Edinburgh Research Explorer

Discovery of a Disrupting Open Cluster Far into the Milky Way Halo: A Recent Star Formation Event in the Leading Arm of the Magellanic Stream?

Discovery of a disrupting open cluster far into the Milky Way halo:
a recent star formation event in the leading arm of the Magellanic stream?

Adrian M. Price-Whelan,[1] David L. Nidever,[2,3] Yumi Choi,[2,4] Edward F. Schlafly,[5] Timothy Morton,[1,6,7]
Sergey E. Koposov,[8,9] and Vasily Belokurov[9,7]

[1]*Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA*
[2]*Department of Physics, Montana State University, P.O. Box 173840, Bozeman, MT 59717, USA*
[3]*National Optical Astronomy Observatory, 950 North Cherry Ave, Tucson, AZ 85719, USA*
[4]*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson AZ, 85721, USA*
[5]*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*
[6]*Department of Astronomy, University of Florida, 211 Bryant Space Science Center, P.O Box 112055, Gainesville, FL 32611-2055 USA*
[7]*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*
[8]*McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
[9]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

ABSTRACT

We report the discovery of a young ($\tau \sim 117$ Myr), low-mass ($M \sim 1200$ M$_\odot$), metal-poor ([Fe/H] ~ -1.14) stellar association at a heliocentric distance $D \approx 28.7$ kpc, placing it far into the Milky Way halo. At its present Galactocentric position $(R, z) \sim (23, 15)$ kpc, the association is (on the sky) near the leading arm of the gas stream emanating from the Magellanic cloud system, but is located $\approx 60°$ from the Large Magellanic Cloud (LMC) center on the other side of the Milky Way (MW) disk. If the cluster is co-located with HI gas in the stream, we directly measure the distance to the leading arm of the Magellanic stream. The measured distance is inconsistent with Magellanic stream model predictions that do not account for ram pressure and gas interaction with the MW disk. The estimated age of the cluster is consistent with the time of last passage of the leading arm gas through the Galactic midplane; We therefore speculate that this star-formation event was triggered by its last disk midplane passage. Most details of this idea remain a puzzle: the Magellanic stream has low column density, the MW disk at large radii has low gas density, and the relative velocity of the leading arm and MW gas is large. However it formed, the discovery of a young stellar cluster in the MW halo presents an interesting opportunity for study. This cluster was discovered with *Gaia* astrometry and photometry alone, but follow-up DECam photometry was crucial for measuring its properties.

Keywords: Galaxy: open clusters and associations – Galaxy: halo – stars: formation – surveys

1. INTRODUCTION

The distant stellar halo of the Milky Way (i.e. far from the disk midplane, $|z| \gtrsim 5$ kpc) is typically characterized by its old ($\gtrsim 10$ Gyr), metal-poor ([Fe/H] ≈ -1.5) stellar population. The typical old age of the stellar halo is understood as a signature of the dominant (in stellar mass) progenitor systems that were accreted early on in the formation of the Galaxy (massive dwarf galaxies Deason et al. 2015; Fiorentino et al. 2015). It is thought that these systems came in with significant

gas reservoirs, but were quenched and stripped through collisional processes that heated and dispersed the gas (e.g., Mayer et al. 2006), thus preventing immediate star formation in the deposited gas. The Milky Way, however, continues to accrete satellite galaxies, as is evidenced by the prominent stellar stream from the Sagittarius dwarf galaxy (Ibata et al. 1994; Majewski et al. 2003), the presence of the Large and Small Magellanic Clouds (LMC and SMC), and about 50 dwarf satellites within the Galactic halo. While Sagittarius was likely stripped of its neutral gas long ago (Burton & Lockman 1999; Tepper-García & Bland-Hawthorn 2018), the LMC–SMC system is associated with $\approx 8 \times 10^8$ M$_\odot$ of HI gas (Brüns et al. 2005), which extends into leading

adrn@astro.princeton.edu

Corresponding author: Adrian M. Price-Whelan

and trailing gas streams (Mathewson et al. 1974; Putman et al. 1998; Brüns et al. 2005; Nidever et al. 2010). Here we report the first discovery of a young star cluster in the Milky Way halo that appears to be associated with this LMC–SMC gas stream, which suggests that young stars can form from tidally stripped gas during low-mass mergers and may therefore exist throughout the otherwise aging stellar halo.

The Magellanic stream (MS), including gas in the leading arm (LA), is a large stream of predominantly hydrogen gas emanating from the LMC–SMC system that wraps nearly $\approx 200°$ around the sky (Mathewson et al. 1974; Putman et al. 1998; Brüns et al. 2005; Nidever et al. 2010) and contains a significant fraction of the total gas mass associated with the LMC–SMC (Brüns et al. 2005). The trailing MS has been studied in great detail by large-area and high-resolution radio sky surveys: Recent surveys have found small-scale structure and gas fragmentation (e.g., Nidever et al. 2008; For et al. 2014) and a large-scale bifurcation, with kinematically (Nidever et al. 2008) and chemically (Fox et al. 2013) distinct "strands" that lead back to the LMC and the SMC. The LA gas has been found to connect to regions of low-column-density gas ($N \sim 10^{18}$–10^{19} cm^{-2}) on the other side of the Galactic disk (Putman et al. 1998; Nidever et al. 2008), and has been decomposed into distinct gas features named LA I–IV (Brüns et al. 2005; Nidever et al. 2008; Venzmer et al. 2012).

The origin and formation of these LA features is still uncertain. Initial studies of the LA argued that the features closest to the LMC–SMC can be traced back to the SMC (Putman et al. 1998), but outer features of the LMC appear to lead directly into the LA I feature (Nidever et al. 2008). Recent chemical abundance measurements along several sight-lines passing through the LA again support an SMC origin for the gas (Fox et al. 2013, 2018; Richter et al. 2018). Whatever the origin of the LA gas, it is clear that tidal stripping by the Milky Way is required to form the LA (Nidever et al. 2008; Besla et al. 2012). However, the LA features deviate from the predicted orbit of the LMC–SMC, implying that ram pressure or interactions the outer Milky Way disk (from the recent disk plane passage) may have removed orbital energy from the LA gas (e.g., Bekki et al. 2008).

The gas in the MS encodes information about the past and future trajectory of the LMC–SMC, and about interactions between this gas and the Milky Way. Combined with recent proper motion measurements of the LMC (Kallivayalil et al. 2006, 2013) and improved models for the LMC–SMC that suggest they are on their first passage through the Galaxy (Besla et al. 2007, 2010, 2012), several groups have used the MS to constrain properties of both the interaction history of the LMC–SMC and of the dark matter halo of the Milky Way (see recent review by D'Onghia & Fox 2016). In general, these models for the formation of the MS rely on past interactions between the LMC and SMC to pre-process the Magellanic gas distribution before infall and eventual stripping by the tidal field of the Milky Way (Besla et al. 2012; Diaz & Bekki 2012); More recent hydrodynamical simulations of the MCs showed that the repeated encounters between the LMC–SMC strip gas both from the SMC and the LMC, and this tidally stripped gas can then create filamentary structures both in the leading and trailing MS (Pardy et al. 2018). Many observational studies (e.g., Olsen et al. 2011; Noël et al. 2013; Mackey et al. 2016; Carrera et al. 2017; Choi et al. 2018a,b; Zivick et al. 2018; Belokurov & Erkal 2019) of the LMC–SMC themselves will provide strong constraints the interaction history of the MCs.

One critical difficulty in using the MS to further improve models of the LMC–SMC interaction and infall is the lack of distance information along the MS. No significant over-density of stars have been found associated with the trailing MS (Guhathakurta & Reitzel 1998), thus leaving distance and tangential velocity information unknown. While a small number of OB stars have been found in the vicinity of the LA gas (Casetti-Dinescu et al. 2014; Zhang et al. 2017), their sparsity and concentration near the Galactic plane make it difficult to unambiguously associate them with the MS and not runaway OB stars from the Milky Way disk.

In this *Article*, we report the discovery of a young stellar cluster at the far edge of the LA II feature ($L_{\rm MS} \sim 65°$) that is located far into the Galactic halo ($D \sim 28.7$ kpc) and therefore plausibly formed from gas in the leading arm of the MS as it crossed the Galactic disk. This provides the first precise distance measurement to the MS leading arm to be compared with simulations of the LMC–SMC in the Milky Way, and provides an opportunity to study recent star formation in a unique environment. The discovery of *Price-Whelan 1* will enable new modeling efforts that track the infall of the LMC–SMC, the tidal stripping of Magellanic gas, and the interaction of this gas with the Milky Way.

In Section 2, we present the initial discovery with *Gaia* DR2 and follow-up observations with DECam to obtain deeper photometry of the region around the association. In Section 3, we use the *Gaia* data to measure the kinematics, and DECam photometry to infer the age, metallicity, and distance to *Price-Whelan 1* In Section 4, we interpret the inferred stellar population

parameters of *Price-Whelan 1* and discuss plausible formation scenarios. We conclude in Section 5.

2. DATA

2.1. *Cluster discovery with Gaia*

We use astrometric data from the *Gaia* mission (Gaia Collaboration et al. 2016), data release 2 (DR2; Gaia Collaboration et al. 2018a; Lindegren et al. 2018) to search for distant, comoving multiplets of blue stars. Our original intent was to search for small, distant, co-moving groups of blue horizontal branch stars to identify new candidate satellites of the Milky Way. We therefore initially select all stars from *Gaia* with parallax $\varpi < 1$, color $-0.5 < (\mathrm{BP}-\mathrm{RP}) < 0$, G-band magnitude $G < 20$, and Galactic latitude $|b| > 20°$ (see Appendix A for the database query). We further exclude stars within a 15° radius from the LMC, and a 8° radius from the SMC — 27,895 stars remain after these cuts. We then cross-match this catalog to itself with both sky positions and proper motions: we search for pairs of stars that have separations $s < 0.5°$ and proper motion differences $|\Delta\mu| < 0.5$ mas yr^{-1}. We then combine mutually-connected comoving pairs into small groups of stars that are colocated on the sky and comoving in proper motions, and remove groups that have < 4 members. We cross-match the mean sky positions of the groups to locations of local group galaxies (McConnachie 2012) and Milky Way globular clusters (2010 edition; Harris 1996) and filter out all groups that lie within 1 degree of these known objects. After these filters, one group of comoving stars remains at $(\mathrm{RA}, \mathrm{Dec}) \sim (179, -29)°$.

We then query all objects from the *Gaia* DR2 catalog within a rectangle centered on the nominal position of this group, with a width of 5° and a height of 5° in the equatorial (ICRS) coordinate system (see Appendix A for the database query). Figure 1 shows the *Gaia* data for this region: The left panel shows the *Gaia* color-magnitude diagram, extinction corrected following the procedure used in Danielski et al. (2018) and the co-efficients from Gaia Collaboration et al. (2018b), with the blue over-density highlighted by the polygon (blue) and under-plotted with a 100 Myr, [Fe/H] $= -1.1$ MIST isochrone (red line; Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015). Surprisingly, this group appears to be a young, distant main sequence, rather than an old population of horizontal branch stars. The middle and right panels of Figure 1 show sky positions and proper motions of all stars in this sky region (grey background density), and only stars in the CMD selection polygon (black markers).

The *Gaia* data reveal the presence of a young, distant, spatially-clustered, and co-moving stellar over-density

— named *Price-Whelan 1* — but the *Gaia* photometry is too shallow to resolve anything but the brightest main sequence stars. The spatial morphology of the cluster is large on the sky, and interestingly substructured with at least two subcomponents (labeled a and b in Figure 1) that are indistinguishable in terms of their proper motions and uncertainties. With the *Gaia* data alone, the distance, age, and metallicity of the cluster cannot be determined, as these quantities are degenerate where the main sequence is nearly vertical. In the next section, we describe deeper DECam imaging obtained over a portion of the cluster.

2.2. *Follow-up with DECam*

We obtained DECam u-, g-, and i-band imaging of a single field centered on the "a" spatial component of the cluster (see Figure 1) discovered using the *Gaia* data (see previous section). Observations were obtained with the Dark Energy Camera (DECam) on the CTIO 4m Blanco telescope on UT 2018 May 20 with 3×300s exposures in $u-$, $g-$, and i-band. The NOAO Community Pipeline (CP; Valdes et al. 2014; Valdes et al., in preparation) InstCal images, which have the instrumental signature removed, were downloaded from the NOAO Archive[1] for further processing. Forced PSF photometry was performed with the PHOTRED pipeline (Nidever et al. 2017) which uses the DAOPHOT ALLSTAR (Stetson 1987) and ALLFRAME (Stetson 1994) suite of programs.

The instrumental photometry (i.e., $-2.5\log(\mathrm{ADU\ sec}^{-1})$) was calibrated using Pan-STARRS1 (PS1; Chambers et al. 2016) photometry for g- and i-band and with SkyMapper (Wolf et al. 2018) for the u-band observations. We use a robust linear fit to cross-matched sources (that appear in both PS1/SkyMapper and our observations) to obtain a zero-point and (for g- and i-band) color-term to derive color transformations relating our instrumental DECam photometry to calibrated PS1 grizy and SkyMapper u-band photometry (see Table 1 for the derived transformation coefficients). We apply this color transformation to the instrumental photometry to put our photometry on the PS1/DECam system. This calibration enables us to treat the photometry as if it were PS1 (g- and i-band) and DECam (u-band) photometry in the isochrone modeling described below. The robust RMS around the linear fit is 0.028 mag for g- and i-band and 0.164 mag for the u-band; We adopt these RMS values as systematic errors that are added in quadrature to the photometric uncertainties for individual sources in the modeling described below. The

---

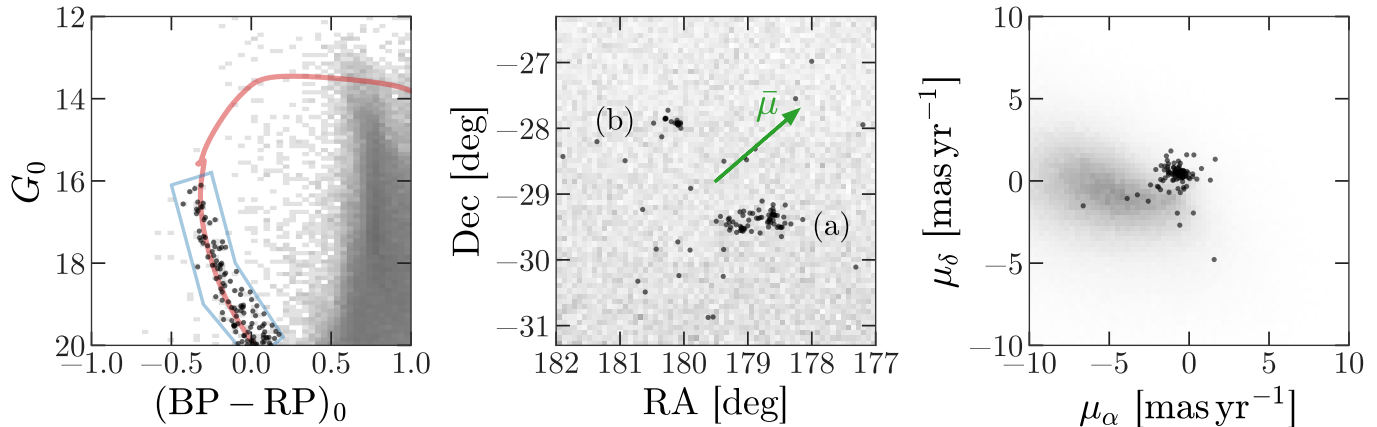[1] http://archive.noao.edu

Discovery of Price-Whelan 1 with Gaia

Figure 1. Color-magnitude diagram, sky positions, and proper motions from *Gaia* DR2 for the region around $(\alpha, \delta) \sim (179, -29)°$. **Left panel**: *Gaia* color-magnitude diagram (CMD) for all sources with $G < 20$ and sky position within 3 degrees of $(\alpha, \delta) = (179.5, -28.8)°$, extinction corrected following Danielski et al. (2018). Red line shows a 100 Myr, [Fe/H] = −1.1 MIST isochrone shifted to a distance of 30 kpc. Shaded pixels (grey 2D histogram) show the density of all sources in the CMD for this sky region, and points (black markers) show only sources in the blue selection polygon shown. **Middle panel**: Sky positions for all sources in the shown sky region (grey 2D histogram). Black markers show only sources in the blue selection polygon shown in the CMD (left panel). Arrow (green) indicates the inferred proper motion direction (Section 3.1). Subcomponents of the cluster (a and b) are indicated. **Right panel**: Proper motions for all sources in the shown sky region (grey 2D histogram). Black markers again show only sources in the blue CMD selection region (left panel).
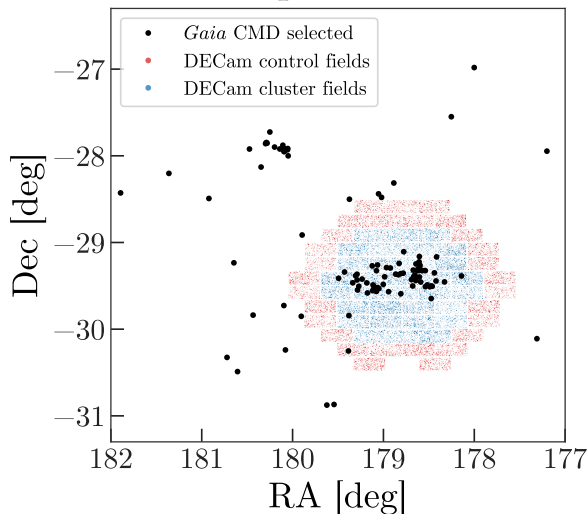

Figure 2. The sky region around *Price-Whelan 1*, showing the same *Gaia* CMD-selected sources as in Figure 1 (black markers), and point sources identified from DECam g-band follow up (blue and red markers). Red markers show sources in the DECam field used as control sources, and blue markers show sources used as cluster member candidates.

larger scatter in u-band—larger than the uncertainty in the individual photometric measurements—is due to the significant differences between the DECam and

SkyMapper u-band passbands as well as the u-band's sensitivity to both temperature and metallicity.

Figure 2 shows the sky positions of the *Gaia* CMD-selected sources (black markers; see same in middle panel, Figure 1), along with point sources identified in the g-band data obtained with DECam: sources in control fields are shown as red markers, and sources in cluster fields are shown as blue markers. We note that the smaller over-density located northeast of the DECam field, component b (see Figure 1), was not followed up in this work.

Figure 3 shows DECam color-magnitude diagrams for the control and cluster sub-fields (the magnitudes here are not extinction-corrected). The blue comoving group, *Price-Whelan 1*, identified in Section 2.1 using *Gaia* data alone shows up as a clear young main sequence in the cluster fields (right panel of Figure 3). Later, we use photometry in the sub-region of the cluster fields identified by the dashed rectangle to infer the cluster parameters.

3. METHODS

In the subsections below, we perform two independent analyses of the data available for *Price-Whelan 1*. First, in Section 3.1, we use astrometric data from *Gaia* DR2 to determine the mean proper motion of the cluster by modeling the kinematics of the cluster and back-
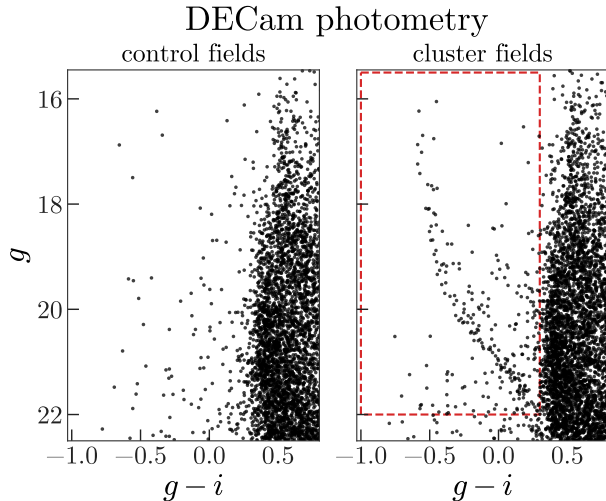
DECam photometry

Figure 3. DECam $g - i$ versus g color-magnitude diagrams for the DECam control fields (left panel) and DECam cluster fields (right panel). Note the prominent young, distant main sequence in the cluster fields: This is the main sequence of *Price-Whelan 1*. The sources in the dashed (red) rectangle in the right panel are later used to measure the cluster stellar population parameters.

Table 1. DECam Photometric Transformation

Band	Color	Zero-point term	Color term
u_{DECam}	$\cdots$	27.178 ± 0.016	$\cdots$
g_{PS1}	$g_{\mathrm{inst.}} - i_{\mathrm{inst.}}$	24.9471 ± 0.0068	-0.03494 ± 0.0051
i_{PS1}	$g_{\mathrm{inst.}} - i_{\mathrm{inst.}}$	24.7534 ± 0.0011	-0.07379 ± 0.0007

ground sources. Then, in Section 3.2, we use photometric data from DECam to assign membership probabilities to sources and simultaneously measure the cluster stellar population parameters (age, metallicity, distance, etc.). We perform these analyses separately because *Gaia* DR2 only includes the most massive members of the cluster, but the addition of lower main sequence members apparent in the DECam photometry provide a much better constraint on the cluster parameters. The parameters derived from this analysis are summarized in Table 4.

3.1. *Inferring the mean proper motion with Gaia*

We measure the mean proper motion of *Price-Whelan 1* by constructing a probabilistic model of the cluster and background populations using astrometric data from *Gaia*. We start by selecting all stars with (BP − RP)$_0$ < 0.35 to remove low-mass and old main sequence star contamination in the region. Figure 4, left (grey points), shows the sky positions of stars that pass this

blue cut in the region around the young cluster. The larger solid-line circle indicates the region we define as the cluster area, and the two smaller dashed-line circles indicate control fields that are combined and used for modeling the background distribution of proper motions. The control fields are designed to, together, have the same total area as the cluster field, and were chosen to have similar latitudes as the cluster field in the Magellanic stream coordinate system (Nidever et al. 2008). We assume that the background density in the cluster field is equivalent to the average background density of the joint control fields: We do not see any gradients or significant differences in the proper motion distribution between these two control fields. From visual inspection of the CMDs of the control fields, we do not see any significant clustered over-density and therefore assume that these fields are dominated by the background stellar density. The right panel of Figure 4 shows the difference of the 2D proper motion distributions in the cluster and summed control fields. The distinct over-density of stars near $(\mu_\alpha, \mu_\delta) \approx (-0.5, 0.5)$ mas yr^{-1} is the identified comoving association of blue stars.[2] Note that Figure 4 is only meant as an illustration: The cluster proper motion is determined probabilistically by taking into account the full *Gaia* covariance matrices for each source, as described below.

To measure the cluster (mean) proper motion, we first construct a model for the error-deconvolved proper motion distribution in the control fields using "extreme deconvolution" (XD; Bovy et al. 2011) with two Gaussian components. XD takes into account the full error distributions for each proper motion measurement $\boldsymbol{\mu} = (\mu_\alpha, \mu_\delta)$, including covariances $\mathbf{C}_\mu$, provided by *Gaia* DR2. After running XD on the proper motion distribution of the control fields, we fix the parameters of the density model and use this as the background model for the cluster field. We model the proper motion distribution in the cluster region using a two-component mixture model with a single, isotropic Gaussian component for the error-deconvolved cluster distribution with mean $\boldsymbol{x}$ and isotropic variance s^2, and the XD-inferred background model, p_{XD}, for the background component. In detail, taking f to be the fraction of blue stars in this region belonging to the young cluster, and $\tilde{\boldsymbol{\mu}}$ to be the

---

[2] Throughout this article, we use μ_α to refer to the proper motion value provided by *Gaia*, which includes the $\cos\delta$ term.
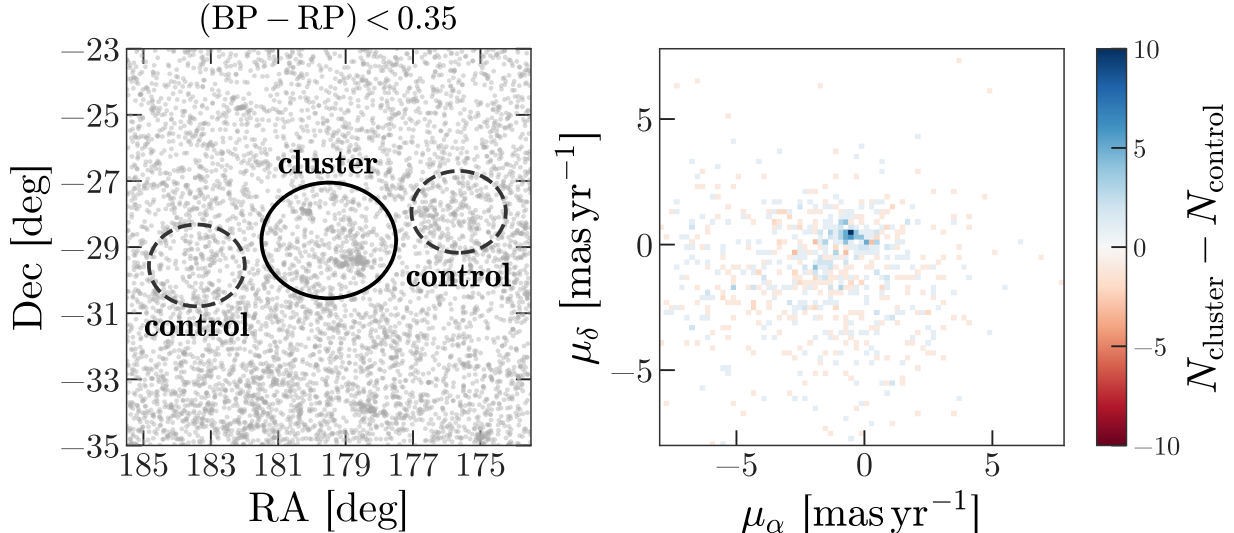
Figure 4. Left panel: Sky positions of all *Gaia* sources in this region with BP − RP < 0.35 (grey markers). The circles show the regions used as cluster and control fields in the proper motion inference (Section 3.1). The control fields, together, have the same area as the cluster field. **Right panel**: The colored 2D density shows the difference in number of sources per pixel in the cluster field versus the two control fields (see left panel). The clear over-density in the cluster field (blue) are stars in *Price-Whelan 1*.

true proper motion for a single star,

$$p(\boldsymbol{\mu} \,|\, \tilde{\boldsymbol{\mu}}, \mathbf{C}_\mu) = \mathcal{N}(\boldsymbol{\mu} \,|\, \tilde{\boldsymbol{\mu}}, \mathbf{C}_\mu) \tag{1}$$

$$p(\tilde{\boldsymbol{\mu}} \,|\, f, \boldsymbol{x}, s) = f\, p_{\mathrm{cl}}(\tilde{\boldsymbol{\mu}} \,|\, \boldsymbol{x}, s) + (1-f)\, p_{\mathrm{XD}}(\tilde{\boldsymbol{\mu}}) \tag{2}$$

$$p_{\mathrm{cl}}(\tilde{\boldsymbol{\mu}} \,|\, \boldsymbol{x}, s) = \mathcal{N}(\tilde{\boldsymbol{\mu}} \,|\, \boldsymbol{x}, s^2\, \mathbb{I}) \tag{3}$$

where $\mathcal{N}(\cdot \,|\, \boldsymbol{y}, \mathbf{C})$ represents the multidimensional normal distribution with mean $\boldsymbol{y}$ and covariance matrix $\mathbf{C}$, and $\mathbb{I}$ is the identity matrix. Because all distributions are Gaussian, the per-star parameters (the true proper motions, $\tilde{\boldsymbol{\mu}}$) can be analytically marginalized out so that the per-star likelihood can be expressed as $p(\boldsymbol{\mu} \,|\, f, \boldsymbol{x}, s, \mathbf{C}_\mu)$. We assume that the measurements for each star, n, are independent so that the full likelihood of all N stars given the parameters $(f, \boldsymbol{x}, s)$ is

$$p(\{\boldsymbol{\mu}_n\}_N \,|\, f, \boldsymbol{x}, s, \{\mathbf{C}_{\mu,n}\}_N) = \prod_n^N p(\boldsymbol{\mu}_n \,|\, f, \boldsymbol{x}, s, \mathbf{C}_{\mu,n}) \quad. \tag{4}$$

We use an ensemble Markov chain Monte Carlo (MCMC) sampler (`emcee`; Foreman-Mackey et al. 2013a; Goodman & Weare 2010) to generate posterior samples over the parameters $(f, \boldsymbol{x}, s)$ using the likelihood defined above (Equation 4), and assuming the following prior probability distributions: uniform over the domain $(-5, 5)$ mas yr^{-1} for each component of $\boldsymbol{x}$, uniform in f, and uniform in log-s over the domain $-6 < \ln s < 4$ (with s in units of mas yr^{-1}). We run the sampler with 32 walkers for 256 steps as burn-in, then reset the sampler and run for an additional 512 steps, after

which the chains appear converged: We compute the Gelman-Rubin (Gelman & Rubin 1992) convergence diagnostic and find that all chains have $R < 1.1$. We then downsample the resulting chains by taking every 16th sample to preserve closer-to-independent samples, leaving a total of 1024 samples; We use these samples to estimate the median posterior parameter values and uncertainties, For the young cluster, we find $\boldsymbol{x} = (-0.56, 0.47) \pm (0.04, 0.02)$ mas yr^{-1}, $\ln s = -3.8 \pm 0.9$, and $f = 0.14 \pm 0.02$. The proper-motion dispersion is consistent with zero, but is unmeasured: 95% of the posterior samples have $s < 0.09$ mas yr^{-1}, indicating that the observed proper motion dispersion is consistent with the proper motion *Gaia* uncertainties, which have minimum and median values of ~ 0.09 mas yr^{-1} and ~ 0.4 mas yr^{-1}, respectively. Table 5 contains a subset of the *Gaia* DR2 data for stars in the cluster field, with kinematic membership probabilities determined from this modeling procedure.

3.2. *Inferring the stellar population parameters with DECamphotometry*

While the *Gaia* data provide exquisite astrometric data for the brightest members of *Price-Whelan 1*, constraints on the stellar population from the *Gaia* data alone are limited by degeneracies in the mapping from the color-magnitude diagram to stellar parameters. In addition, uncertainty in the *Gaia* passbands can cause significant differences in *Gaia* BP − RP color for very blue sources (Maíz Apellániz & Weiler 2018). To measure the cluster population parameters, we therefore use

the deeper photometry from DECam, which provides a clear view of the main sequence of *Price-Whelan 1* down to stellar masses $M \sim 0.9$ M$_\odot$. We note that this analysis and the subsequent discussion only considers subregion (a) of *Price-Whelan 1*, which dominates the total mass and number of stars in the cluster. Figure 3 shows the DECam g- and i-band color-magnitude diagrams for control fields (left) and cluster fields (right) selected from the DECam footprint, once again showing the young main sequence (coherent stellar population in right panel). We use the g- and i-band photometry—and u-band, though only the brightest $\sim$15 cluster field members are robustly detected, however for those sources, the u-band significantly helps to constrain the extinction for the cluster—for individual stars in a sub-section of the cluster field CMD (red dashed outlined region in Figure 3) to infer the stellar population parameters of *Price-Whelan 1*. To summarize our methodology, we first generate independent posterior samplings over the stellar parameters of each individual source under an interim prior (over age, stellar mass, distance, etc.), then use these individual samplings to construct a Bayesian hierarchical model for a two-component mixture model of the cluster and a background population. This methodology is more robust than conventional isochrone-fitting methods that require by-hand fitting of stellar population parameters and naturally propagates uncertainties in the photometry and cluster membership. If you prefer to skip the details, the results of this modeling are presented in Section 4.1.

In detail, we start by using the `isochrones` package (Morton 2015) to generate posterior samplings over stellar parameters for each individual source given its photometry and an interim prior. We use the MIST (Dotter 2016; Choi et al. 2016; Paxton et al. 2011, 2013, 2015) isochrone grid, and `isochrones` automatically performs interpolation between the provided grid of stellar isochrones to predict photometry given a set of stellar parameters. Here, the stellar parameters for each source are its "equal evolutionary point" number EEP (see MIST documentation[3]), age τ, metallicity [Fe/H], extinction A_V, and distance D, which can be uniquely mapped to a point in the observed CMD; The likelihood of a given set of these parameters, $\boldsymbol{\theta} = (\text{EEP}, \tau, [\text{Fe/H}], A_V, D)$, is then computed from the photometry and (assumed Gaussian) photometric uncertainties given the predicted photometry.

[3] http://waps.cfa.harvard.edu/MIST/

To generate posterior samples, we must also specify prior probability distributions over the parameters $\boldsymbol{\theta}$. The priors for each parameter are summarized in Table 2. The bounds on EEP limit the isochrone to evolutionary phases between the zero-age main sequence to the terminal-age main sequence, but this prior is actually computed using the stellar mass computed from the isochrone parameters (see caption of Table 2). The prior and bounds on A_V are set to prefer small and reasonable values of extinction for this moderately high Galactic latitude region. The prior on distance assumes a uniform space density of stars.

Parameter	Prior	Bounds
τ	uniform	$(10 \text{ Myr}, 15 \text{ Gyr})$
[Fe/H]	uniform	$(-2, 0.5)$
EEP	(see caption)	$(202, 355)$
A_V	$\propto A_V^{-1}$	$(0.001, 1)$ mag
D	$\propto D^2$	$(1, 100)$ kpc

Table 2. Prior probability distributions for each of the stellar parameters defined in Section 3.2, used for the independent (per star) posterior samplings. The stellar parameters here are the "equal evolutionary point" number EEP, age τ, metallicity [Fe/H], extinction A_V, and distance D. The prior on EEP is computed by calculating the stellar mass, m, corresponding to a given point in the isochrone parameter space, $(\text{EEP}, \tau, [\text{Fe/H}])$, and computing the probability from a Salpeter (Salpeter 1955) mass prior such that $p(\text{EEP}) = p(m) \left| \frac{\text{d}m}{\text{dEEP}} \right|$, following the prescription used in `isochrones` (Morton 2015).

We find that the photometry for individual sources in the lower main sequence are very poorly constrained in all parameters, and the prior tends to pull the posterior samplings to prefer closer, older stellar parameters. This is a weakness of our methodology for performing the hierarchical inference: Each source is considered independently, even though there is clear structure in the CMD (i.e. Figure 3), and inferring isochronal parameters for individual lower main sequence stars is a fundamentally degenerate problem. A more correct way to do this would be to infer the stellar parameters of *all* stars, the cluster hyperparameters, and the background simultaneously. However, for the 417 stars we are using (in the red box in Figure 3), this model would have $\sim$2,000 free parameters if left unmarginalized. We are developing tools to perform star cluster parameter inference in this way (Morton et al., in prep.), but here we adopt a simple hack to allow us to instead perform individual posterior samplings and then combine those samplings into a hierarchical inference.

The brightest stars in the cluster have very precise DECam photometry, and have precise astrometry from *Gaia*. Given their location in the CMD, these stars must be young. We therefore use the brightest *kinematic* member of *Price-Whelan 1 — Gaia* source ID 3480046557809199616 — as an "anchor" star: We model every other source in the DECam selection region (i.e. excluding this one) by fitting the photometry of it and the anchor star simultaneously, assuming they have the same age, metallicity, distance, and extinction but different EEP values. Motivated by the possible signature of an (unresolved) binary sequence in the DECam CMD, we add one further piece of complexity to the model by allowing the photometry of each non-anchor source to be fit as an unresolved binary star system. This adds an additional parameter, the unresolved binary mass ratio q, to the list of inferred stellar parameters for each source, but, for this work, we ignore the binary companions and implicitly marginalize over q in what follows. The details of this model are handled by the `isochrones` package.

We generate posterior samplings over the parameters $(\boldsymbol{\theta}, q)$ for each of the 417 sources in the selected region of the DECam CMD using `PyMultinest` (Buchner et al. 2014; Feroz & Hobson 2008; Feroz et al. 2009), and store the value of the prior evaluated at the location of each sample. We then use these samples and interim prior values to construct our hierarchical model, as described below.

In the hierarchical model, we assume that the stellar parameters of each primary star in the selected region are either drawn from the cluster, or a background (stellar halo) population. For the cluster, we assume that the values are drawn from delta functions in age, metallicity, distance, and extinction, with the centroids of the delta functions $\boldsymbol{\alpha} = (\tau^*, [\mathrm{Fe/H}]^*, D^*, A_V^*)$ as hyperparameters of the hierarchical inference—that is, we assume that the cluster is a single stellar population (SSP). We note that we have run the same hierarchical model allowing for Gaussian spreads in age, metallicity, and distance where we simultaneously infer the variances in each cluster parameter, but we have found that the variances are unconstrained, and we therefore instead treat the cluster as an SSP.

For the background model, we assume the same priors as specified in Table 2. The one additional parameter that must be included in this hierarchical model is the mixture weight: The global fraction of sources that are likely cluster members, f. To compute the likelihood for the hierarchical model, we use the individual posterior samplings to marginalize over the per-source stellar parameters $\boldsymbol{\theta}_n$ to compute the marginal likelihood $p(\boldsymbol{m}_n \,|\, \boldsymbol{\alpha}, f)$, where $\boldsymbol{m}_n = (g, i)_n$ is the vector of photometric data for source n; This likelihood for a single source given a set of hyperparameters $(\boldsymbol{\alpha}, f)$ is then

$$p(\boldsymbol{m}_n \,|\, \boldsymbol{\alpha}, f) = \int \mathrm{d}\boldsymbol{\theta}_n \, p(\boldsymbol{m}_n \,|\, \boldsymbol{\theta}_n) \, p(\boldsymbol{\theta}_n \,|\, \boldsymbol{\alpha}, f) \quad . \quad (5)$$

We employ the "importance sampling trick" (see, e.g., Appendix of Price-Whelan et al. 2018 or Hogg et al. 2010; Foreman-Mackey et al. 2014 for other examples) to re-write an approximate form for this marginal likelihood as

$$p(\boldsymbol{m}_n \,|\, \boldsymbol{\alpha}, f) \approx \frac{\mathcal{Z}_n}{K} \sum_k^K \frac{p(\boldsymbol{\theta}_{nk} \,|\, \boldsymbol{\alpha}, f)}{p(\boldsymbol{\theta}_{nk} \,|\, \boldsymbol{\alpha}_0, f)} \quad (6)$$

where the index k specifies the index of one of K posterior samples generated from the independent samplings (described above), $\mathcal{Z}_n$ is a constant, and the denominator, $p(\boldsymbol{\theta}_{nk} \,|\, \boldsymbol{\alpha}_0, f)$, are the values of the interim prior used to do the independent samplings. In this work, $N = 417$ and we adopt $K = 2048$.

With the marginal likelihood (Equation 6), we then need to specify prior probability distributions for the hyperparameters $(\boldsymbol{\alpha}, f)$, and we can then generate posterior samples for the hyperparameters. We use uniform priors for all of these, as summarized in Table 3. We use `emcee` (Foreman-Mackey et al. 2013a; Goodman & Weare 2010) to sample from the posterior probability distribution for the hyperparameters given all of the photometric data,

$$p(\boldsymbol{\alpha}, f \,|\, \{\boldsymbol{m}_n\}) \propto p(\boldsymbol{\alpha}) \, p(f) \prod_n^N p(\boldsymbol{m}_n \,|\, \boldsymbol{\alpha}, f) \quad . \quad (7)$$

Here we use 64 walkers and run for an initial 128 steps to burn-in the sampler before running for a final 1024 steps. We again compute the Gelman-Rubin (Gelman & Rubin 1992) convergence diagnostic and find that all chains have $R < 1.1$ and are thus likely converged. Figure 5 shows a corner plot with all 1D and 2D marginal posterior probability distributions estimated from the samples.

4. RESULTS

4.1. *Stellar population and physical characteristics*

We use the posterior samples from the hierarchical inference described in Section 3.2 to compute the posterior probabilities that each source in the DECam CMD selection box (red box, Figure 3) is a member of the cluster. Figure 6 shows the DECam photometry for all sources with membership probability > 0.5 (circle

Parameter	Prior	Bounds
τ^*	uniform	$(1\,\mathrm{Myr}, 1\,\mathrm{Gyr})$
$[\mathrm{Fe/H}]^*$	uniform	$(-2, 0)$
A_V^*	uniform	$(0, 1)$ mag
D^*	uniform	$(1, 100)$ kpc
f	uniform	$(0, 1)$

Table 3. Prior probability distributions for the hyperparameters $(\boldsymbol{\alpha}, f)$, used for the hierarchical inference of the cluster stellar population parameters. The cluster parameters here are the cluster age τ^*, metallicity $[\mathrm{Fe/H}]^*$, extinction A_V^*, and distance D^*, and the fraction of stars in the field that belong to the cluster f.
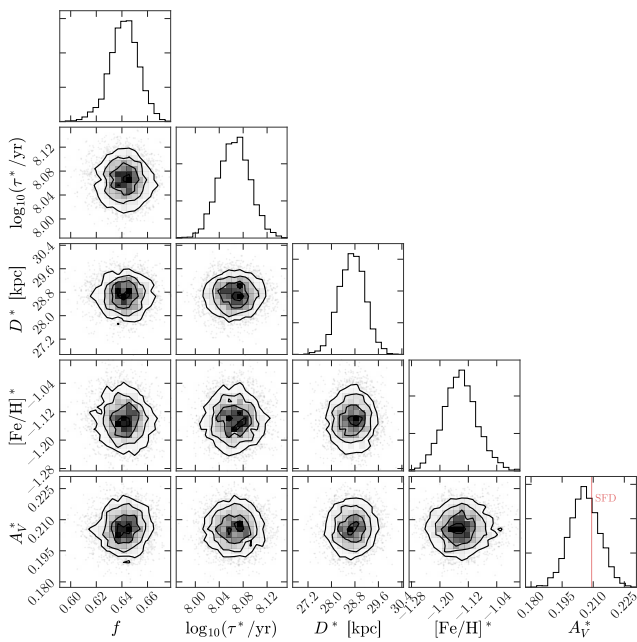
Figure 5. A corner plot showing marginal posterior probability distributions estimated from the the posterior samples generated from the hierarchical inference of the cluster population parameters.

markers). Over-plotted in Figure 6 (blue, solid line) is a MIST isochrone with the median posterior parameters derived from the hierarchical modeling, shifted to the median distance and extincted given the median A_V value. Under the assumption that *Price-Whelan 1* has a single stellar population, we find that *Price-Whelan 1* is indeed young, distant, and metal poor, with median posterior values and standard deviations of age $\tau = 117 \pm 23$ Myr, distance $D = 28.7 \pm 0.5$ kpc, and metallicity $[\mathrm{Fe/H}] = -1.14 \pm 0.05$ (see also Table 4 for a summary). We find that the inferred extinction, $A_V = 0.205 \pm 0.011$, is consistent with the value $A_{V,\mathrm{SFD}} \approx 0.2$ from the (recalibrated) SFD dust map

DECam data & isochrone models
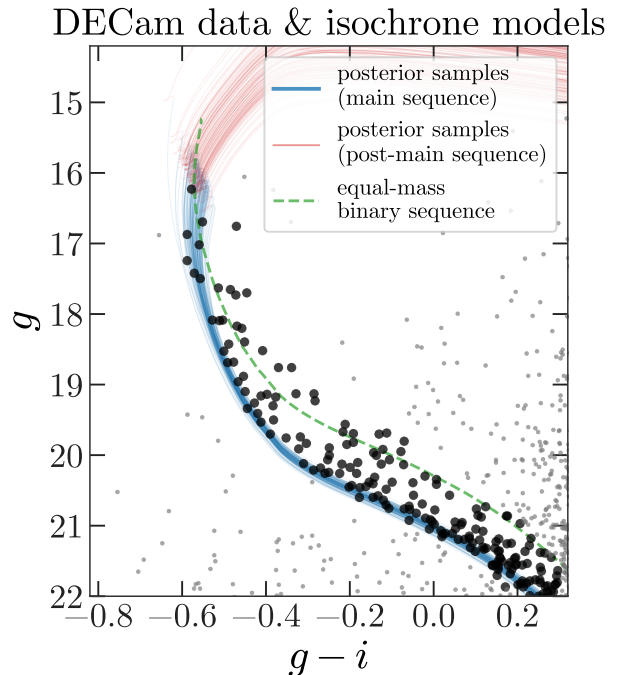
Figure 6. The DECam color-magnitude diagram for sources with probability > 0.5 of belonging to *Price-Whelan 1* (dark, larger markers), and sources with probability < 0.5 (lighter, smaller markers). The solid line (blue) shows the MIST isochrone for the median posterior sample from the hierarchical inference of the cluster population parameters, shifted to the inferred distance to *Price-Whelan 1*, and extincted with the inferred extinction. The dashed line (green) shows the same isochrone, shifted ≈ 0.75 magnitudes brighter, representing the expected location of the equal-mass binary sequence of the cluster.

(Schlegel et al. 1998; Schlafly & Finkbeiner 2011). We note again that, with the DECam photometry, we are limited to studying the subcomponent a of *Price-Whelan 1* — follow-up imaging of component b (see Figure 1) would enable a similar analysis for the full cluster.

Also plotted in Figure 6 is the equal-mass binary sequence (green, dashed line) computed from the median posterior sample: The abundance of sources between the nominal isochrone and the binary sequence highlights the fact that the cluster may contain a significant number of binary or multiple star systems, but we leave a detailed study of multiplicity to future work. We note that, similar to young open clusters in the Milky Way disk (e.g., Gaia Collaboration et al. 2018b), the bluest stars in the CMD have more apparent scatter than the lower main sequence. This is likely a combination of many things that are not addressed in this work, such as binarity, convective core overshooting (e.g., Yang & Tian 2017; Johnston et al. 2019), or an intrinsic spread in stellar parameters.

We use the isochrone corresponding to the median posterior sample to estimate the total stellar mass of the cluster. By assuming that the DECam imaging is 100% complete to stars with $(g - i) < 0.3$ and $g < 22$, and by assuming a Kroupa initial mass function (Kroupa 2001), we use the number of observed stars and the isochrone to compute the total mass, $M_{\text{tot},*} \approx 1200 \, M_\odot$.

The mass and age of *Price-Whelan 1* are comparable to Milky Way disk open clusters, but with a much lower metallicity, and a much larger spatial extent. For example, the Pleiades has an age ~ 135 Myr (Gossage et al. 2018), but a physical size ~ 5 pc. At a distance of 28.7 kpc, *Price-Whelan 1* spans $\sim 1.5°$, corresponding to a physical size ~ 700–800 pc. This is more comparable to (but still larger than) recent star formation sites in the Magellanic bridge (e.g., Mackey et al. 2017), which likely formed as a result of the violent interaction between the MCs. If *Price-Whelan 1* formed unbound, but with an initial size comparable to the present size of the Pleiades, this corresponds to an expansion velocity ~ 6 km s^{-1}, which would be detectable with precise radial velocity measurements of stars on either side of *Price-Whelan 1*.

Inferred properties of *Price-Whelan 1*

Name	Value	Description
$\langle \alpha \rangle$	$178.8°$	right ascension
$\langle \delta \rangle$	$-29.4°$	declination
$\langle D \rangle$	28.7 ± 0.4 kpc	Heliocentric distance
$\langle \mu_\alpha \rangle$	-0.56 ± 0.04 mas yr^{-1}	proper motion in RA
$\langle \mu_\delta \rangle$	0.47 ± 0.02 mas yr^{-1}	proper motion in Dec
τ	116 ± 7 Myr	age
M	$1200 \, M_\odot$	total stellar mass
[Fe/H]	-1.14 ± 0.05	metallicity
A_V	0.21 ± 0.02	extinction

Table 4. This table summarizes the measured or inferred kinematic and stellar population parameters of the young halo stellar cluster *Price-Whelan 1*. Formal precisions on all inferred parameters are very small, typically one to a few per cent, but we suspect that systematic errors with the photometry and isochrone models limit the accuracy of these measurements to $\sim 5\%$.

4.2. *Relation to the Magellanic stream*

At the sky location and distance of *Price-Whelan 1*, i.e. well into the Galactic halo, the only plausible gas reservoirs that could have formed a young cluster are the MS, or a previously unknown high velocity cloud (HVC). HVCs are thought to either be accreted and therefore lower metallicity than typical present-day Milky Way
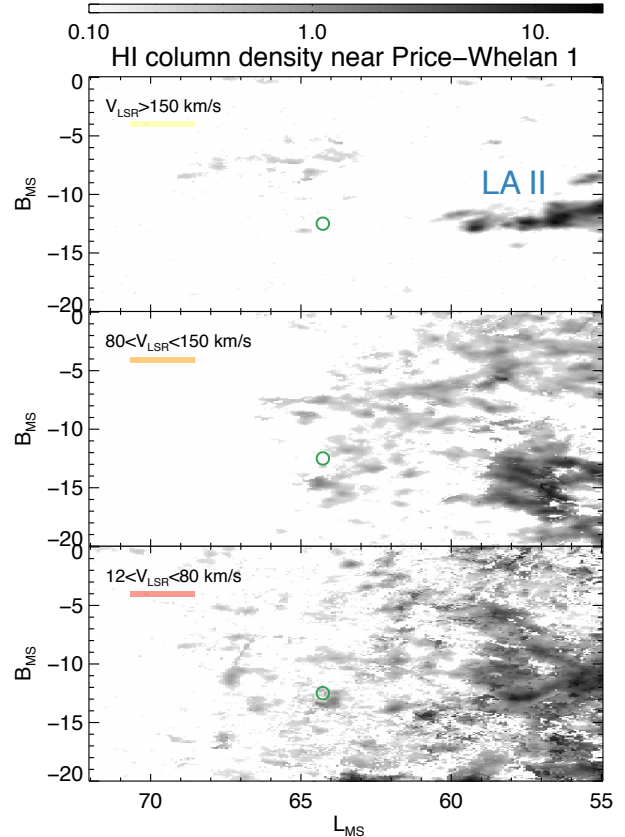
Figure 7. GASS HI column density in the region of LA II, shown in three different velocity slices (corresponding to and indicated in each panel). The colored rectangles under each velocity label correspond to the colored ranges in Figure 8. The coordinate system is the Magellanic stream spherical coordinate system defined in Nidever et al. (2008), and the units of L_{MS} and B_{MS} are degrees. The horizontal colorbar shows the column density of HI in units of 10^{19} atoms cm^{-2}. The open circle (green) marks the position of *Price-Whelan 1*, and the feature commonly attributed to the LA is indicated in the top panel.

gas, or ejected from the Milky Way through a "Galactic fountain"-like process and therefore comparable metallicity to disk gas. Both processes clearly occur: the MS itself is evidence of gas accretion into the Milky Way halo, and the mysterious Smith Cloud (Smith 1963) is a metal-rich ([Fe/H] ~ 0.5 Fox et al. 2016) HVC that plausibly originated from the Galactic disk (e.g., Bregman 1980). Given the low metallicity of the stars in *Price-Whelan 1*, the gas it formed from was likely extragalactic, as any violent star-forming regions in the Milky Way disk that could have driven gas so far out into the halo have, at present-day, significantly higher metallicity than this cluster.

We therefore take a closer look at the HI gas in the vicinity of *Price-Whelan 1* to assess its possible
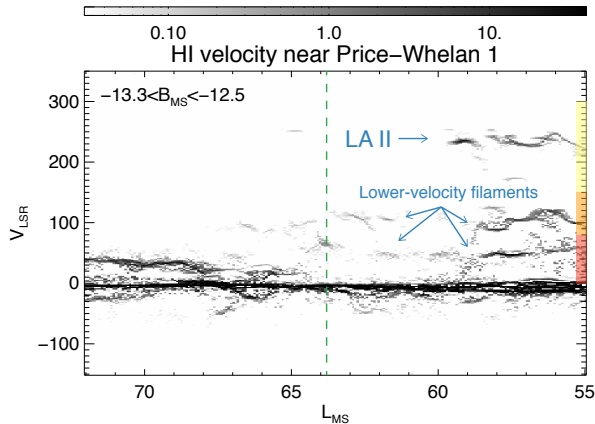
Figure 8. The position-velocity diagram of GASS HI Gaussian centers in a slice of Magellanic latitudes (indicated) in the same region as Figure 7, showing the line-of-sight velocity (in the local standard of rest frame) of gas near *Price-Whelan 1*. The colored rectangles (red, orange, yellow) indicate the three velocity slices used in Figure 7. Here the colorbar shows the gas density in units of 10^{18} atoms cm^{-2} deg. Note that the densest gas at the longitude of *Price-Whelan 1* (vertical line) appears to be associated with the lower velocity filaments seen at lower longitudes $L_{MS} \lesssim 60°$, which may be decelerated gas associated with the Magellanic stream.

relation to the MS. While no bulk metallicity measurements exist for the LA gas, recent measurements of oxygen and silicon abundances ([O/H] ~ -1) suggest that the gas has a comparable abundance pattern to the SMC gas (Fox et al. 2013, 2018; Richter et al. 2018). If *Price-Whelan 1* formed from this gas, we would then expect that the stellar metallicity should be similar to young stars in the SMC. In the outskirts of the SMC, the metallicity distribution of the youngest stars (age < 2 Gyr) has a peak around [Fe/H] ~ -1 and a spread of $\sigma_{[Fe/H]} \sim 0.2$ dex, which at least means that the metallicity of *Price-Whelan 1* ([Fe/H] ≈ -1.1) is consistent with young SMC stars (Dobbie et al. 2014).

We next look at the kinematics of high velocity gas in this region. Figure 7 shows the HI column density of the catalog of Gaussian centers (produced with the techniques and software from Nidever et al. 2008) from the Galactic All Sky Survey (GASS; McClure-Griffiths et al. 2009; Kalberla et al. 2010) in the region around the LA and plotted in the MS coordinate system (L_{MS}, B_{MS}; Nidever et al. 2008). The location of *Price-Whelan 1* is marked (green circle), and the three panels show three different velocity slices (indicated on each panel). The MS gas is most visible and densest at the highest velocities where the LA II feature is found, around $L_{MS} \lesssim 61°$ and $B_{MS} \sim -12°$, but the lower panels (i.e. slower gas) still displays some of the same morphology as the highest velocity slice.

This is also apparent in a position-velocity diagram of the same range of longitudes: Figure 8 shows the velocity of gas in this region (all longitudes, but near $B_{MS} \sim -13°$) as a function of L_{MS}. Here, the longitude of *Price-Whelan 1* is marked as the dashed vertical (green) line. The LA II feature is again visible as the $V_{LSR} > 200$ km s^{-1} filament of gas at $L_{MS} \lesssim 60°$ (and a faint extension near $L_{MS} \sim 65°$), but two other prominent filaments with similar morphologies are apparent at lower velocities $V_{LSR} \sim 60$ km s^{-1} and $V_{LSR} \sim 100$ km s^{-1} from $L_{MS} \lesssim 65°$.

The entire region shown in Figure 7 is on the other side of the Galactic disk with respect to the Magellanic clouds, meaning that all of the MS gas in this region has passed through the Galactic midplane. The similar filamentary structure in the high and lower velocity gas therefore raises the question: Did the low velocity filaments originate in the MS, but were decelerated through interactions with Milky Way gas? This scenario has been suggested (i.e., that the LA material has interacted with gas in the Galactic disk) based on the kinematics and morphology of a LA HVC (McClure-Griffiths et al. 2008) and presents a possible formation scenario for *Price-Whelan 1*. If the cluster formed from the LA gas, and some of the gas interacted with the Milky Way disk, the dynamics of the stars and the gas that interacted would have decoupled as soon as the stars formed. Given the geometry of the current location of *Price-Whelan 1* (i.e. our viewing angle from the solar position), in this scenario we would expect a moderate velocity difference between the gas that interacted and stars to primarily appear as a difference in the tangential velocities of the stars and the gas (which is not measured). We can therefore use the line-of-sight velocity of the gas as a proxy for the line-of-sight velocity of the stars in *Price-Whelan 1* to study the Galactic orbit of the cluster, as discussed in the next section, and because we do not yet have radial velocity information for the stars.

4.3. *The Galactic orbit of Price-Whelan 1*

Since the gas distribution in Figure 8 shows excess at velocities $V_{LSR} \approx (60, 110, 230)$ km s^{-1} near the location of the cluster (ignoring the disk gas between $V_{LSR} \sim -20$–20 km s^{-1}), it seems that there are three qualitatively different line-of-sight velocities that the gas (and plausibly *Price-Whelan 1*) could have. If we adopt these velocity measurements as the line-of-sight (LOS) velocity of *Price-Whelan 1*, we have measurements of all 6 phase-space components and can compute orbits for the cluster. We compute Galactic orbits of the cluster using each of the three values of the line-of-sight veloc-
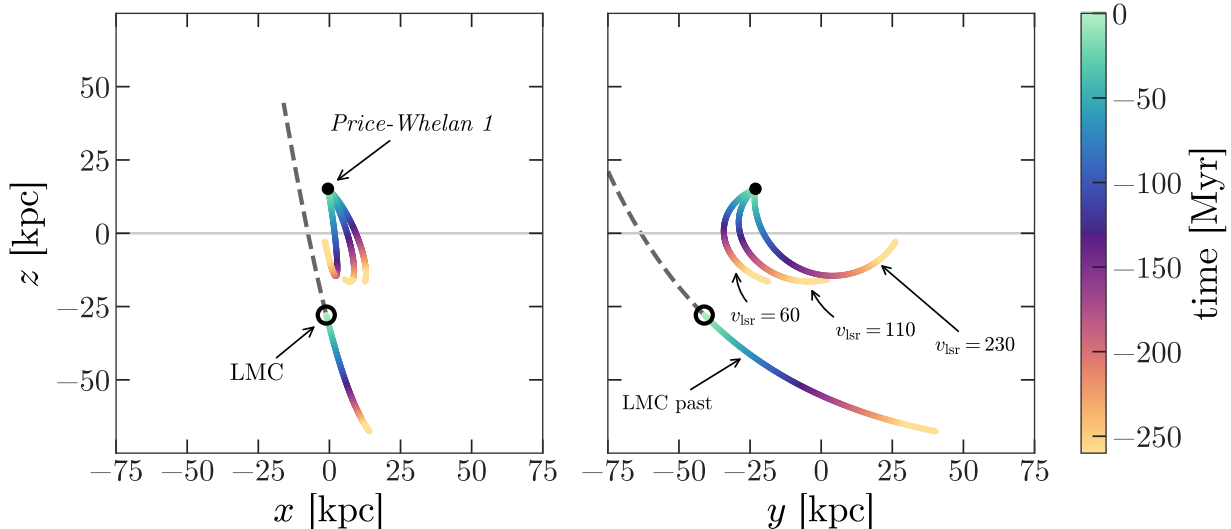
Figure 9. The past orbit of *Price-Whelan 1* in Galactocentric Cartesian coordinates, computed by assuming a line-of-sight velocity equal to each of three different possible gas filaments in the LA II region of the MS (as indicated in right panel). In these coordinates, the Sun is at $(x, y, z) = (-8.1, 0, 0)$ kpc, and the present-day locations of *Price-Whelan 1* (filled black marker) and the LMC (open black marker) are indicated. The time at each point in the orbits is indicated by the colorbar, with the darkest regions corresponding to $t = -130$ Myr, the inferred age of *Price-Whelan 1*. The past and future orbit of the LMC are shown as the colored line and dashed line, respectively.

ity from the LA II region filaments. Figure 9 shows these three orbits computed by integrating the position of *Price-Whelan 1* backwards in time: The two panels of Figure 9 show the Galactocentric Cartesian trajectories of *Price-Whelan 1* over the last 260 Myr by assuming each of the three plausible line-of-sight velocities discussed above (as labeled in the right panel). To compute the orbits, we use a three-component Milky Way mass model consisting of a disk (Miyamoto & Nagai 1975), a bulge (Hernquist 1990), and a spherical dark matter halo (Navarro et al. 1996). The parameters of this model are set to match the circular velocity profile and disk properties of Bovy (2015), and is implemented in the `gala` package (Price-Whelan 2017) as the `MilkyWayPotential`.

We tried including an LMC component as a Hernquist sphere (Hernquist 1990) with masses between 10^{10}–2.5×10^{11} M$_\odot$, but find that the orbit of *Price-Whelan 1* does not change much over the integration period. If the LMC mass were as high as 8×10^{11} M$_\odot$, all three orbits of *Price-Whelan 1* return to and end up bound to the LMC, but at this large of a mass, the assumption of a fixed Milky Way reference frame is invalid. We therefore neglect LMC–SMC component in what follows, and present an alternate interpretation below.

Also shown in Figure 9 is the present-day location of the LMC (open black circle), the present-day location of *Price-Whelan 1* (filled black circle), the past orbit of the LMC over the same period, and the projected future orbit of the LMC (dashed line), all excluding the

LMC–SMC as a mass component. For the two lower LOS velocity cases, the orbit of *Price-Whelan 1* crosses the midplane of the Galactic disk at a time comparable to the age of the cluster (i.e. when the color is darkest, corresponding to a time ≈ -130 Myr). However, while the orbits of *Price-Whelan 1* and the LMC are generally in the same orbital plane and have the same sense about the Galaxy, none of the orbits closely approach (within the half-mass radius of our Hernquist model for the LMC) or cross the orbit of the LMC (even when including the LMC mass component). This is expected if the gas was shocked or otherwise lost kinetic energy through, e.g., ram pressure — as is expected to explain the morphology of the LA material (e.g., Hammer et al. 2015) — during or before the star formation event that produced *Price-Whelan 1*. We therefore posit that the orbit of *Price-Whelan 1* can be explained as having originated from the LMC–SMC system with some loss of orbital energy during the star formation event.

Precise radial velocity measurements of stars in *Price-Whelan 1* will definitively rule on this scenario. For one, with a precision $\sigma_v \lesssim 2$ km s^{-1}, the internal kinematics of *Price-Whelan 1* could be resolved, and in particular the expansion rate could be measured. But even low-precision velocity measurements would either confirm the association between *Price-Whelan 1* and one of the three LA II-region gas filaments, or would challenge this scenario.

4.4. *Other possible formation scenarios*

While the spatial proximity of *Price-Whelan 1* to LA gas in the Magellanic Stream suggests a tentative connection, radial velocity data for the stars is required to confirm that the velocity of the cluster is consistent with the gas and thereby confirm the connection (Nidever et al., in prep.). However, even if the cluster can be confidently, kinematically associated with the LA gas, a number of mysteries would remain. First, it is not clear what could have triggered this star formation event, or why this appears to be the only young cluster associated with the LA gas. Second, the origin of the LA gas itself is not well understood, and some studies have questioned its connection to the Magellanic Clouds (Tepper-García et al. 2019).

Other possible formation scenarios for *Price-Whelan 1* do not help to resolve these issues. One possible alternative scenario is that the *Price-Whelan 1* birth cloud was unassociated with the MS and instead a lone HVC. However, this still raises questions about what triggered the star formation, given that other HVCs in the Milky Way are observed to be devoid of stars (e.g., Stark et al. 2015). Another possible origin for *Price-Whelan 1* is that it formed from gas stripped from a fully-destroyed dwarf galaxy (an unknown satellite of the Magellanic Clouds). However, we have not found a significant over-density of comoving older stars associated with the LA in this region.

One final possibility is that these stars were formed in or associated with the Galactic disk. While there have been some claimed discoveries of young, embedded clusters at high Galactic latitude (within $\lesssim 5$ kpc of the sun, e.g., Camargo et al. 2015, 2016, these may not be real (Turner et al. 2017). Still, stars formed in the Galactic disk have been observed far from the disk midplane with heights up to $|z| \sim 10$ kpc (e.g., the TriAnd over-density; Price-Whelan et al. 2015; Bergemann et al. 2018)). However, TriAnd has an old stellar population, and no young disk stars have been since found associated with this or other features in the outer Galactic disk (Deason et al. 2018; Laporte et al. 2019).

5. SUMMARY

We have identified a young, metal-poor stellar association in the Galactic halo — named *Price-Whelan 1* — with an age $\tau \approx 117$ Myr, heliocentric distance $D \approx 28.7$ kpc, and metallicity [Fe/H] ≈ -1.14. At its present-day sky position, and at large distances, all significant quantities of H I are associated with the leading arm of the Magellanic stream, and thus it plausibly formed from this gas. The age of the cluster is broadly consistent with the time it would have most recently crossed the Galactic midplane, suggesting the possibility

that interaction with the Milky Way disk or tidal compression could have triggered this star formation event. Previous studies have detected young stars in the leading arm and in the periphery of the LMC (Casetti-Dinescu et al. 2014; Moni Bidin et al. 2017), however, this is the first time that an entire young star cluster has been detected so far from the Clouds.

The discovery of *Price-Whelan 1* provides a critical distance constraint to the leading arm Magellanic stream and will aid future Magellanic system and Milky Way modeling efforts. It also provides an opportunity to study star formation in a unique environment (i.e. low gas density, low metallicity, and large velocity), unlike that of the Milky Way disk or any other local cluster-forming region. The serendipitous discovery of this cluster is a reminder that the combined value of the *Gaia* data with deep, large-area imaging surveys provides a wealth of information about our Galaxy and stellar halo.

APPENDIX

A. QUERIES

Initial query to select very blue stars away from the Galactic plane:

```
SELECT * FROM gaiadr2.gaia_source
WHERE parallax < 1
AND (bp_rp > -0.5) AND (bp_rp < 0)
AND phot_g_mean_mag < 20
AND ABS(b) > 20
```

Query to retrieve *Gaia* data around the blue, comoving group found and discussed in Section 2:

```
SELECT *
FROM gaiadr2.gaia_source
WHERE (parallax < 1 OR parallax IS NULL)
    AND ra > 177 AND ra < 182
    AND dec > -31.3 AND dec < -26.3
```

B. DATA

Catalog of *Gaia* sources and kinematic probability

source_id	ra	dec	pmra	pmdec	G0	BP0	RP0	member_prob
3462048793708019200	184.799	-34.971	0.09	-1.61	19.79	19.915	19.61	0.00
3462056696448285440	184.511	-34.988	-4.24	-1.32	20.80	20.498	20.31	0.09
3462062670747949312	184.426	-34.816	-0.56	-2.77	20.33	20.363	20.06	0.00
3462062709402484224	184.407	-34.809	-7.50	-3.16	20.67	20.437	20.09	0.01
3462065904858152832	184.705	-34.783	0.91	-1.95	20.64	20.625	20.37	0.12

...

(5983 rows)

Table 5. Select *Gaia* DR2 data and membership probabilities computed for all stars in the cluster region (see Figure 4).

Catalog of DECam photometry for point sources

name	ra	dec	g	g_err	i	i_err	control_mask	cluster_mask	member_mask
J115311.06-283133.4	178.296	-28.525	24.09	0.46	23.99	0.33	True	False	False
J115403.33-283553.8	178.513	-28.598	25.27	0.29	22.70	0.04	True	False	False
J115424.21-283749.6	178.600	-28.630	24.61	0.25	23.83	0.28	True	False	False
J115413.93-283343.6	178.558	-28.562	25.37	0.41	22.76	0.05	True	False	False
J115419.62-283635.8	178.581	-28.609	24.20	0.13	21.89	0.02	True	False	False

...

(51932 rows)

Table 6. DECam photometry and sky positions for point sources in the observed DECam field. The boolean columns control_mask and cluster_mask are True when a given source is in the control or cluster CCDs, respectively (see Figure 2). The boolean column member_mask is True when the source has a photometric membership probability > 0.5 of belonging to the cluster stellar population (see Section 3.2).

Facility: Blanco (DECam)

Software: *Astropy* (Astropy Collaboration et al. 2013, 2018), *dustmaps* (Green 2018), *emcee* (Foreman-Mackey et al. 2013a,b), *gala* (Price-Whelan 2017), *IPython* (Pérez & Granger 2007), *isochrones* (Morton 2015), *matplotlib* (Hunter 2007),

numpy (Walt et al. 2011), *PyMultinest* (Buchner et al. 2014), *schwimmbad* (Price-Whelan & Foreman-Mackey 2017), *scipy* (Jones et al. 2001–)

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipócz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bekki, K., Chiba, M., & McClure-Griffiths, N. M. 2008, ApJ, 672, L17, doi: 10.1086/526456

Belokurov, V. A., & Erkal, D. 2019, MNRAS, 482, L9, doi: 10.1093/mnrasl/sly178

Bergemann, M., Sesar, B., Cohen, J. G., et al. 2018, Nature, 555, 334, doi: 10.1038/nature25490

Besla, G., Kallivayalil, N., Hernquist, L., et al. 2007, ApJ, 668, 949, doi: 10.1086/521385

—. 2012, MNRAS, 421, 2109, doi: 10.1111/j.1365-2966.2012.20466.x

—. 2010, ApJ, 721, L97, doi: 10.1088/2041-8205/721/2/L97

Bovy, J. 2015, The Astrophysical Journal Supplement Series, 216, doi: 10.1088/0067-0049/216/2/29

Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, Ann. Appl. Stat., 5, 1657, doi: 10.1214/10-AOAS439

Bregman, J. N. 1980, ApJ, 236, 577, doi: 10.1086/157776

Brüns, C., Kerp, J., Staveley-Smith, L., et al. 2005, A&A, 432, 45, doi: 10.1051/0004-6361:20040321

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125, doi: 10.1051/0004-6361/201322971

Burton, W. B., & Lockman, F. J. 1999, A&A, 349, 7. https://arxiv.org/abs/astro-ph/9908012

Camargo, D., Bica, E., & Bonatto, C. 2016, A&A, 593, A95, doi: 10.1051/0004-6361/201628710

Camargo, D., Bica, E., Bonatto, C., & Salerno, G. 2015, MNRAS, 448, 1930, doi: 10.1093/mnras/stv092

Carrera, R., Conn, B. C., Noël, N. E. D., Read, J. I., & López Sánchez, Á. R. 2017, MNRAS, 471, 4571, doi: 10.1093/mnras/stx1932

Casetti-Dinescu, D. I., Moni Bidin, C., Girard, T. M., et al. 2014, ApJL, 784, L37, doi: 10.1088/2041-8205/784/2/L37

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints. https://arxiv.org/abs/1612.05560

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102, doi: 10.3847/0004-637X/823/2/102

Choi, Y., Nidever, D. L., Olsen, K., et al. 2018a, ApJ, 866, 90, doi: 10.3847/1538-4357/aae083

—. 2018b, ArXiv e-prints. https://arxiv.org/abs/1805.00481

Danielski, C., Babusiaux, C., Ruiz-Dern, L., Sartoretti, P., & Arenou, F. 2018, A&A, 614, A19, doi: 10.1051/0004-6361/201732327

Deason, A. J., Belokurov, V., & Koposov, S. E. 2018, Monthly Notices of the Royal Astronomical Society, 473, 2428, doi: 10.1093/mnras/stx2528

Deason, A. J., Belokurov, V., & Weisz, D. R. 2015, MNRAS, 448, L77, doi: 10.1093/mnrasl/slv001

Diaz, J. D., & Bekki, K. 2012, ApJ, 750, 36, doi: 10.1088/0004-637X/750/1/36

Dobbie, P. D., Cole, A. A., Subramaniam, A., & Keller, S. 2014, MNRAS, 442, 1680, doi: 10.1093/mnras/stu926

D'Onghia, E., & Fox, A. J. 2016, Annual Review of Astronomy and Astrophysics, 54, 363, doi: 10.1146/annurev-astro-081915-023251

Dotter, A. 2016, ApJS, 222, 8, doi: 10.3847/0067-0049/222/1/8

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449, doi: 10.1111/j.1365-2966.2007.12353.x

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601, doi: 10.1111/j.1365-2966.2009.14548.x

Fiorentino, G., Bono, G., Monelli, M., et al. 2015, ApJ, 798, L12, doi: 10.1088/2041-8205/798/1/L12

For, B. Q., Staveley-Smith, L., Matthews, D., & McClure-Griffiths, N. M. 2014, ApJ, 792, 43, doi: 10.1088/0004-637X/792/1/43

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013a, Publications of the Astronomical Society of the Pacific, 125, 306, doi: 10.1086/670067

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64, doi: 10.1088/0004-637X/795/1/64

Foreman-Mackey, D., Conley, A., Meierjurgen Farr, W., et al. 2013b, emcee: The MCMC Hammer, Astrophysics Source Code Library. http://ascl.net/1303.002

Fox, A. J., Richter, P., Wakker, B. P., et al. 2013, ApJ, 772, 110, doi: 10.1088/0004-637X/772/2/110

Fox, A. J., Lehner, N., Lockman, F. J., et al. 2016, ApJ, 816, L11, doi: 10.3847/2041-8205/816/1/L11

Fox, A. J., Barger, K. A., Wakker, B. P., et al. 2018, ApJ, 854, 142, doi: 10.3847/1538-4357/aaa9bb

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, ArXiv e-prints. https://arxiv.org/abs/1804.09365

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018b, A&A, 616, A10, doi: 10.1051/0004-6361/201832843

Gelman, A., & Rubin, D. B. 1992, Statistical Science, 7, 457, doi: 10.1214/ss/1177011136

Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65, doi: 10.2140/camcos.2010.5.65

Gossage, S., Conroy, C., Dotter, A., et al. 2018, ApJ, 863, 67, doi: 10.3847/1538-4357/aad0a0

Green, G. 2018, The Journal of Open Source Software, 3, 695, doi: 10.21105/joss.00695

Guhathakurta, P., & Reitzel, D. B. 1998, in Galactic Halos, Vol. 136, 22

Hammer, F., Yang, Y. B., Flores, H., Puech, M., & Fouquet, S. 2015, ApJ, 813, 110, doi: 10.1088/0004-637X/813/2/110

Harris, W. E. 1996, AJ, 112, 1487, doi: 10.1086/118116

Hernquist, L. 1990, ApJ, 356, 359, doi: 10.1086/168845

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166, doi: 10.1088/0004-637X/725/2/2166

Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Ibata, R. A., Gilmore, G., & Irwin, M. J. 1994, Nature, 370, 194, doi: 10.1038/370194a0

Johnston, C., Tkachenko, A., Aerts, C., et al. 2019, MNRAS, 482, 1231, doi: 10.1093/mnras/sty2671

Jones, E., Oliphant, T., Peterson, P., et al. 2001–, SciPy: Open source scientific tools for Python. http://www.scipy.org/

Kalberla, P. M. W., McClure-Griffiths, N. M., Pisano, D. J., et al. 2010, A&A, 521, A17, doi: 10.1051/0004-6361/200913979

Kallivayalil, N., van der Marel, R. P., Alcock, C., et al. 2006, ApJ, 638, 772, doi: 10.1086/498972

Kallivayalil, N., van der Marel, R. P., Besla, G., Anderson, J., & Alcock, C. 2013, ApJ, 764, 161, doi: 10.1088/0004-637X/764/2/161

Kroupa, P. 2001, MNRAS, 322, 231, doi: 10.1046/j.1365-8711.2001.04022.x

Laporte, C. F. P., Belokurov, V., Koposov, S. E., Smith, M. C., & Hill, V. 2019, arXiv e-prints, arXiv:1907.10678. https://arxiv.org/abs/1907.10678

Lindegren, L., Hernandez, J., Bombrun, A., et al. 2018, ArXiv e-prints. https://arxiv.org/abs/1804.09366

Mackey, A. D., Koposov, S. E., Da Costa, G. S., et al. 2017, MNRAS, 472, 2975, doi: 10.1093/mnras/stx2035

Mackey, A. D., Koposov, S. E., Erkal, D., et al. 2016, MNRAS, 459, 239, doi: 10.1093/mnras/stw497

Maíz Apellániz, J., & Weiler, M. 2018, A&A, 619, A180, doi: 10.1051/0004-6361/201834051

Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, ApJ, 599, 1082, doi: 10.1086/379504

Mathewson, D. S., Cleary, M. N., & Murray, J. D. 1974, ApJ, 190, 291, doi: 10.1086/152875

Mayer, L., Mastropietro, C., Wadsley, J., Stadel, J., & Moore, B. 2006, MNRAS, 369, 1021, doi: 10.1111/j.1365-2966.2006.10403.x

McClure-Griffiths, N. M., Staveley-Smith, L., Lockman, F. J., et al. 2008, ApJ, 673, L143, doi: 10.1086/528683

McClure-Griffiths, N. M., Pisano, D. J., Calabretta, M. R., et al. 2009, ApJS, 181, 398, doi: 10.1088/0067-0049/181/2/398

McConnachie, A. W. 2012, AJ, 144, 4, doi: 10.1088/0004-6256/144/1/4

Miyamoto, M., & Nagai, R. 1975, PASJ, 27, 533

Moni Bidin, C., Casetti-Dinescu, D. I., Girard, T. M., et al. 2017, MNRAS, 466, 3077, doi: 10.1093/mnras/stw3242

Morton, T. D. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library. http://ascl.net/1503.010

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563, doi: 10.1086/177173

Nidever, D. L., Majewski, S. R., & Butler Burton, W. 2008, ApJ, 679, 432, doi: 10.1086/587042

Nidever, D. L., Majewski, S. R., Butler Burton, W., & Nigra, L. 2010, ApJ, 723, 1618, doi: 10.1088/0004-637X/723/2/1618

Nidever, D. L., Olsen, K., Walker, A. R., et al. 2017, AJ, 154, 199, doi: 10.3847/1538-3881/aa8d1c

Noël, N. E. D., Conn, B. C., Carrera, R., et al. 2013, ApJ, 768, 109, doi: 10.1088/0004-637X/768/2/109

Olsen, K. A. G., Zaritsky, D., Blum, R. D., Boyer, M. L., & Gordon, K. D. 2011, ApJ, 737, 29, doi: 10.1088/0004-637X/737/1/29

Pardy, S. A., D'Onghia, E., & Fox, A. J. 2018, ApJ, 857, 101, doi: 10.3847/1538-4357/aab95b

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3, doi: 10.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4, doi: 10.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15, doi: 10.1088/0067-0049/220/1/15

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21, doi: 10.1109/MCSE.2007.53

Price-Whelan, A. M. 2017, The Journal of Open Source Software, 2, 388, doi: 10.21105/joss.00388

Price-Whelan, A. M., & Foreman-Mackey, D. 2017, The
    Journal of Open Source Software, 2, 357,
    doi: 10.21105/joss.00357

Price-Whelan, A. M., Johnston, K. V., Sheffield, A. A.,
    Laporte, C. F. P., & Sesar, B. 2015, Monthly Notices of
    the Royal Astronomical Society, 452, 676,
    doi: 10.1093/mnras/stv1324

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2018,
    AJ, 156, 18, doi: 10.3847/1538-3881/aac387

Putman, M. E., Gibson, B. K., Staveley-Smith, L., et al.
    1998, Nature, 394, 752, doi: 10.1038/29466

Richter, P., Fox, A. J., Wakker, B. P., et al. 2018, ArXiv
    e-prints.   https://arxiv.org/abs/1808.09455

Salpeter, E. E. 1955, ApJ, 121, 161, doi: 10.1086/145971

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103,
    doi: 10.1088/0004-637X/737/2/103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ,
    500, 525, doi: 10.1086/305772

Smith, G. P. 1963, Bulletin of the Astronomical Institutes
    of the Netherlands, 17, 203

Stark, D. V., Baker, A. D., & Kannappan, S. J. 2015,
    MNRAS, 446, 1855, doi: 10.1093/mnras/stu2182

Stetson, P. B. 1987, PASP, 99, 191, doi: 10.1086/131977

—. 1994, PASP, 106, 250, doi: 10.1086/133378

Tepper-García, T., & Bland-Hawthorn, J. 2018, MNRAS,
    478, 5263, doi: 10.1093/mnras/sty1359

Tepper-García, T., Bland-Hawthorn, J., Pawlowski, M. S.,
    & Fritz, T. K. 2019, MNRAS, 1575,
    doi: 10.1093/mnras/stz1659

Turner, D. G., Carraro, G., & Panko, E. A. 2017, MNRAS,
    470, 481, doi: 10.1093/mnras/stx1258

Valdes, F., Gruendl, R., & DES Project. 2014, in
    Astronomical Society of the Pacific Conference Series,
    Vol. 485, Astronomical Data Analysis Software and
    Systems XXIII, ed. N. Manset & P. Forshay, 379

Venzmer, M. S., Kerp, J., & Kalberla, P. M. W. 2012,
    A&A, 547, A12, doi: 10.1051/0004-6361/201118677

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011,
    Computing in Science and Engg., 13, 22,
    doi: 10.1109/MCSE.2011.37

Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, PASA,
    35, e010, doi: 10.1017/pasa.2018.5

Yang, W., & Tian, Z. 2017, ApJ, 836, 102,
    doi: 10.3847/1538-4357/aa5b9d

Zhang, L., Moni Bidin, C., Casetti-Dinescu, D. I., et al.
    2017, ApJ, 835, 285, doi: 10.3847/1538-4357/835/2/285

Zivick, P., Kallivayalil, N., van der Marel, R. P., et al. 2018,
    ApJ, 864, 55, doi: 10.3847/1538-4357/aad4b0