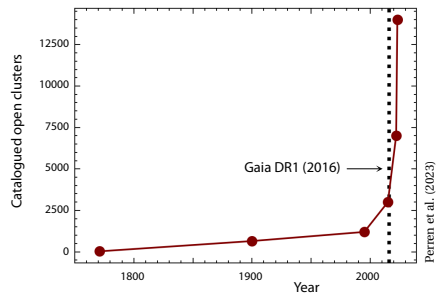
Firstly, some context. Open clusters are groups of, perhaps, several hundred stars, sites of star formation that were formed from the same giant molecular cloud, and therefore all of roughly the same age. They are still being formed in our Galaxy, at an estimated rate of one every few thousand years. Before Gaia, more than a thousand were known within our Galaxy, and many more, perhaps tens of thousands, are thought to exist.

Being only loosely bound by gravity when formed, open clusters slowly disperse as gas (and therefore mass) is stripped by the radiation pressure of their hot young stars. They are further disrupted by close encounters with other cluster stars, and other external structures, as they orbit the Galaxy. They may survive as recognisable clumps for a few hundred million years, with the most massive surviving for a few billion.

In essay #74, I summarised some of the new clusters being discovered by Gaia, on studies of their ages and Galactic orbits, on their role in the structure of spiral arms, and on their dynamical evolution, including their dissolving 'tidal tails' and associated runaway stars.

With DR3 in June 2022, more stars (more than 1.8 billion having astrometry and G magnitudes) and improved distances and space motions, have together enabled many new insights.

But I will concentrate here on just one metric: the *number* of open clusters known today. This figure, from Perren et al. (2023), tells a key part of the story. Since the availability of Gaia DR1 in 2016, and especially with DR2 in 2018, the number of known Galactic open clusters has increased dramatically. From more than 3000 open clusters catalogued pre-Gaia, the current number stands at nearly 14 000, the large majority of these having been discovered with Gaia.



But there are a number of challenges in keeping track of the total number of known clusters in such a rapidly developing field. Given that some studies are now reporting hundreds of new clusters at a time, which of the new discoveries reported in any given paper have already been claimed as new discoveries, under a different name, and probably with some different members?

Another important boundary condition is this: what is the minimum number of comoving stars required to claim a new cluster, given that at the very sparsest levels, even pairs of comoving stars may point to a long-disrupted cluster (e.g. Kamdar et al., 2019). And, given the different discovery methods now being used to identify spatial and kinematic clustering, which of the claimed discoveries are unlikely to be real clusters?

On the first point, and already making use of Data Release 2, Cantat-Gaudin et al. (2018) examined the extensive pre-Gaia open cluster catalogue compilations of Dias et al. (2002) and Kharchenko et al. (2013). They found that, from more than 3000 identified clusters, they could confirm only around 1200.

THE DISCREPANCY arises because many of the previously suggested (pre-Gaia) clusters are now known to be simply 'asterisms', over-densities arising (for example) from strong extinction patterns mainly in the direction of the Galactic bulge (Kos et al., 2018).

But it is not simply some 'numbers' that are at stake. For example, excluding the non-clusters results in a cluster age function in better agreement with theoretical models of their formation and dissolution (Anders et al., 2021). And a cleaner sample of true clusters allowed Cantat-Gaudin et al. (2018) to conclude that the youngest clusters are concentrated near the Galactic plane and trace the Galaxy's spiral arms, while older objects are more uniformly distributed, typically further from the plane, and at larger Galactocentric distances.

A COMPILATION of new open cluster discoveries with Gaia has been made by Perren et al. (2023), and their Table 1 is included here (in slightly revised form).

Amongst the discoveries made with Gaia DR2, a number of papers have reported more than 200 new clusters: Sim et al. (2019) reported 207 within 1 kpc; Castro-Ginard et al. (2020) used machine-learning to find 570 with $|b| < 20°$; Cantat-Gaudin et al. (2020) reported 2017 out to 4 kpc; and Dias et al. (2021) found 1742, also making use of Gaia radial velocities.

Papers reporting 200 or more new open clusters increased further with EDR3: Castro-Ginard et al. (2022) found 628, mostly beyond 1 kpc; He et al. (2022a) found 541 mostly within 3 kpc; He et al. (2022b) reported 836, including 46 with $|b| > 20°$, resulting in a nearly threefold increase in the number at high Galactic latitudes; Hao et al. (2022) found 703; Tarricq et al. (2022) found 467 searching out to 50 pc from the cluster centres; and He et al. (2023) discovered 1656 beyond 1.2 kpc, also using unsupervised machine-learning.

And with DR3, Chi et al. (2023b) discovered 1179 new clusters within 5 kpc; and Hunt & Reffert (2023) discovered 6272. The latter used a blind, all-sky search using 729 million sources to $G = 20$, creating a homogeneous catalogue of clusters including many new objects, and making use of the popular *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) algorithm (Campello et al., 2013).

INDIVIDUAL DISCOVERIES can also be of particular interest. As one example which illustrates the difficulties of cluster identification in the pre-Gaia era as a result of the very high density of background sources, even at moderate distances, Negueruela et al. (2021) used Gaia DR2 astrometry to identify a new massive young cluster at 2.3 kpc in the Sagittarius arm with more than 30 stars brighter than $G = 13$, which they inferred to be one of the most massive clusters in the solar neighbourhood with an initial mass of $\sim 10^4 M_\odot$.

Reference	Gaia	N
Kharchenko et al. (2012)	–	2854
Loktin & Popova (2017)	–	1050
Bica et al. (2019)	–	3555
Castro-Ginard et al. (2018)	DR2	23
Castro-Ginard et al. (2019)	DR2	53
Ferreira et al. (2019)	DR2	3
Liu & Pang (2019)	DR2	76
Sim et al. (2019)	DR2	207
Castro-Ginard et al. (2020)	DR2	570
Cantat-Gaudin et al. (2020)	DR2	2017
Ferreira et al. (2020)	DR2	25
Hao et al. (2020)	DR2	16
Casado (2021)	DR2	20
Dias et al. (2021)	DR2	1742
Ferreira et al. (2021)	DR2	34
He et al. (2021)	DR2	74
Hunt & Reffert (2021)	DR2	41
Jaehnig et al. (2021)	DR2	11
Santos-Silva et al. (2021)	DR2	5
Castro-Ginard et al. (2022)	EDR3	628
He et al. (2022a)	EDR3	541
He et al. (2022b)	EDR3	836
Hao et al. (2022)	EDR3	703
Li et al. (2022)	EDR3	61
Tarricq et al. (2022)	EDR3	467
Chi et al. (2023a)	EDR3	82
Chi et al. (2023c)	EDR3	46
He et al. (2023)	EDR3	1656
Li & Mao (2023)	EDR3	35
Chi et al. (2023b)	DR3	1179
Hunt & Reffert (2023)	DR3	6272
Qin et al. (2023)	DR3	101

THE TABLE ABOVE is the result of the ambitious compilation by Perren et al. (2023). Their *Unified Cluster Catalogue* is by far the largest catalogue of open clusters in the Milky Way, with almost 14 000 clusters listed. Derived from Gaia DR3 data up to $G = 20$, each cluster is processed with their specific probability membership algorithm, incorporating each star's coordinates, parallax, proper motions, and associated uncertainties.

Their compilation of previous open cluster catalogues results in a combined list of 24 983 open clusters, which reduced to 13 684 *unique* open clusters after cross-matching, with more than a million probable members identified in total.

Their Unified Cluster Catalogue is available online at https://ucc.ar, and I urge interested readers to take a look. From its intuitive interface, enter the name of your chosen cluster, using whatever alias you know it by, for example Westerlund 1 (essay #106). You get, amongst other resources, a zoomable sky map, various plots for selected probable members, and links to Simbad.

WOULD ANYONE care to guess how many will be in the final Gaia catalogue, a decade from now?!