



# Host-star Properties of Hot, Warm, and Cold Jupiters in the Solar Neighborhood from Gaia Data Release 3: Clues to Formation Pathways

Bihan Banerjee<sup>1</sup>, Mayank Narang<sup>1,2</sup>, P. Manoj<sup>1</sup>, Thomas Henning<sup>3</sup>, Himanshu Tyagi<sup>1</sup>, Arun Surya<sup>1,4</sup>,  
Prasanta K. Nayak<sup>1,5</sup>, and Mihir Tripathi<sup>1,2</sup>

<sup>1</sup>Tata Institute of Fundamental Research, Mumbai 400005, India; [banerjeebihan@gmail.com](mailto:banerjeebihan@gmail.com)

<sup>2</sup>Academia Sinica Institute of Astronomy & Astrophysics, 11F of Astro-Math Bldg., No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan, R.O.C.

<sup>3</sup>Max-Planck-Institut für Astronomie (MPIA), Königstuhl 17, D-69117 Heidelberg, Germany

<sup>4</sup>Indian Institute of Astrophysics, Bangalore, India

<sup>5</sup>Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, 7820436, Macul, Santiago, Chile

Received 2024 February 16; revised 2024 April 11; accepted 2024 April 22; published 2024 June 5

## Abstract

Giant planets exhibit diverse orbital properties, hinting at their distinct formation and dynamic histories. In this paper, using Gaia Data Release 3 (DR3), we investigate if and how the orbital properties of Jupiters are linked to their host star properties, particularly their metallicity and age. We obtain metallicities for main-sequence stars of spectral type F, G, and K, hosting hot, warm, and cold Jupiters with varying eccentricities. We compute the velocity dispersions of the host stars of these three groups using kinematic information from Gaia DR3 and obtain average ages using a velocity dispersion–age relation. We find that the host stars of hot Jupiters are relatively metal rich ( $[Fe/H] = 0.18 \pm 0.13$ ) and young (median age of  $3.97 \pm 0.51$  Gyr) compared to the host stars of cold Jupiters in nearly circular orbits, which are relatively metal poor ( $0.03 \pm 0.18$ ) and older (median age of  $6.07 \pm 0.79$  Gyr). The host stars of cold Jupiters in high-eccentricity orbits, on the other hand, show metallicities similar to those of the hosts of hot Jupiters, but are older, on average (median age of  $6.25 \pm 0.92$  Gyr). The similarity in metallicity between the hosts of hot Jupiters and the hosts of cold Jupiters in high-eccentricity orbits supports high-eccentricity migration as the potential origin of hot Jupiters, with the latter serving as the progenitors of hot Jupiters. However, the average age difference between them suggests that the older hot Jupiters may have been engulfed by their host star over timescales  $\sim 6$  Gyr. This allows us to estimate the value of stellar tidal quality factor,  $Q'_* \sim 10^{6 \pm 1}$ .

*Unified Astronomy Thesaurus concepts:* [Gaia \(2360\)](#); [hot Jupiters \(753\)](#); [Metallicity \(1031\)](#); [Exoplanet migration \(2205\)](#); [Tidal interaction \(1699\)](#)

## 1. Introduction

Outside our solar system, more than 5000 planets have been detected and the number increases every day. The properties of the discovered exoplanets are often very different from those of the solar system planets. For example, many of the first detected exoplanets were hot Jupiter (HJs), i.e., giant planets with an orbital period shorter than 10 days (e.g., Mayor & Queloz 1995; Butler et al. 1997).

Planet properties, though seemingly diverse and different, are found to be tightly correlated with their host-star properties (e.g., Mulders 2018; Narang et al. 2018; Petigura et al. 2018; Hsu et al. 2019; Yang et al. 2020; Gaudi et al. 2021; Zhu & Dong 2021). For example, stellar spectral type and stellar mass, and accompanying planet mass and size, are found to be correlated (e.g., Howard et al. 2010, 2012; Fressin et al. 2013; Dressing & Charbonneau 2015; Mulders et al. 2015; Hedges-Ullman et al. 2019; Yang et al. 2020). Another robust correlation exists between host-star metallicity and planet mass (e.g., Gonzalez 1997; Santos et al. 2003, 2006, 2017; Fischer & Valenti 2005; Udry & Santos 2007; Reffert et al. 2015; Narang et al. 2018). The occurrence rate of giant planets increases with host-star metallicity, and on average Jupiter-hosting stars are likely to be of supersolar metallicity ( $[Fe/H] \sim 0.18 \pm 0.05$ ; e.g., Mulders 2018; Narang et al. 2018; Petigura et al. 2018).

However, many of these results are derived from the Kepler sample, which is complete for planets with orbital periods  $\leq 1$  yr (e.g., Narang et al. 2018; Petigura et al. 2018). Occurrence rate studies with Kepler data reveal that although short-period Jupiters are easier to detect, they are much rarer than small planets with similar orbital periods (e.g., Mulders 2018; Petigura et al. 2018; Hsu et al. 2019; Gaudi et al. 2021). Recently many more warm Jupiters (WJs) have been detected with the Transiting Exoplanet Survey Satellite (e.g., Eberhardt et al. 2023; Lubin et al. 2023), and long-period Jupiters have been discovered with radial velocities (RVs) of higher precision. Occurrence rate studies with combined data of transit and RV surveys show that the giant planet occurrence rate increases beyond 1 au and peaks around an orbital distance  $\sim 3$  au (e.g., Fernandes et al. 2019; Fulton et al. 2021; Kunitomo & Bryson 2021; Wolthoff et al. 2022). The most well-studied gas-giant, Jupiter in the solar system, is located at a distance of 5.2 au from the Sun and orbits a star with a relatively lower metallicity ( $[Fe/H] = 0$ ) than HJ hosts ( $[Fe/H] \approx 0.18$  dex; Mulders 2018; Narang et al. 2018; Petigura et al. 2018). Very long-period Jupiters, observed through the direct-imaging method, are found to show no particular preference toward host-star metallicity, and the average metallicity is around solar to subsolar (e.g., Swastik et al. 2021). Several high-resolution RV surveys report that the average metallicity of long-period Jupiter hosts is close to the solar value (e.g., Fulton et al. 2021; Wolthoff et al. 2022). However, not all long-period Jupiters are Jupiter analogs. A large fraction of the long-period Jupiters detected through RV

observations have large orbital eccentricity (Zakamska et al. 2011; Bitsch et al. 2020; Rosenthal et al. 2024), unlike the solar system’s Jupiter. On the other hand, the orbits of short-period Jupiters are predominantly circular (e.g., Jackson et al. 2023). Do these orbital diversities of Jupiters hint at different formation and evolution scenarios? A strong connection between the orbital architecture of planets and host-star properties could provide clues to the formation history. Therefore, it is only natural to ask, how do the host-star properties correlate with the orbital properties (orbital period and eccentricity) of the giant planets?

Buchhave et al. (2018) and Maldonado et al. (2018) have attempted to answer this question; however, they arrive at slightly different conclusions. Maldonado et al. (2018), with metallicities measured from high-resolution spectra for a sample of 88 host stars of giant planets, found that HJ hosts are relatively more metal rich than the hosts of cold Jupiters (CJs). Maldonado et al. (2018) also showed that the host stars of CJs are relatively richer in the  $\alpha$  elements than HJs. They argued that their results suggest a different formation mechanism for HJs and CJs. On the other hand, Buchhave et al. (2018) performed their analysis with 65 Jupiter hosts, and showed that the hosts of both HJs and CJs with high orbital eccentricity have supersolar metallicity and likely come from the same population; the host stars of low-eccentricity CJs have solar or subsolar metallicity, and belong to a different population. Some independent high-resolution spectroscopic surveys also indicate both of Jupiter hosts and high-eccentric planet hosts are metal-rich and have low X/Fe (e.g., Biazzo et al. 2022).

The simple existence of the host-star metallicity–orbital period correlation has deep implications for the formation pathways of Jupiters. It is still unclear how Jupiters are formed at different orbital locations (e.g., Dawson & Johnson 2018). The increase of the giant planet occurrence rate with host-star metallicity is best explained by core-accretion theory (e.g., Bodenheimer & Pollack 1986; Pollack et al. 1996; Ikoma et al. 2001; Mordasini et al. 2008; D’Angelo & Lissauer 2018). According to this theory, a solid core of a critical mass forms first ( $\sim 10 M_{\oplus}$  at 5 au under minimum mass solar nebula conditions; see also, e.g., Piso & Youdin 2014; Piso et al. 2015), which then accretes gas from the disk to form a Jupiter-like planet. However, in all protoplanetary disks, the available solid mass at a given radius ( $r$ ) increases with  $r$  (e.g., Powell et al. 2019; Armitage 2020). The sticking efficiency of grains also increases beyond the snow line (Okuzumi et al. 2012; Armitage 2020). Simulations show the efficiency of forming such massive cores is highest beyond the snow line of the disk, i.e., at 2–5 au from a Sun-like star (e.g., Mordasini et al. 2012). To complete this process within the disk dispersal timescale, a metal-rich disk is necessary (e.g., Ida & Lin 2005; Kornet et al. 2005; Wyatt et al. 2007; Boss 2010; Mordasini et al. 2012). Since the star and the circumstellar disk form from the same interstellar cloud material, we do expect more giant planets in metal-rich systems. However, even in metal-rich systems, the formation mechanisms of HJs and WJs are still not clearly understood. Two major theories, widely discussed in the literature for the formation of close-in giant planets, are:

(1) *In-situ formation*. A good fraction of the detected giant planets have an orbital period  $\leq 1$  yr. These planets are called HJs and WJs (the rest of them, with orbital periods  $> 1$  yr are called CJs). In in-situ formation theory,

HJs and WJs form at their present locations (e.g., Lee et al. 2014; Lee & Chiang 2017). Solid grains coagulate to form a massive core in the inner disk. If the core mass crosses a certain threshold value (typically  $\sim 10 M_{\oplus}$ ), it can accrete gas very rapidly (e.g., Piso & Youdin 2014; Piso et al. 2015). The main challenge for this mechanism is the lack of available solid mass at the inner disk, under standard disk conditions. However, an enhanced radial drift of pebbles can provide an additional large fraction of mass (Johansen & Lambrechts 2017). Since stellar metallicity traces the total available solid mass of the disk, this mechanism can only be active in extremely metal-rich systems (Dawson & Johnson 2018). Maldonado et al. (2018) found HJ hosts to be significantly more metal rich than CJ hosts and argued in favor of the in-situ formation of HJs. If in-situ formation indeed is the dominant channel of formation, then we expect to find a gradual decrease in the metallicity of host stars with the orbital period of the Jupiter they host, irrespective of their orbital eccentricity.

(2) *Migration*. Jupiters are formed beyond the snow line, and then they eventually migrate inward. Two possible migration channels are:

(2.1) *Disk migration*. In this scenario, the Jupiters forming in the outer disk interact with the gas disk and migrate inwards, exchanging angular momentum and dissipating energy (e.g., Goldreich & Tremaine 1980; Lin & Papaloizou 1986; D’Angelo et al. 2003; Baruteau et al. 2014). The final location of the planet is likely decided by mass loss, tidal interactions, the inner disk edge location, and many other factors (Chang et al. 2010). However, in this migration scenario, there is no clear host-star metallicity dependence (e.g., Armitage 2020), and planets undergoing disk migration do not get excited to highly eccentric orbits (Duffell & Chiang 2015). If disk migration plays a dominant role in forming close-in Jupiters, we expect the Jupiters to be in circular to low-eccentricity orbits, with a uniform distribution of host-star metallicity regardless of the Jupiter’s orbital period (e.g., Goldreich & Sari 2003; Duffell & Chiang 2015).

(2.2) *High-eccentricity tidal migration*. In this scenario, a fraction of CJs is excited to some highly eccentric orbits, either by planet–planet scattering (e.g., Chatterjee et al. 2008), by Kozai–Lidov cycles between planets (e.g., Kozai 1962; Lidov 1962), by secular interactions (e.g., Petrovich 2015; Hamers et al. 2017), or by interacting with external bodies (e.g., stellar flybys; e.g., Shara et al. 2016; Xiang-Gruess 2016), and then tidally circularize to a smaller orbit to become an HJ. The first three mechanisms are preferable in metal-rich systems where multiple giant planets are expected to form (e.g., Bitsch et al. 2015; Buchhave et al. 2018). Buchhave et al. (2018) found eccentric CJ hosts to have similar metallicities as HJ hosts, which supported this mechanism of formation.

Host-star metallicity can help distinguish between these formation mechanisms and hence investigating the host-star metallicity–orbital period correlation is vital for improving our understanding of giant planet formation. Although Buchhave et al. (2018) and Maldonado et al. (2018) have investigated this correlation, their sample sizes were small, and the host-star

metallicities were derived from different surveys and techniques and therefore were not homogeneous. In addition, as discussed earlier, they arrive at somewhat contradictory conclusions. Maldonado et al. (2018) arguing in favor of in-situ formation and Buchhave et al. (2018) finding support for high-eccentricity tidal migration. We revisit this problem with a significantly larger sample and homogeneous set of host-star metallicities derived from Gaia DR3. Metallicities derived from different observations and different instruments tend to suffer from offsets. Therefore, when making comparisons it is important to have a homogeneously measured sample of metallicities. In this paper, we have used metallicities of planet hosts uniformly determined with the Gaia Radial Velocity Spectrograph (RVS). The wavelength range covered by RVS is 846–870 nm, with medium resolving power  $R = \lambda/\Delta\lambda \sim 11,500$  (Cropper et al. 2018). Following the *GSP-Spec* module (Recio-Blanco et al. 2016), the chemo-physical parameters of nearly 5.6 million stars of our Galaxy have been derived from these spectra and reported in Gaia Data Release 3 (DR3; Gaia Collaboration et al. 2016, 2023; Recio-Blanco et al. 2023; Babusiaux et al. 2023). This sample of stars with homogeneously determined metallicities includes many planet hosts and allows us to compare the host-star metallicity of different planet samples.

Stellar metallicity and stellar age are found to be correlated (e.g., Carlberg et al. 1985; Meusinger 1991; Nordström et al. 2004). As the interstellar medium (ISM) eventually gets enriched in metals, younger stars tend to be born metal rich. Several works have argued that Jupiter hosts, in addition to being metal rich, are also younger than field stars and smaller-planet hosts (e.g., Swastik et al. 2021; Mustill et al. 2022; Unni et al. 2022; M. Narang et al. 2024, under review). There have been suggestions that Jupiters can only be formed only after a threshold chemical enrichment of the ISM has taken place (e.g., Mordasini et al. 2012; Narang 2022; M. Narang et al. 2024, under review). It has been also argued in the literature that HJ hosts seem younger because older HJs are getting destroyed by stellar tides (e.g., Hamer & Schlaufman 2019; Chen et al. 2023; Miyazaki & Masuda 2023). Therefore, it is an interesting exercise to compare the average metallicity and age of host stars of Jupiters in various orbital distances and eccentricities. Gaia DR3 also provides parallax ( $\pi$ ), proper motion in R.A. and decl. (pmRA and pmDE, respectively), and RV information of the stars. Using these measured quantities, the velocity dispersion for a group of stars can be obtained, which is a proxy for age (e.g., Binney et al. 2000; Aumer & Binney 2009; see Section 6).

In Section 2 we describe our sample and the selection criteria. Calibration of Gaia metallicities and comparison with metallicities derived from other surveys (e.g., GALAH and LAMOST) are discussed in Section 3. We present our results and discuss their statistical significance in Sections 4 and 5, respectively. We discuss the plausible physical origins of the observed results in Section 7, and finally we summarize in Section 8.

## 2. Sample Selection

Our analysis requires a sample of main-sequence stars that host giant planets at various orbital distances, with (1) stellar metallicity ([Fe/H]) homogeneously determined and (2) RV, parallax, and proper motion measured with acceptable accuracy. Gaia DR3 provides a homogeneous data set for both. From

the planetary systems composite table in the NASA exoplanet archive (Akeson et al. 2013; NASA Exoplanet Science Institute 2020), we selected confirmed planets (`controversial flag=0`) with measured orbital period (Orbital period = not null) on 11th December 2022. This gives us 4871 detected planets around 3612 host stars. We crossmatched the positions of these planet hosts with Gaia DR3 sources (I/355/gaiadr3<sup>6</sup>) and astrophysical parameters catalogs (I/355/paramp<sup>7</sup>) simultaneously, using a search radius of  $1''$ . Using the distances measured by Gaia parallaxes (Bailer-Jones et al. 2021; Gaia Collaboration et al. 2023), with parallax/parallax error  $> 10$ , we only keep sources within 250 pc, with robust measurements of their distances. We want to keep only the main-sequence host stars in our sample. Pecaut & Mamajek (2013) have provided an online table<sup>8</sup> for standard main-sequence stars, which is updated with new photometric observations (the last update was on 2022 April 16). In this table, dereddened colors and absolute  $G$ -band magnitudes of the standard stars of different spectral types from Gaia Data Release 2 (DR2) are given. Following Narang (2022), we performed a second-order polynomial fit between Gaia DR2 BP/RP colors and Gaia DR2  $G$ -band magnitude, restricting ourselves between F4 and M4 spectral types. The best-fit solution is shown in Equation (1) (also see Narang 2022). It has been verified by Narang (2022) that the median difference between the values of  $M_G$ ,  $G_{BP}$ , and  $G_{RP}$  derived from Gaia DR2 and Gaia DR3 are 0.01, 0.02, and 0.01, respectively. Therefore, the following equation holds for Gaia DR3 within  $\leq 3\%$ :

$$M_G = -0.43 \times (G_{BP} - G_{RP})^2 + 4.72 \times (G_{BP} - G_{RP}) + 1.00. \quad (1)$$

Gaia DR3 provides apparent magnitude ( $m_G$ ), observed color index ( $(G_{BP} - G_{RP})_{\text{obs}}$ ), distances ( $d$ ), the  $G$ -band extinction coefficient ( $A_G$ ), and color excess values ( $E(BP - RP)$ ) for all these sources. We note that 99% of the planet host stars in our sample within 250 pc have an ( $A_G$ ) value  $< 0.5$ , and an  $E(BP - RP)$  value  $< 0.3$ . To find the absolute magnitude and intrinsic color, we use the following equations:

$$M_G = m_G - 5 \log_{10}(d - 1) - A_G, \quad (2)$$

$$(G_{BP} - G_{RP}) = (G_{BP} - G_{RP})_{\text{obs}} - E(BP - RP). \quad (3)$$

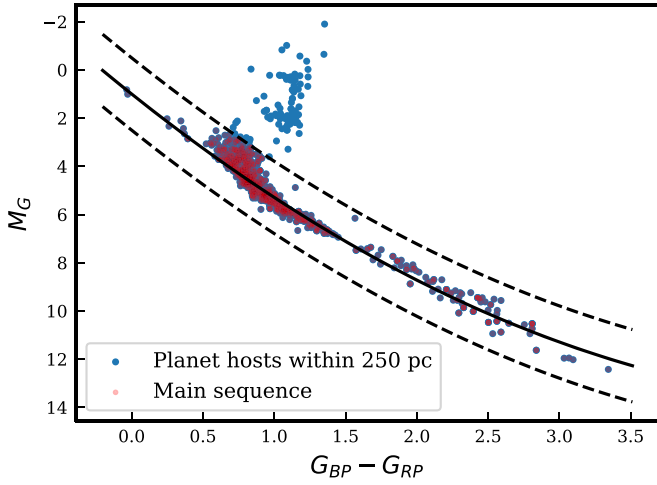
We select stars within  $\pm 1.5$  mag of the expected value of  $M_G$  for a given  $G_{BP} - G_{RP}$ , following Equation (1). In Figure 1 all the planet-hosting stars within 250 pc are shown as blue points. The black solid line represents Equation (1), and the two dashed lines are  $\pm 1.5$  mag of the expected  $M_G$  value for a given  $G_{BP} - G_{RP}$ . Therefore the stars within the two dashed lines are main-sequence stars and are plotted as red points. We find 610 main-sequence stars hosting 816 planets from this sample.

Gaia has observed all of these main-sequence, planet-hosting stars, but metallicities derived from RVS data are available for only a subset of them. In addition, the quality of the RVS data is not the same for all of them within this subset. To select only the best-quality data products in [Fe/H], we follow Recio-

<sup>6</sup> <https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=I/355/gaiadr3>

<sup>7</sup> <https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=I/355/paramp>

<sup>8</sup> [https://www.pas.rochester.edu/~emamajek/EEM\\_dwarf\\_UBVIJHK\\_colors\\_Teff.txt](https://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt)



**Figure 1.** Gaia DR3 color–magnitude diagram of planet-hosting stars in the solar neighborhood. All stars within 250 pc with detected planets are shown as blue points. Red points indicate main-sequence stars. The black solid line represents Equation (1), and the dashed lines represent  $\pm 1.5$  mag of the expected  $M_G$  value for a given  $G_{BP} - G_{RP}$ .

Blanco et al. (2023) and set the first 13 values of the quality flag chain to zero (see Appendix C of Recio-Blanco et al. 2023). We find 380 main-sequence stars among the 610 to have  $[\text{Fe}/\text{H}]$  values measured with the highest quality. These 380 stars host 519 planets in total.

In this work, we are primarily interested in Jupiters, and their host stars. We define Jupiters as planets with mass ( $M \sin i$  or  $M$ ) in between  $100 M_{\oplus}$  and  $1200 M_{\oplus}$ . From our sample, we restrict ourselves only to planets with a mass measured with  $3\sigma$  accuracy ( $M/\sigma_M \geq 3$ ), and orbital period measured with  $7\sigma$  accuracy ( $P/\sigma_P \geq 7$ ); we end up having 239 Jupiters, around 209 host stars. There are 26 systems with multiple Jupiters.

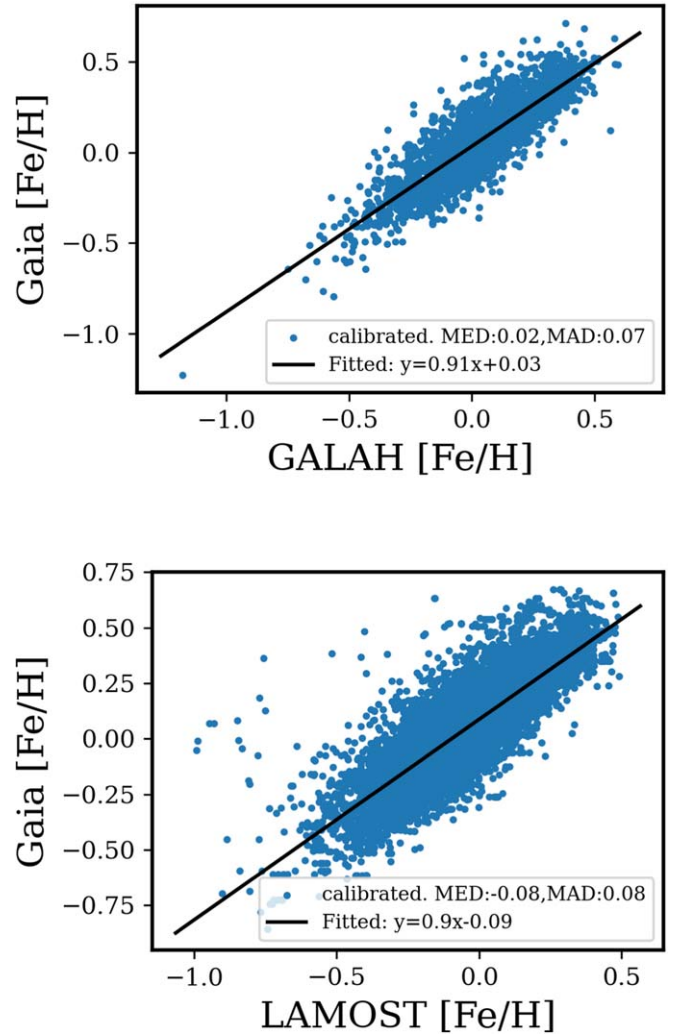
### 3. Calibration of Gaia Data Release 3 Metallicities

Gaia DR3 has provided the values of  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{M}/\text{H}]$ , and  $[\text{Fe}/\text{H}]$  derived from RVS spectra using GSP-Spec Matisse-Gauguin (Recio-Blanco et al. 2023). However, this method consistently finds lower surface gravity ( $\log g$ ) values for all the stars, and there is an overall offset compared to the literature. Therefore, a correction needs to be added to recover the true values of  $\log g$ . In addition, there is a weak dependence of metallicity values on  $\log g$ . Hence, applying a global correction to the  $[\text{M}/\text{H}]$ ,  $[\text{Fe}/\text{H}]$ , and  $[\alpha/\text{H}]$  values, by adding a polynomial of uncalibrated  $\log g$ , is suggested by Recio-Blanco et al. (2023). Following the same, we perform the following correction:

$$[\text{Fe}/\text{H}]_{\text{calibrated}} = [\text{Fe}/\text{H}]_{\text{uncalibrated}} + \sum p'_n (\log g)^n.$$

The values for  $p'_n$  are taken from Recio-Blanco et al. (2023).

After making these corrections to the  $[\text{Fe}/\text{H}]$  values from Gaia DR3 we compare the calibrated metallicities to their corresponding values in GALAH DR3 (e.g., Buder et al. 2021) and LAMOST Data Release 7 (DR7; e.g., Wang et al. 2020). For this comparison, we select all the main-sequence stars within 250 pc from Gaia DR3 (See Section 2), with best-quality RVS data (first 13 entries of the quality flag chain is zero; see Recio-Blanco et al. 2023) and crossmatch them with GALAH DR3 and LAMOST DR7 sources within a  $1''$  search radius around each of the Gaia DR3 sources. To get the best-quality data products of GALAH DR3, we follow the standard practice



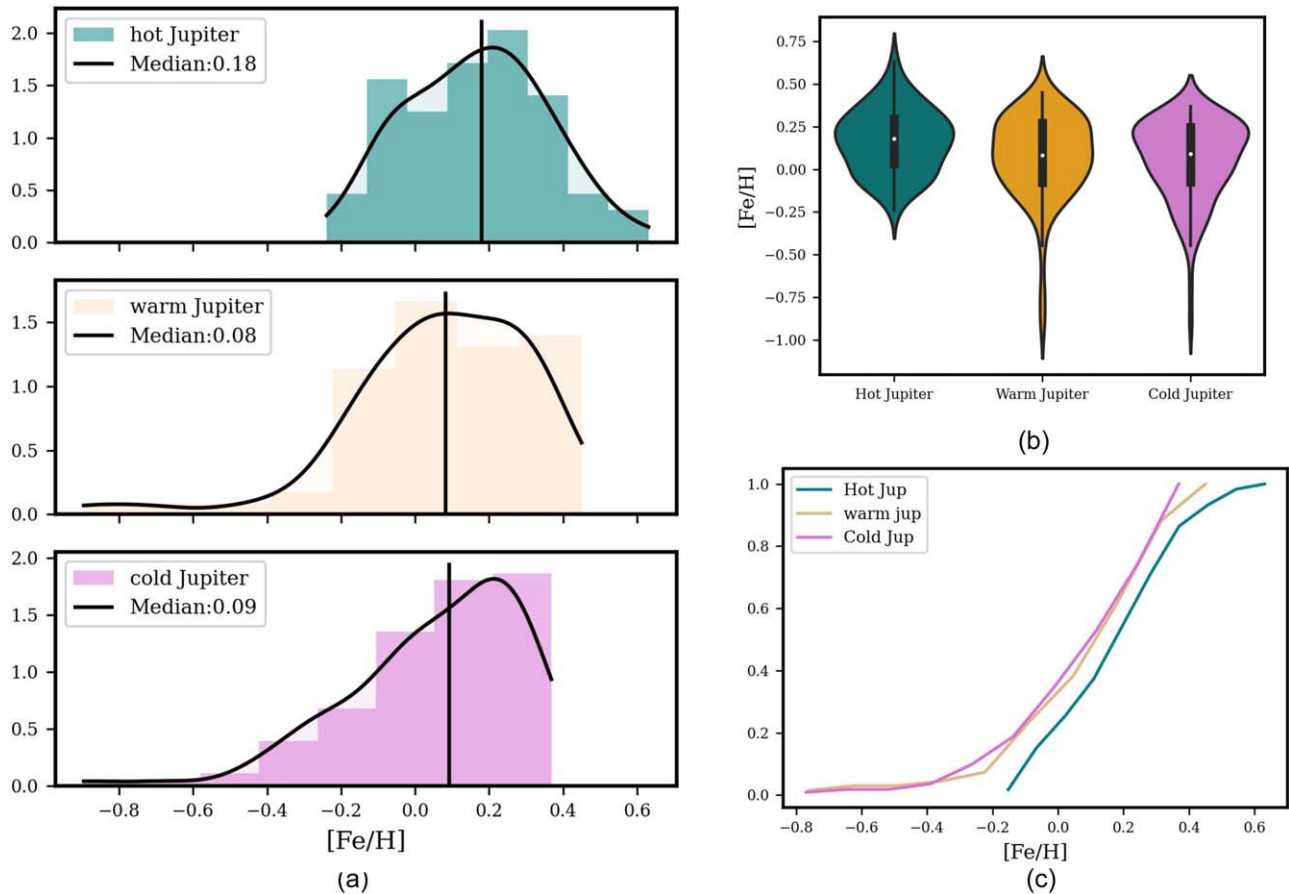
**Figure 2.** Comparison of the calibrated Gaia  $[\text{Fe}/\text{H}]$  with (a) GALAH and (b) LAMOST metallicities. The blue points represent the  $[\text{Fe}/\text{H}]$  values of stars common to each pair of instruments. The black solid line is a fitted straight line. The median of the difference (MED) of  $[\text{Fe}/\text{H}]$  values between (a) Gaia and GALAH and (b) Gaia and LAMOST, i.e., the median of  $(y - x)$  coordinates of the blue points is shown. The median absolute deviation (MAD) of  $[\text{Fe}/\text{H}]$  differences is also shown.

and use the recommended flag settings from the GALAH\_DR3\_main\_allstar\_v2.fits catalog: (1) `snr_c3_irc3_>30` and (2) `flag_sp=0` and `flag_fe_h=0`<sup>9</sup>. Between the Gaia and GALAH best-quality data, we find 2593 main-sequence sources within 250 pc. Similarly, for the best-quality data products from LAMOST, we set the  $[\text{Fe}/\text{H}]$  error  $< 0.3$  dex, and two quality flags for the  $R$ - and  $B$ -band spectral data, `bad_b` and `bad_r`, respectively, to zero<sup>10</sup>. We find 10,505 main-sequence stars common to Gaia DR3 and LAMOST DR7 within 250 pc.

Figure 2 compares the calibrated  $[\text{Fe}/\text{H}]$  values from Gaia DR3 to those from GALAH and LAMOST for main-sequence stars within 250 pc. The data can be fitted with a straight line with a slope close to one, indicating that the calibrated stellar metallicity values from Gaia DR3 are consistent with the GALAH and LAMOST values. The median  $[\text{Fe}/\text{H}]$  difference

<sup>9</sup> [https://www.galah-survey.org/dr3/using\\_the\\_data/#recommended-flag-values](https://www.galah-survey.org/dr3/using_the_data/#recommended-flag-values)

<sup>10</sup> <https://dr7.lamost.org/v2.0/doc/mr-data-production-description>



**Figure 3.**  $[\text{Fe}/\text{H}]$  distributions for HJs, WJs, and CJs. (a) Colored histograms and KDEs, which show HJs are relatively metal rich compared to WJs and CJs. Any difference between WJs and CJs in terms of host-star metallicity is not apparent. The black vertical lines denote the median  $[\text{Fe}/\text{H}]$ . (b) Violin plots showing the distribution of  $[\text{Fe}/\text{H}]$  for HJ, WJ, and CJ hosts (c) Cumulative distributions of  $[\text{Fe}/\text{H}]$  for HJ, WJ, and CJ hosts. The plots clearly show HJ hosts are distinctly metal rich.

between two instruments, i.e., the offset, is small for both cases, 0.02 between Gaia DR3 and GALAH DR3 and  $-0.08$  between Gaia DR3 and LAMOST DR7.

#### 4. Host-star Metallicity–Orbital Period Connection for Jupiters

As mentioned in Section 2, we have 239 Jupiters around 209 main-sequence stars in our final sample. The metallicities ( $[\text{Fe}/\text{H}]$ ) of all these stars are calibrated following the procedure described in Section 3. These 239 Jupiters are located at various orbital distances from their host stars. Based on their orbital periods ( $P$ ), these Jupiters can be subdivided into three categories: hot, warm, and cold. We define (a) HJs as Jupiters with an orbital period ( $P$ ) shorter than 10 days, (b) WJs with  $10 \text{ days} < P < 365 \text{ days}$ , and (c) CJs as  $P > 365 \text{ days}$ . In our sample, we have 59 HJs, 68 WJs, and 112 CJs. We investigated whether the host-star metallicities ( $[\text{Fe}/\text{H}]$ ) of these three groups show any similarities or differences. In Figure 3(a) we show histograms of  $[\text{Fe}/\text{H}]$  distributions for HJ, WJ, and CJ hosts. We have fitted each normalized histogram with a kernel density estimate (KDE) and those are shown as black curves. The vertical black lines represent the medians of these distributions. Inspection of these distributions reveals that the  $[\text{Fe}/\text{H}]$  distributions of the WJ and CJ hosts are similar, but that of the HJ hosts is different (see Figure 3(a)). WJ and CJ hosts have a low-metallicity tail, which is absent in the case of HJ hosts.

The metallicity distribution of HJ hosts is flatter and has a smaller spread. The medians of the WJ and CJ hosts are very close to each other, but the median  $[\text{Fe}/\text{H}]$  of the HJ hosts is higher.

In Figure 3(b) the metallicity distributions of HJ, WJ, and CJ hosts are shown as violin plots. As can be seen, the HJ hosts are relatively metal rich while the WJ and CJ hosts, on the other hand, have low-metallicity tails and have similar median metallicities.

In Figure 3(c) we have shown the cumulative distributions of  $[\text{Fe}/\text{H}]$  of the three types of Jupiter hosts. The cumulative distributions also show the HJ hosts are distinctly more metal rich than the WJ and CJ hosts. As one can see from Figure 3(c),  $\sim 40\%$  of the WJ and CJ hosts have subsolar metallicity ( $[\text{Fe}/\text{H}] < 0$ ) while that is true for only  $\sim 20\%$  of the HJ hosts. To characterize these distributions quantitatively, we computed and compared the following quantities: median as a measure of the central tendency, MAD as a measure of the dispersion, kurtosis to characterize the flatness of the distribution, and skewness to characterize any asymmetry and longer tails compared to a normal distribution toward a particular direction. For all the groups within our sample, these quantities are computed and summarized in Table 1.

We find that, if we only divide Jupiters into these three bins based on orbital period, HJ hosts seem to be more relatively metal rich than WJ and CJ hosts (similar to what has been reported in the literature by e.g., Narang et al. 2018;

**Table 1**

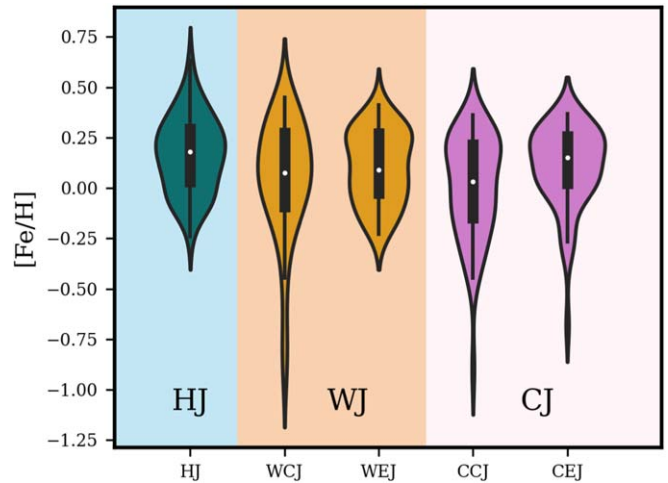
Statistical Properties of the Distribution of Host Star [Fe/H] for Different Classes of Jupiters

| Sample | Median | MAD  | Kurtosis | Skewness |
|--------|--------|------|----------|----------|
| HJ     | 0.18   | 0.13 | -0.51    | 0.08     |
| WJ     | 0.08   | 0.18 | 3.14     | -1.38    |
| CJ     | 0.09   | 0.16 | 1.71     | -1.15    |
| CEJ    | 0.15   | 0.12 | 2.66     | -1.37    |
| CCJ    | 0.03   | 0.18 | 1.06     | -0.92    |

Petigura et al. 2018). The median metallicities (vertical black lines in Figure 3(a)) of WJ and CJ hosts are similar to each other, and lower than those of HJ hosts (see Table 1). We also see that the metallicity distributions of the HJ hosts are flatter and have a smaller spread about a high-metallicity value (0.18 dex). The spread is also almost symmetric about the median. This is reflected in the small values of kurtosis and skewness (See Table 1). However, in the case of the WJ and CJ hosts, the distributions have low-metallicity tails, and they are asymmetric about the median. This is reflected in positive kurtosis and negative skewness scores as shown in Table 1.

However, in addition to their orbital periods, these Jupiters also show diversity in orbital eccentricity. Many independent works on planetary dynamics suggest eccentricity ( $e$ ) is a tracer of the dynamic history of planets (e.g., Rasio & Ford 1996; Chatterjee et al. 2008; Ford & Rasio 2008; Ghosh & Chatterjee 2023). Typically, planets gain a large eccentricity either by interacting with each other (e.g., Kozai 1962; Chatterjee et al. 2008) or with an external body (Muñoz et al. 2016; Shara et al. 2016). In the first scenario, metal-rich disks are preferred, while in the latter no such dependence on metallicity is expected (e.g., Shara et al. 2016; Dawson & Johnson 2018). Therefore, it is interesting to investigate if host-star metallicity is also correlated with the eccentricity of accompanying planets. With this in mind, we further subdivided the HJs, WJs, and CJs into two bins with high and low orbital eccentricity. We chose  $e = 0.2$  as the dividing criterion between the two bins. Hereafter, we will call the Jupiters with  $e < 0.2$  as the ones in “circular” orbits, and those with  $e > 0.2$  as planets in “eccentric” orbits. We find 56 CJs in “eccentric” orbits, and 56 CJs in “circular” orbits; hereafter we will call them CJs in eccentric orbits (CEJs) and CJs in circular orbits (CCJs). WJs also show a smaller but significant variation in eccentricity. We find 33 WJs in eccentric orbits and 35 WJs in circular orbits. On the other hand, most of the HJs are in circular orbits. We find only two HJs in eccentric orbits.

In Figure 4 host-star [Fe/H] distributions for the low- and high-eccentricity subgroups of WJs and CJs are shown as violin plots. We do not show the eccentric subgroups of HJ, because only two HJs have high eccentricities in our sample. The color palette is the same as in Figure 3. The low-eccentricity, or “circular,” subgroup is shown on the left and the eccentric subgroup is shown on the right. We find that the difference of median host-star [Fe/H] between the eccentric and circular subgroups is greater than 0.1 dex, but only for CJs.



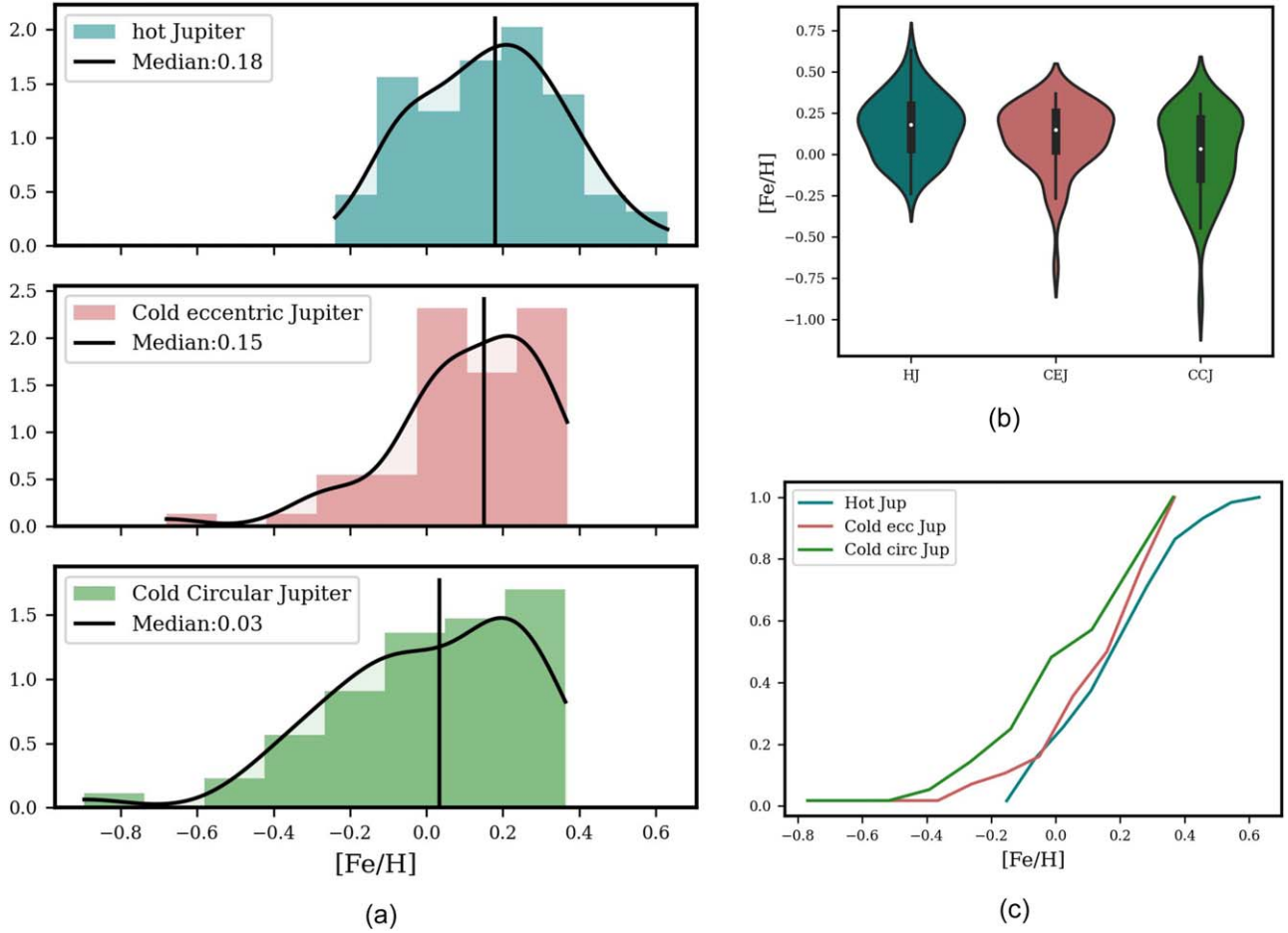
**Figure 4.** Host-star [Fe/H] for high- and low-eccentricity planets among HJs, WJs, and CJs. For each group, the low-eccentricity, or “circular,” subgroup is shown on the left, and the eccentric subgroup is shown on the right. HJs are not subdivided into two subgroups due to the lack of eccentric HJs in our sample. We note that, only for CJs, the difference of the median host-star [Fe/H] between the eccentric and circular subgroups is greater than 0.1 dex.

Therefore, for the rest of the paper, we treat CEJs and CCJs as two separate groups but combine the eccentric and circular subgroups for HJs and WJs.

In Figure 5 we have shown the [Fe/H] distributions of CEJ and CCJ hosts along with HJ hosts. Figure 5(a) shows the three histograms and KDEs, and the vertical line corresponds to the median of each. Figure 5(b) is a violin plot of the three distributions. It is evident from the figure that the host stars of HJs and CEJs are more metal rich than CCJ hosts. We find that CEJs and CCJs have different distributions of [Fe/H], with CEJs being relatively more metal rich on average (median [Fe/H] = 0.15). The average host-star metallicity of the CCJs, however, is close to the solar value (median [Fe/H] = 0.03; see Table 1). The distribution of [Fe/H] also seems to be flatter in the case of CCJs. We note that with a larger sample and homogeneous data set our result qualitatively follows the findings of Buchhave et al. (2018).

Comparing the [Fe/H] distributions of the HJs, WJs, CEJs, and CCJs we arrive at the following conclusions:

1. The median difference between the [Fe/H] distributions of the HJ and CJ hosts is 0.09, and between the HJ and WJ hosts it is 0.1. On the other hand, the median difference between the WJ and CJ hosts is only  $-0.01$ .
2. The distributions of [Fe/H] among the CJ hosts show a dichotomy, dependent on the eccentricity of the accompanying Jupiter. The median difference between the CEJ and CCJ hosts is 0.12. The median difference of [Fe/H] distributions between the HJ hosts and CCJ hosts is 0.15. The median [Fe/H] values of the HJ and CEJ hosts are very similar, with a difference of only 0.03.
3. Based on the median differences and shapes of the distributions, the [Fe/H] distributions of the CCJ hosts differ from those of the HJ hosts and CEJ hosts. On the other hand, the [Fe/H] distributions of the HJ and CEJ hosts are very similar. This hints at the possibility that CEJ hosts and HJ hosts belong to the same underlying population, and CCJ hosts are from a different population.



**Figure 5.** (a)  $[\text{Fe}/\text{H}]$  distributions of HJ, CEJ, and CCJ hosts. The black curved lines are the KDEs from the colored histograms. The median  $[\text{Fe}/\text{H}]$  distributions of HJs and CEJs are similar, while the CCJ hosts have a lower median  $[\text{Fe}/\text{H}]$ . We also note that for HJ hosts, the  $[\text{Fe}/\text{H}]$  distribution has a longer tail toward high  $[\text{Fe}/\text{H}]$ , and CCJs have a long tail toward low metallicity. Also, the  $[\text{Fe}/\text{H}]$  distribution for the CCJs is much flatter. (b) Violin plots showing the  $[\text{Fe}/\text{H}]$  distributions of HJ, CCJ, and CEJ hosts. CCJs and CEJs are denoted by green and red, respectively. (c) Cumulative distributions of  $[\text{Fe}/\text{H}]$  for HJs, CEJs, and CCJs. We can see that the central part of the host-star  $[\text{Fe}/\text{H}]$  distributions of HJs and CEJs resemble each other.

The immediate questions that arise from these results are the following:

1. How statistically significant are these results?
2. Are these results coming from the true nature of these distributions, or are they due to a random sampling of points from a different underlying distribution?
3. Do other properties of the host stars (e.g., age) of the sample also follow the similarities and differences between the HJ, CEJ, and CCJ populations?
4. What formation and evolution channels of Jupiter formation do these results support?
5. Do these results suffer from selection effects and observational biases?

We address these questions in the following sections: we discuss questions 1 and 2 in Section 5, question 3 in Section 6, and questions 4 and 5 in Section 7.

## 5. Statistical Analysis of the Results

In this section, several tests to compute the statistical significance of these results are discussed.

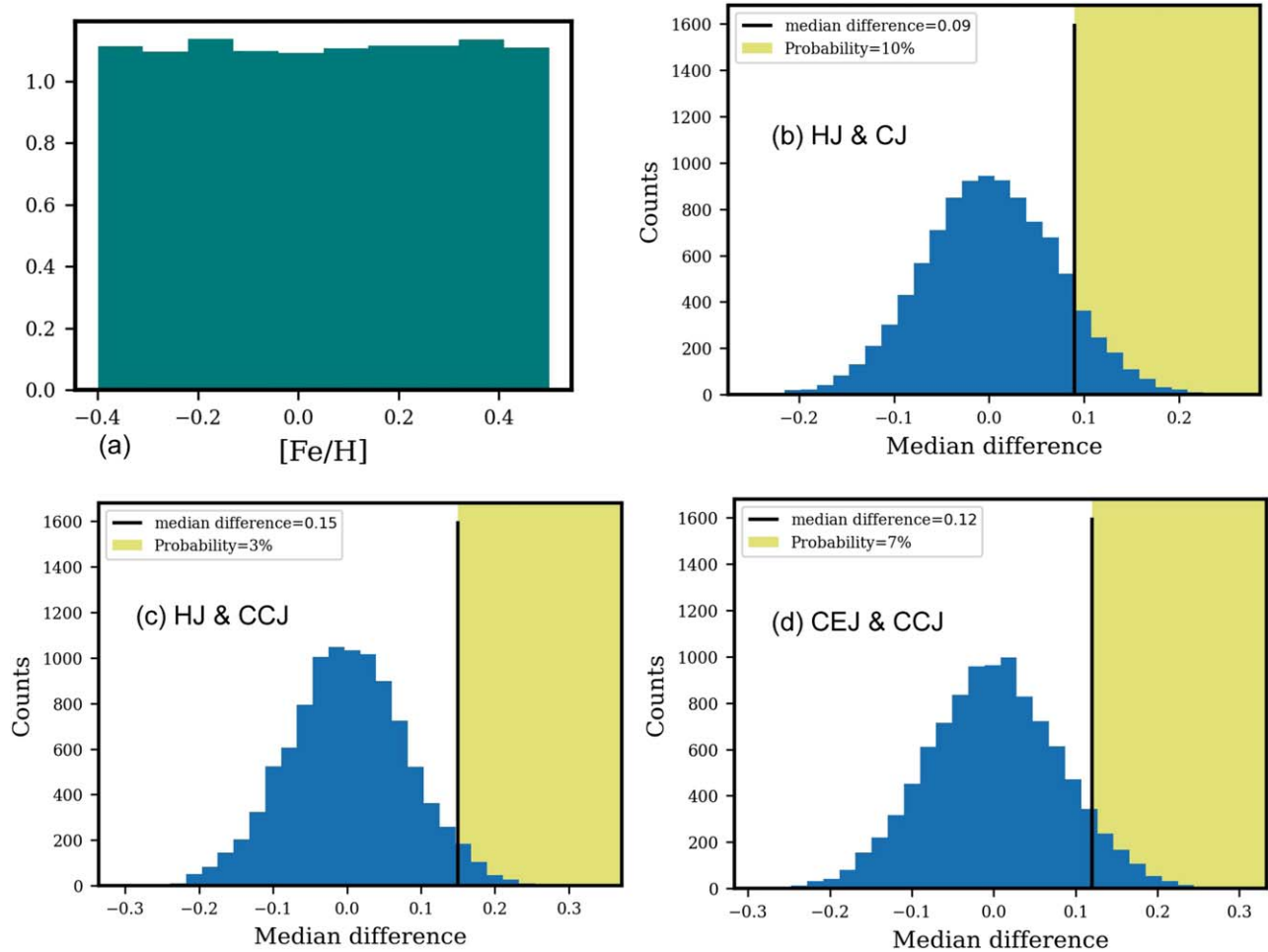
First, we approach the problem using a Monte Carlo analysis. We assume all Jupiter hosts belong to the same underlying population, i.e., they have one common  $[\text{Fe}/\text{H}]$

distribution for all Jupiter hosts, HJ, WJ, and CJ alike. Then we randomly draw samples from this distribution, divide them into separate groups, and label them as HJ, CJ, CEJ, or CCJ. We keep the sample size of each group equal to the observed sample size of that group. Then we compute the probability of finding the difference of the medians of the two groups greater or equal to the observed median differences. (See Section 5.1.) If this probability is greater than 5%, we do not rule out the possibility that our observed median difference is an outcome of random draws of subsamples from a common metallicity distribution of host stars. Otherwise, we rule out the possibility.

In Section 5.2, we will perform several nonparametric null hypothesis tests between the observed  $[\text{Fe}/\text{H}]$  distributions of HJ, CEJ, and CCJ hosts.

### 5.1. Monte Carlo Analysis

First, we assume the underlying distribution of host-star  $[\text{Fe}/\text{H}]$  is uniform (see Figure 6(a)). We note that  $\geq 95\%$  of all Jupiter hosts have  $[\text{Fe}/\text{H}]$  values between  $-0.4$  and  $0.5$ . We assume all Jupiter hosts have an equal probability of having an  $[\text{Fe}/\text{H}]$  value anywhere in this range. The median difference of host-star  $[\text{Fe}/\text{H}]$  between the HJs and CJs from our observed sample is  $0.09$  dex. We randomly draw two samples of the exact sizes of HJs and CJs in observed samples from this



**Figure 6.** Monte Carlo Analysis 1: can random sampling from an underlying uniform distribution of host-star  $[\text{Fe}/\text{H}]$  produce the observed results? (a) A histogram of the uniform sample. The blue histograms are the distributions of median differences for (b) HJs and CJs, (c) HJs and CCJs, and (d) CEJs and CCJs. The vertical black solid lines in all these figures represent the observed median differences. The region right to this black solid line is colored yellow, where here the median difference  $\geq$  the observed median difference. The probability of median difference  $\geq$  observed median difference is the area under the histogram in the yellow shaded region divided by the total number of counts.

uniform distribution and compute the median difference between the two samples. This exercise is repeated 10,000 times, so we obtain a distribution of median differences of two subsamples drawn from the same underlying distribution. In Figure 6(b) we show this distribution as the blue histogram. The vertical black line in Figure 6(b) represents the observed median difference of host-star  $[\text{Fe}/\text{H}]$  between HJs and CJs (0.09 in this case). The area right of this vertical black line is shaded yellow. The obtained median difference is greater than the observed median difference in this region. We find that the probability of occurrence of the median difference  $\geq$  the observed difference by simply taking the ratio of the number of occurrences where the obtained median difference is larger than the observed median difference, and the total number of draws (10,000). For the HJs and CJs, the probability is 10% (see Figure 6(b)).

In the case of the CCJs and CEJs the sample size is 56 for both. The observed median difference of host-star  $[\text{Fe}/\text{H}]$  distributions between the HJs and CCJs is 0.15 and between the CEJs and CCJs it is 0.13. Following the same procedure, we compute the probability that a random draw from the underlying uniform distribution produces this result or a larger median difference. Between the HJs and CCJs, the probability

is only 3%, and between the CEJs and CCJs the probability is 7%.

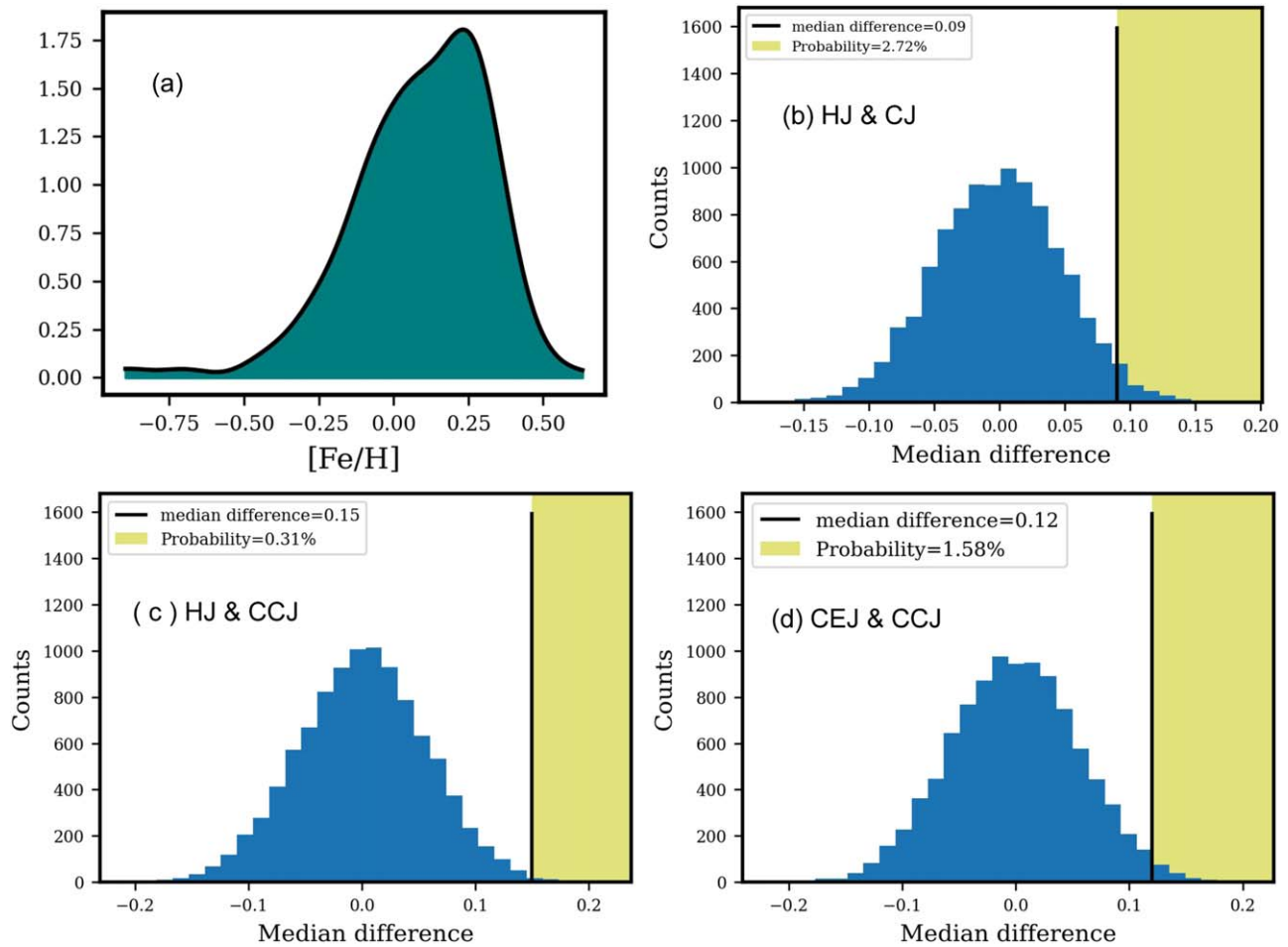
Therefore, if the  $[\text{Fe}/\text{H}]$  values of all Jupiter hosts belong to the same uniform distribution, the probability of getting the observed median differences or greater is (1) 10% for HJs and CJs, (2) 3% for HJs and CCJs, and (3) 7% for CEJs and CCJs.

Therefore, we can only rule out the possibility of having a common parent population of metallicities only for HJs and CCJs, according to the threshold we have set.

However, the underlying distribution of  $[\text{Fe}/\text{H}]$  of the Jupiter hosts is very unlikely to be uniform. As multiple works suggest, Jupiter hosts are preferentially metal rich (e.g., Santos et al. 2017; Mulders 2018; Narang et al. 2018; Zhu & Dong 2021). Therefore, to represent a more realistic scenario, we take the KDE of the  $[\text{Fe}/\text{H}]$  distribution of our observed sample to be the underlying distribution for all Jupiter hosts. Then we repeat the same exercises as above. In Figure 7 we summarize our findings.

Figure 7(a) shows the KDE of the  $[\text{Fe}/\text{H}]$  of all Jupiter hosts in our sample. We assume this to be the underlying distribution for all Jupiters. First, we draw samples of the exact sizes as the HJ and CJ samples. Figure 7(b) shows the resulting distributions of the median difference between the HJs and CJs. We find that the probability of the median difference being greater





**Figure 7.** Monte Carlo Analysis 2: can random sampling from an underlying uniform distribution of host-star  $[\text{Fe}/\text{H}]$  produce the observed results? (a) The underlying distribution is a KDE of the  $[\text{Fe}/\text{H}]$  distribution of Jupiter hosts in our sample. We draw random samples from this distribution repeatedly of the exact sizes of our observed sample and compute the median difference. The blue histograms are the distributions of median differences for (b) HJs and CJs, (c) HJs and CCJs, and (d) CEJs and CCJs. The vertical black solid lines in all these figures represent the observed median differences. The region right to this black solid line is colored yellow, where where the median difference is greater than the observed median difference. The probability of the median difference  $\geq$  the observed median difference is the area under the histogram in the yellow shaded region divided by the total number of counts.

than the observed is 2.72%. In the case of the HJs and CCJs (see Figure 7(c)) this probability is only 0.31%. Finally, between the CEJs and CCJs the probability is 1.58% (see Figure 7(d)).

Therefore, if the  $[\text{Fe}/\text{H}]$  values of all Jupiter hosts belong to the same distribution as approximated by the KDE of our sample, the probability of getting the observed median differences or greater is (1) 2.72% for the HJs and CJs, (2) 0.31% for the HJs and CCJs, and (3) 1.58% for the CEJs and CCJs.

From the preceding analysis, we can safely conclude that:

1. If we assume a common, uniform  $[\text{Fe}/\text{H}]$  distribution of the Jupiter-hosting stars, and the  $[\text{Fe}/\text{H}]$  distributions of the HJs, CCJs, and CEJs are randomly drawn subsamples from this common parent distribution, we find the probability of getting the observed or a greater median difference between them is  $<5\%$  only for HJ and CCJ hosts. Therefore we can rule out the possibility that HJ and CCJ hosts have a common uniform parent  $[\text{Fe}/\text{H}]$  distribution with  $>95\%$  confidence.
2. If we assume the  $[\text{Fe}/\text{H}]$  distribution of the Jupiter host stars follows the KDE of the  $[\text{Fe}/\text{H}]$  values in our sample then we can rule out the possibility that HJs and CCJs

belong to the same parent population with 99.99% confidence. We can do the same for HJ and CJ hosts with 97.28% confidence and CEJ and CCJ hosts with 98.42% confidence. This hints strongly that CEJ and HJ hosts are well-separated populations from CCJ hosts, in terms of metallicity.

## 5.2. Statistical Tests

We have found in Section 4 that the Gaia DR3 metallicities ( $[\text{Fe}/\text{H}]$ ) of the host stars of HJs and CEJs show similar distributions, however, these distributions differ from the  $[\text{Fe}/\text{H}]$  distribution of CCJ hosts. This indicates perhaps the HJ and CEJ hosts belong to the same underlying population, whereas the CCJ hosts come from a different population. Now we will use various statistical tests to compare the central tendencies, dispersions, and tails of these distributions, and determine the statistical significance of these results.

Since we do not have any prior information about the true metallicity distribution of these stars, to compare the samples we use nonparametric, distribution-free tests (e.g., Corder & Foreman 2009). All these tests check if the null hypothesis is true, i.e., if the two populations underlying the two samples are

**Table 2**

Results of Two-sample, Two-sided Statistical Tests, Mann–Whitney Tests, Kruskal–Wallis Tests, and Kolmogorov–Smirnov Tests

| Test      | Sample 1 | Sample 2 | Statistic          | $p$ -value        |
|-----------|----------|----------|--------------------|-------------------|
| MW U Test | HJ       | CJ       | $4063.5 \pm 200.5$ | $0.014 \pm 0.013$ |
|           | HJ       | CCJ      | $2222.0 \pm 111.0$ | $0.001 \pm 0.001$ |
|           | CEJ      | CCJ      | $1944.0 \pm 112.5$ | $0.025 \pm 0.024$ |
|           | HJ       | CEJ      | $1846.5 \pm 118.5$ | $0.264 \pm 0.217$ |
| KW Test   | HJ       | CJ       | $6.09 \pm 3.12$    | $0.014 \pm 0.013$ |
|           | HJ       | CCJ      | $10.18 \pm 3.92$   | $0.001 \pm 0.001$ |
|           | CEJ      | CCJ      | $4.79 \pm 2.77$    | $0.024 \pm 0.023$ |
|           | HJ       | CEJ      | $1.253 \pm 1.108$  | $0.263 \pm 0.215$ |
| K-S Test  | HJ       | CJ       | $0.22 \pm 0.04$    | $0.03 \pm 0.03$   |
|           | HJ       | CCJ      | $0.32 \pm 0.05$    | $0.004 \pm 0.004$ |
|           | CEJ      | CCJ      | $0.32 \pm 0.05$    | $0.006 \pm 0.006$ |
|           | HJ       | CEJ      | $0.198 \pm 0.03$   | $0.175 \pm 0.132$ |

**Note.** We must have a  $p$ -value  $< 0.05$  to reject the null hypothesis. The standard `scipy` packages have been used to obtain the test results. Mann–Whitney (MW) U, Kruskal–Wallis (KW), and Kolmogorov–Smirnov (K-S) test results are given in this table.

identical. The alternative hypothesis is that they are not identical.

In these tests, the first step is to compute the test statistic, which quantifies the difference between two or more data groups (the data sets of [Fe/H] of HJ, CJ, CEJ, and CCJ hosts in our case). Then, we compute the  $p$ -value, which represents the probability of observing a test statistic as extreme as, or more extreme than, the one computed from the data, assuming that the null hypothesis is true (i.e., there is no real difference between the data groups). A smaller  $p$ -value suggests stronger evidence against the null hypothesis, indicating that the observed data are unlikely to have occurred by random chance alone.

To reject the null hypothesis, we set the confidence level to 95%, or the threshold  $p$ -value to 0.05. If the resultant  $p$ -value of all the tests for two samples is  $\leq 0.05$  we reject the null hypothesis with  $\geq 95\%$  confidence.

We do not know if the observed samples are the true representatives of the underlying distributions. Therefore, instead of comparing the observed data sets directly, we use bootstrapped resampling, i.e., we repeatedly draw random samples from the data sets with replacement. We compute the statistical tests between the bootstrapped samples in each iteration and finally get distributions of test statistics and  $p$ -values. We obtain the median and MAD values of these distributions, and these results are summarized in Table 2.

To compare the central tendencies and dispersion of two distributions, we use the MW U test (e.g., Mann & Whitney 1947) and the KW test (e.g., Kruskal & Wallis 1952). For comparison of the cumulative distributions, we use the K-S test (e.g., Smirnov 1948; Hodges 1958). A short description of these tests can be found in the Appendix.

The comparison between host-star [Fe/H] distributions of HJs, CJs, CCJs, and CEJs is compiled in Table 2. The null hypothesis in each case is the two samples (Samples 1 and 2 in Table 2) belong to the same parent population. If we look at the MW U test and KW test results, the data set pair {HJ, CJ} has  $p$ -value  $< 0.05$ . However, the obtained  $p$ -value for the K-S test statistic between the pair {HJ, CJ} is  $0.03 \pm 0.03$ , so the upper limit is marginally greater than our threshold of 0.05. This

implies the medians of HJs and CJs are well separated and the difference is significant, but the difference between their cumulative distributions is not significant.

For the data set pairs {HJ, CCJ} and {CEJ, CCJ}, we have  $p$ -value  $< 0.05$  in all cases. Therefore, for {HJ, CCJ} and {CEJ, CCJ} we can safely reject the null hypothesis and conclude they do not come from the same underlying population.

However, for the pair {HJ, CEJ} we have the smallest test statistic for all the tests, and also a large  $p$ -value  $> 0.05$  corresponding to the test statistic. Therefore we cannot reject the null hypothesis in this case.

All of the above statistical test results strengthen our findings of Section 4 and we can safely conclude the following:

1. Regarding the [Fe/H] distributions, host stars of HJs and CCJs do not belong to the same parent population. Similarly, the host stars of CEJs and CCJs belong to different parent populations.
2. In terms of the [Fe/H] distributions, host stars of HJs and CEJs likely belong to the same parent population. There is no evidence in the data to suggest otherwise.

Therefore, our results so far, support the possibility that HJ hosts and CEJ hosts have similar properties, and CEJs might be the progenitors of HJs, both forming from disks with supersolar metallicity. On the other hand, CCJs represent a separate class and form in relatively lower-metallicity disks (Also see Buchhave et al. 2018).

## 6. Stellar Age and Orbital Period Correlation

As the Universe evolves, the ISM gets enriched in metals. Therefore stars that form at a later epoch, i.e., younger stars, have a higher metallicity compared to older stars, on average. This results in a positive correlation between stellar metallicity and age over Galactic timescales (e.g., Carlberg et al. 1985; Meusinger 1991; Nordström et al. 2004). On the other hand, as metallicity increases in the molecular cloud, the efficiency in giant planet formation also increases (e.g., Ida & Lin 2005; Wyatt et al. 2007; Mordasini et al. 2012; Piso et al. 2015). It has been adequately demonstrated in the literature that giant planet host stars are more metal rich compared to field stars and small-planet hosts (e.g., Gonzalez 1997; Santos et al. 2003, 2006; Fischer & Valenti 2005; Udry & Santos 2007; Narang et al. 2018). One might also expect these stars to be relatively young, from the metallicity–age correlation of the stars, and some works in the literature suggest the same (e.g., Swastik et al. 2022; Swastik et al. 2024; M. Narang et al. 2024, under review). Several independent works have found HJ hosts to be younger on average than field stars (e.g., Hamer & Schlaufman 2019; Mustill et al. 2022; Blaylock-Squibbs & Parker 2023). However, the reason for the relative age difference is debated, as it might be a consequence of the age–metallicity correlation (e.g., Swastik et al. 2021, 2022; Narang et al. 2023; M. Narang et al. 2024, under review), or the destruction of the older population of HJs by their host stars (e.g., Hamer & Schlaufman 2019), or a combination of both. In this work, we are comparing the host stars of Jupiters only, but these Jupiters are located at various orbital distances. In the previous sections, we have discussed their differences in metallicity. Now we ask, do these Jupiter hosts follow a similar trend in terms of age?

However, it is nontrivial to measure the individual ages of main-sequence stars, especially for solar or lower-mass stars

(e.g., Soderblom 2010). Methods of obtaining the ages of individual stars include isochrone fitting (e.g., Takeda et al. 2007; Soderblom 2010), asteroseismology measurements (e.g., Bazot et al. 2008), using an empirical relation from stellar spindown (e.g., Barnes 2009). However, an isochrone-fitted age has large uncertainties (e.g., Takeda et al. 2007). Ages determined by modeling asteroseismology measurements have much less uncertainty, but these observations are resource intensive, and homogeneous asteroseismology measurements for our sample are not available yet (e.g., Bazot et al. 2008). Similarly, a homogeneous data set of stellar rotation periods is also not available for the stars in our sample. Nevertheless, we are not interested in the ages of individual stars, rather we want to know how the groups of HJ, WJ, CEJ, and CCJ host stars differ in terms of their average age. One of the ways to obtain an ensemble age for a group of stars is to use the dispersion in their Galactic space velocities as a proxy for age. This has been well established and widely used in the literature (e.g., Wielen 1977; Carlberg et al. 1985; Meusinger 1991; Dehnen & Binney 1998; Binney et al. 2000; Manoj & Bhatt 2005; Aumer & Binney 2009; Schönrich et al. 2010; Sharma et al. 2014; Yu & Liu 2018). Using this method, M. Narang et al. (2024, under review) have shown that the hosts of Jupiters are metal rich and younger than field stars and small-planet hosts. Similar findings, using different methods are reported in, e.g., Swastik et al. (2022, 2024) and Miyazaki & Masuda (2023). We applied a similar analysis to our sample of stars hosting HJs, WJs, and CJs in circular and eccentric orbits.

Gaia DR3 provides parallax ( $\pi$ ), proper motion in R.A. (pmRA,  $\mu_\alpha$ ) and decl. (pmDE,  $\mu_\delta$ ), and RV ( $\gamma$ ) information for all the stars in our sample. First, we calculate the space velocity components ( $U$ ,  $V$ ,  $W$ ) from the observed RV and proper motion:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = B \begin{bmatrix} \gamma \\ \frac{k\mu_\alpha}{\pi} \\ \frac{k\mu_\delta}{\pi} \end{bmatrix}, \quad (4)$$

where  $k$  is a constant,  $k = 4.740470 \text{ km s}^{-1}$  (e.g., Ujjwal et al. 2020; Narang 2022) and  $B$  is a  $3 \times 3$  matrix that transforms the equatorial velocity vector to Cartesian system with the  $z$ -axis pointing toward the Galactic north pole. Equation (4) gives us heliocentric space velocities. Subtracting the solar velocity gives  $(U_\odot, V_\odot, W_\odot) = (11.1, 12.24, 7.25) \text{ km s}^{-1}$  (Schönrich et al. 2010), giving us space velocities with respect to our local standard of rest.

Now, for each of the ensembles of HJ, WJ, and CJ hosts we compute the dispersion in Galactic velocity as the sum of the dispersions in individual components:

$$\sigma_{\text{tot}} = \sqrt{\sigma_U^2 + \sigma_V^2 + \sigma_W^2}, \quad (5)$$

where  $\sigma_U^2 = \frac{1}{N} \sum_{i=0}^N (U_i - \bar{U})^2$

In this way, we get one value of the velocity dispersion for each bin (HJs, WJs, CJs, CEJs, and CCJs). However, there are two major sources of errors in this computed dispersion:

1. The errors in the measured values of proper motions and RVs.

2. Sampling errors. Since our sample in each bin is only a subset of a larger population, the velocity dispersion we compute may not represent the true velocity dispersion of the population.

For this reason, we compute a distribution of  $\sigma_{\text{tot}}$  with bootstrap resampling. Suppose, in a bin, we have  $N$  number of host stars. We select  $N$  host stars randomly out of them with replacement, i.e., the same star can be selected twice. Measurements of parallax, pmRA, pmDE, and RVs are available for all these stars, and each measurement has an associated error. We consider each measurement to be a Gaussian with the mean ( $\mu$ ) as the reported value and the standard deviation ( $\sigma$ ) as the associated error, and we draw random values out of these distributions. Using these values, following Equations (4) and (5) we compute  $\sigma_{\text{tot}}$ . We repeat this exercise 10,000 times to get a distribution of  $\sigma_{\text{tot}}$ . These distributions can be seen in Figure 8 (left column) We report the median and MAD values of these distributions below.

We find that the velocity dispersions of the host stars of CCJs and CEJs are similar,  $47.5 \pm 2.3 \text{ km s}^{-1}$  and  $47.9 \pm 2.6 \text{ km s}^{-1}$ , respectively. HJ hosts have a smaller velocity dispersion ( $40.6 \pm 1.9 \text{ km s}^{-1}$ ). On the other hand, WJ hosts have a velocity dispersion in between ( $45.1 \pm 1.9 \text{ km s}^{-1}$ ). The smaller velocity dispersion indicates that HJ hosts are younger compared to CJ and WJ hosts (see Figure 8).

This velocity dispersion can be used as an age estimator. Following Narang (2022) we use the relation from Aumer & Binney (2009) to find the average age ( $\tau$ ) of an ensemble of stars from the total velocity dispersion as:

$$\tau = (10 \text{ Gyr} + \tau_1) \left( \frac{\sigma_{\text{tot}}}{\nu_{10}} \right)^{(1/\beta)} - \tau_1. \quad (6)$$

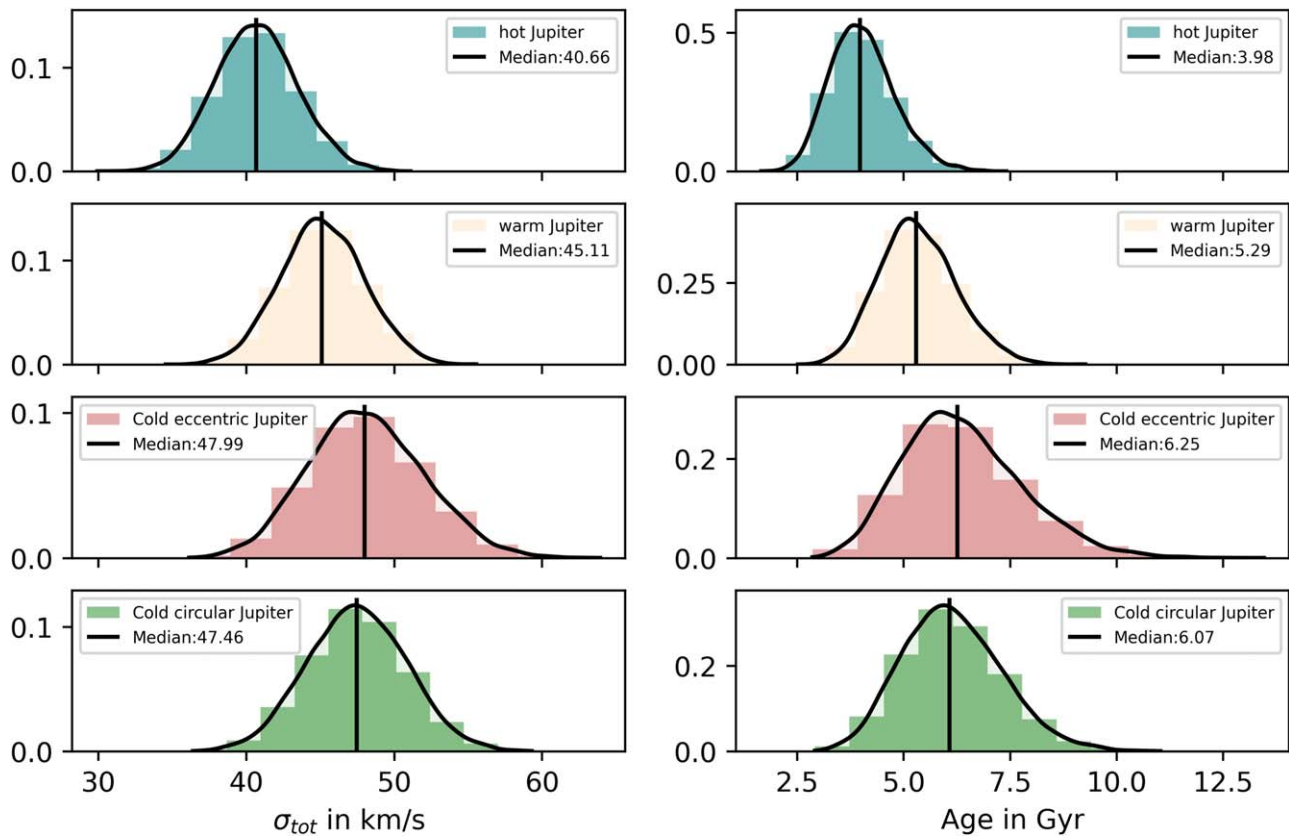
From Aumer & Binney (2009),  $\nu_{10} = 57.157 \text{ km s}^{-1}$ ,  $\beta = 0.385$ , and  $\tau_1 = 0.261 \text{ Gyr}$ . Using Equation (6), we convert the distribution of  $\sigma_{\text{tot}}$  to the corresponding age distributions (see Figure 8). We report the median and MAD of those age distributions below as MED  $\pm$  MAD.

CCJ and CEJ hosts have median ages of  $6.07 \pm 0.79 \text{ Gyr}$  and  $6.25 \pm 0.92 \text{ Gyr}$ , respectively. HJs are slightly younger, with  $3.97 \pm 0.51 \text{ Gyr}$ , and WJs have an average age of  $5.28 \pm 0.6 \text{ Gyr}$ . Therefore, despite having similar [Fe/H] values on average, CEJ and HJ hosts have different average ages. Our results are in good agreement with other works in literature (e.g., Chen et al. 2023; Miyazaki & Masuda 2023; Swastik et al. 2024).

## 7. Discussion

We chose a well-curated sample of main-sequence stars hosting Jupiter-like planets in the solar neighborhood with reliable measurements of metallicity ([Fe/H]) and kinematics (proper motion, parallax, and RV) from Gaia DR3. Based on the orbital distances from their host star, and orbital eccentricities, we subdivided these Jupiters into four groups (HJs, WJs, CEJs, and CCJs). We have compared the host-star age and metallicities of these groups.

To summarize the results of the last sections, we find that, on average, HJ hosts are metal rich ([Fe/H] =  $0.18 \pm 0.13$ ) and young (average age  $\sim 3.97 \pm 0.51 \text{ Gyr}$ ). On the other hand, CCJ hosts are relatively metal poor, around solar metallicity ([Fe/H] =  $0.03 \pm 0.18$ ), and relatively older (average age  $\sim 6.07 \pm 0.79 \text{ Gyr}$ ). However, CEJ hosts, despite being



**Figure 8.** Velocity dispersion (left column) and age distributions (right column) of the HJ, WJ, CEJ, and CCJ hosts. The distributions are obtained using the bootstrapped resampling method described in the text. We find that HJ hosts are younger with the smallest velocity dispersion.

metal rich on average ( $[\text{Fe}/\text{H}] = 0.15 \pm 0.12$ ) are relatively older (average age  $\sim 6.25 \pm 0.92$  Gyr).

The outcomes of our statistical tests on the  $[\text{Fe}/\text{H}]$  distributions of the host stars unveil that HJ hosts and CCJ hosts do not belong to the same population. Similarly the  $[\text{Fe}/\text{H}]$  distributions of the CEJ and CCJ hosts most likely do not come from the same distribution. On the other hand, the CEJ and HJ hosts likely come from the same population. However, we see an age difference between the CEJ and HJ hosts, despite their metallicities being similar.

### 7.1. Implications of the Results

As we have discussed in the [Introduction](#), the correlation between orbital properties and host-star metallicities of giant planets has deep implications for the understanding of the formation and evolution pathways of giant planets. Here we discuss how our findings favor or disfavor each scenario.

#### 7.1.1. In-situ Formation

The plausibility of forming HJs and WJs at their present locations is one of the major open questions (e.g., Batygin et al. 2016; Lee & Chiang 2017; Dawson & Johnson 2018; Poon 2021). In order to form a Jupiter, a massive core needs to be formed before disk gas dissipates. However, the inner disk itself has a small amount of dust mass, and to form Jupiter the solid mass of the inner disk should be enhanced by 2 orders of magnitude (e.g., Dawson & Johnson 2018). This enhancement should also reflect in the metallicity of the host star. As a result, if in-situ formation is at play, HJ hosts should be very metal rich on average. However, if we go further from the host

star, massive core can grow within the disk dispersal time even in disks with gradually less solid mass (e.g., Mordasini et al. 2012; Piso et al. 2015). Therefore, Jupiters located in further orbits should have host stars with gradually decreasing metallicity, on average. If in-situ formation was the dominant mechanism for the formation of Jupiters, we should have seen a gradual decrease of the median metallicity of Jupiter hosts with orbital period (Maldonado et al. 2018). Although we see HJ hosts to be metal rich, we do not notice any significant difference in metallicity between the WJ and CJ hosts. In addition, the in-situ formation scenario cannot account for the similarities in metallicity between the CEJ and HJ hosts.

#### 7.1.2. Disk Migration

In gas-disk migration, a massive planet perturbs the nearby gas and sends it onto horseshoe orbits via corotation torques, deflecting distant gas by Lindblad torques and exchanging angular momentum in the process (Goldreich & Tremaine 1980; Lin & Papaloizou 1986; Baruteau et al. 2014). As a result, a net inward torque acts on the planet and the planet starts to migrate inwards. However, this mechanism is independent of the metallicity of the system (e.g., Dawson & Johnson 2018). Therefore, if gas-disk migration were the dominant mechanism to sculpt all the close-in HJs and WJs, we would not have seen a metallicity and orbital period correlation. On the other hand, gas-disk migration cannot excite the eccentricity of a planetary orbit to a high value (e.g., Duffell & Chiang 2015). Therefore WJs in eccentric orbits (See Figure 4) are hard to explain by this mechanism.

### 7.1.3. High-eccentricity Tidal Migration

The most popular theory of forming HJs is by the tidal migration of CEJs (e.g., Socrates et al. 2012; Petrovich 2015; Hamers et al. 2017; Dawson & Johnson 2018; Teysandier et al. 2019). Above a threshold eccentricity, CEJs can experience rapid tidal decay of their orbit. The first step is to excite the giant planets to high eccentricity. This can be attained if a giant planet interacts with a perturber. The perturber can be another giant planet or an external flyby. If the excitation is due to an external flyby, we do not expect any dependence on host-star metallicity (e.g., Shara et al. 2016). However, in metal-rich disks, giant planet formation is very efficient, due to the larger amount of solid mass present (e.g., Ida & Lin 2005; Mordasini et al. 2012; Bitsch et al. 2015). If multiple giants are formed in sufficiently close orbits, it may lead to planet–planet scattering, and one of them can be excited to very high eccentricity (e.g., Rasio & Ford 1996; Chatterjee et al. 2008; Ford & Rasio 2008). Therefore, the metallicity enhancement of the host stars of CEJs might be an indicator of planet–planet scattering happening in such initially metal-rich disks (e.g., Buchhave et al. 2018; Dawson & Johnson 2018).

If the eccentricity is very high, and the tidal migration timescale is very short, the CEJs tidally circularize in a shorter orbit and end up as HJs (e.g., Petrovich 2015). But if the eccentricity is below a threshold value, and the tidal migration timescale is comparable to the stellar lifetime, the CEJs remain in their eccentric orbit with a large semimajor axis (e.g., Petrovich 2015).

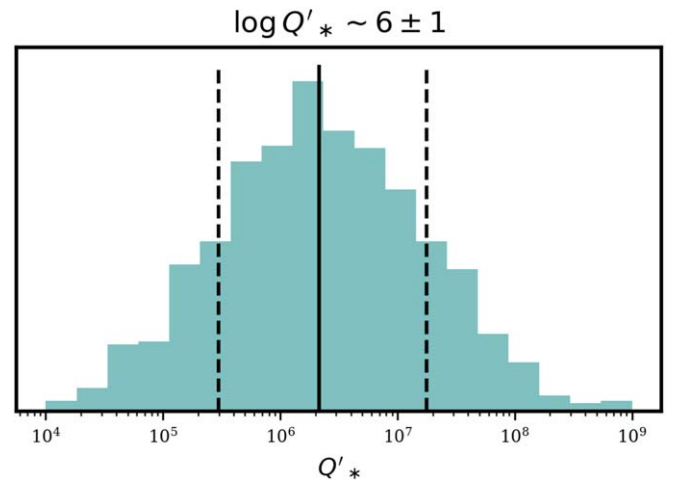
This theory predicts that CEJs are the progenitors of close-in HJs (e.g., Socrates et al. 2012; Petrovich 2015). Therefore we would expect similar host-star properties between HJs and CEJs, on average. We do find that the [Fe/H] distributions of the HJ and CEJ hosts are similar, and the null hypothesis of them being drawn from the same parent population could not be rejected. Both of the host-star populations are also metal rich, as we expect for planet–planet scattering to occur (e.g., Chatterjee et al. 2008; Bitsch et al. 2015).

On the other hand, we find that the average ages, derived from the velocity dispersions of the HJ and CEJ hosts, are different. HJ hosts are relatively younger than CEJ hosts. But if CEJs indeed are the progenitors of HJs, the ages of their host stars should be similar.

One possible explanation of our findings might be the destruction of older HJs (e.g., Hamer & Schlaufman 2019; Miyazaki & Masuda 2023). If a significant fraction of older HJs are engulfed by their host star, the remaining HJ hosts would appear younger on average. Even after orbital eccentricity is sufficiently damped, the HJs can fall into the host star by tidal interactions. The semimajor axis of a circularized Jupiter shrinks because of the tides raised by the planet on the star (e.g., Goldreich & Soter 1966; Jackson et al. 2008; Barker 2020). The infall timescale ( $t_{\text{in}}$ ), for an HJ with orbital period  $P$ , semimajor axis  $a$ , and mass  $M_p$  around a star of mass  $M_*$  is given by (e.g., Barker & Ogilvie 2009; Lai 2012; Hamer & Schlaufman 2019):

$$t_{\text{in}} = \frac{2}{13} \left| \frac{a}{\dot{a}} \right| = \frac{2}{13} \frac{2Q'_* M_* P}{9 M_p 2\pi} \left( \frac{a}{R_*} \right)^5, \quad (7)$$

where  $Q'_* = 3Q_*/2k$  is the modified stellar tidal quality factor,  $Q_*$  is the ratio of the maximum energy stored in tides to the



**Figure 9.** Possible distribution of  $\log Q'_*$ . Here we assume the inspiral timescale is  $\sim 6$  Gyr and the HJs that have been destroyed had similar planetary and stellar properties. The median and 16th and 84th percentiles are  $\log Q'_* \approx 6 \pm 1$ .

energy dissipated in one orbital cycle, and  $k$  denotes the tidal Love number.

The value of  $Q'_*$  depends on the internal structure of the star, tidal forcing frequency, and amplitude in general (e.g., Barker & Ogilvie 2009; Barker 2020; Miyazaki & Masuda 2023). There have been attempts to constrain the value of  $Q'_*$  from theory (e.g., Barker & Ogilvie 2009; Penev et al. 2018) and observed data sets (e.g., Bonomo et al. 2017; Hamer & Schlaufman 2019; Labadie-Bartz et al. 2019; Miyazaki & Masuda 2023). Assuming CEJs are the true progenitors of HJs, we try to estimate  $Q'_*$  of Jupiter hosts in the following way.

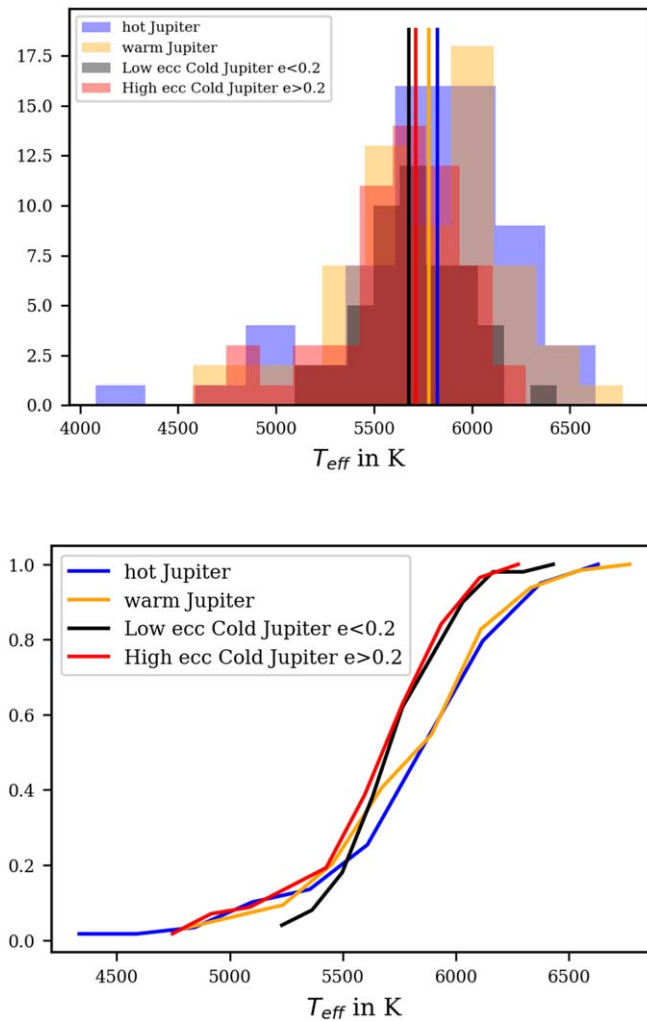
We note from Figure 8 that median age of CEJ hosts is  $\sim 6$  Gyr, and the upper limit of the ages of the HJ hosts is  $\sim 6$  Gyr. Therefore, the probability of finding an HJ host older than 6 Gyr is very small. If CEJs are the progenitors of HJs, and over a timescale  $\sim 6$  Gyr most of the HJs are engulfed by the star, this scenario may occur.

If we equate the tidal inspiral timescale ( $t_{\text{in}}$ ) to 6 Gyr, we will get an estimate of  $Q'_*$ , depending on the stellar and planetary parameters. However, we note that for a CEJ to become an HJ and then to be engulfed by the star, two processes are involved: (1) high-eccentricity tidal migration of CEJs to a shorter orbit and (2) tidal inspiral of the HJ onto the star. Figure 8 indicates that sum of the two timescales should be  $\sim 6$  Gyr. We are assuming that  $t_{\text{in}}$  is much larger compared to the high-eccentricity tidal migration timescale. In other words, we are placing an upper limit on  $Q'_*$ .

If we assume that the HJs that have been engulfed by their host star have similar planetary and host-star properties to the observed HJs, estimation of  $Q'_*$  is straight forward. We draw 1000 random points from the planetary masses, orbital periods, stellar radii, and stellar masses and plug them into Equation (7) to get a distribution of  $Q'_*$ . The distribution of  $Q'_*$  is shown in Figure 9. We find median and 16th and 84th percentiles of:

$$\log Q'_* \approx 6 \pm 1. \quad (8)$$

This range of values of  $Q'_*$  is consistent with those existing in the literature (e.g., Penev et al. 2018; Hamer & Schlaufman 2019; Miyazaki & Masuda 2023) and can lead to the destruction of a significant fraction of the older HJ population, leading to the average young age, and an average age difference with the CEJs.



**Figure 10.**  $T_{\text{eff}}$  distributions of the host stars of HJs, WJs, CEJs, and CCJs. In terms of  $T_{\text{eff}}$  we do not see a significant difference.

However, this value is only an order of magnitude estimate of  $Q'_*$ , and holds only in the premise of our assumptions. To explain the stability of short-period HJs, Lai (2012) argues for a larger value of  $Q'_* \sim 10^8$ – $10^9$ .

For consistency check, and to check for possible bias in the sample, we also compared the effective temperature ( $T_{\text{eff}}$ ) of the host stars of HJs, WJs, CEJs, and CCJs (Figure 10). It is apparent from Figure 10 that in terms of  $T_{\text{eff}}$  we do not see a significant difference between the host stars of different groups. We find that the HJ and WJ hosts are marginally hotter, with  $\delta T_{\text{eff}} < 200$  K, and the CEJ and CCJ hosts have similar  $T_{\text{eff}}$  on average (Figure 10(b)). We note that the observed metallicity–age distribution cannot be due to this marginal difference of  $T_{\text{eff}}$ .

## 8. Summary

1. We started with a well-curated sample of 702 planet hosts on the main sequence within 250 pc, with metallicity ( $[\text{Fe}/\text{H}]$ ) and kinematics homogeneously measured and reported in Gaia DR3. We have used only the best-quality data products from Gaia DR3 following Recio-Blanco et al. (2023). We made consistency checks of Gaia DR3 metallicities with GALAH and LAMOST. For our analysis we used only Jupiter hosts, defining Jupiter as a planet with mass ( $M_J$ ) between ( $100 M_{\oplus} < M_J < 1200 M_{\oplus}$ ).

2. We subdivided the Jupiters into three groups based on their orbital period, HJs, WJs, and CJs, and compared the host-star properties, namely metallicity ( $[\text{Fe}/\text{H}]$ ) and age, of these three groups. We find that HJ hosts are the more metal rich ( $[\text{Fe}/\text{H}] = 0.18 \pm 0.13$ ) while WJ and CJ hosts have similar metallicities on average ( $[\text{Fe}/\text{H}] = 0.08 \pm 0.18$ ). Most of the HJs are in circular orbits, but we also find that a significant fraction of WJs and CJs are in eccentric orbits. We find no difference in host-star metallicity between low- and high-eccentricity WJs. However, for CJs, the eccentric population (CEJ) have high host-star metallicities ( $[\text{Fe}/\text{H}] = 0.15 \pm 0.12$ ) on average compared to the low-eccentricity population (CCJ). CCJ hosts have  $[\text{Fe}/\text{H}]$  values close to solar ( $[\text{Fe}/\text{H}] = 0.03 \pm 0.18$ ). This finding, from a larger and homogeneous data set, agrees with previous work in the literature by Buchhave et al. (2018).
3. To test the statistical significance of the observed results, we have performed Monte Carlo analyses and several nonparametric statistical tests, as described in Section 5.2. Our findings suggest that the observed difference in host-star metallicity between HJs and CCJs, and between CEJs and CCJs, cannot be a consequence of random sampling from an underlying parent population. Based on the  $[\text{Fe}/\text{H}]$  distributions of the host stars, we can conclude that HJs and CCJs, as well as CEJs and CCJs, do not come from the same parent population. However, HJs and CEJs likely come from the same underlying population, and the null hypothesis cannot be ruled out. In other words, CEJs might be the progenitors of HJs in very metal-rich systems. Whereas, CCJs are born separately, in relatively metal-poor environments.
4. We find that the host stars of HJs and CEJs have similar  $[\text{Fe}/\text{H}]$  values on average, but their average ages are different. The similarity in  $[\text{Fe}/\text{H}]$  distributions supports the theory of high-eccentricity migration of CEJs as the progenitors of HJs, triggered by planetary perturbation (e.g., Bitsch et al. 2015; Buchhave et al. 2018). However, the difference in average age indicates that older HJs might be getting destroyed, hence the HJs appear younger on average. If we assume CEJs are indeed the progenitors of HJs and a fraction of HJs are getting destroyed because of tidal interactions with their host star, we find the average value stellar modified tidal quality factor ( $Q'_*$ ) to be  $Q'_* \approx 10^{6 \pm 1}$ . This range of values agrees with other works in the literature (e.g., Penev et al. 2018; Hamer & Schlaufman 2019; Miyazaki & Masuda 2023).

## Acknowledgments

We thank Prof. Thomas Henning for his insightful suggestions. The authors thank the anonymous referee for their valuable feedback and useful insights. This research has made use of the NASA Exoplanet Archive (NASA Exoplanet Science Institute 2020), which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program. 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. We have made use of several packages from `Scipy` (Virtanen et al. 2020) for statistical tests. We sincerely thank the Infosys Foundation for funding the travel to PPVII in Kyoto to present this work as a poster. We acknowledge support of the Department of Atomic Energy, Government of India, under Project Identification No. RTI 4002. PKN acknowledges TIFR's postdoctoral fellowship. PKN also acknowledges support from the Centro de Astrofísica y Tecnologías Afines (CATA) fellowship via grant Agencia Nacional de Investigación y Desarrollo (ANID), BASAL FB210003.

*Software:* Numpy (Harris et al. 2020), Topcat (Taylor et al. 2005), Scipy (Virtanen et al. 2020), Matplotlib (Hunter 2007), Seaborn (Waskom 2021)

*Facilities:* NASA Exoplanet Archive, Centre de Données astronomiques de Strasbourg; Strasbourg astronomical Data Center, European Space Agency (ESA) Gaia Satellite Mission

## Appendix Nonparametric Statistical Tests

If the true nature of the original distribution is unknown, the best way to compare two samples is to make use of nonparametric statistical tests. In this work, we have used several tests; here we describe them briefly. For details please see Corder & Foreman (2009).

1. *MW U test.* This test compares between the central tendencies of two distributions. In this test, the two samples are merged and the combined sample is sorted in ascending order and ranked (e.g., Mann & Whitney 1947; Corder & Foreman 2009). The strategy is to see if the ranks are randomly mixed or if the ranks of the two groups are clustered at opposite ends. The test computes the individual sums of the ranks of the two samples. The smaller of the two sums is called the U-statistic. For large samples (at least  $>8$  for each group) the distribution of the U-statistic can be approximated with a normal distribution. The test then evaluates whether the observed U-statistic significantly deviates from what would be expected under the null hypothesis of no difference between the distributions.
2. *KW test.* The KW H test (e.g., Kruskal & Wallis 1952; Corder & Foreman 2009) tests the null hypothesis that the population medians of all the data groups are equal. It is a nonparametric extension of the Analysis of Variance test. This test is used to compare two or more samples. First, all the samples are combined and ranked in ascending order. For comparison between  $k$  groups, the KW H-statistic is defined by:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1),$$

where  $R_i$  is the sum of the ranks from a particular group and  $n_i$  is the number of values from the corresponding rank sum. If there are ties between the ranks, the statistic is divided by the following correction ( $C_H$ ), and a new statistic is calculated:

$$C_H = 1 - \frac{\sum(T^3 - T)}{N^3 - N}.$$

If the sample size of the individual groups is large, the KW statistic asymptotically approaches the chi-squared

distribution. In our case, all groups have large samples ( $>50$ ) for this assumption to hold. The  $p$ -value of the test is then obtained by computing the survival function from the chi-square distribution of  $H$ .

3. *K-S test.* In this test (e.g., Smirnov 1948; Hodges 1958; Corder & Foreman 2009), two empirical distribution functions (EDFs) are constructed from two data sets, and the maximum vertical distance between the two functions is computed. If we have two samples of size  $m$  and  $n$  from CDFs  $F$  and  $G$  and wish to test the null hypothesis that  $F(x) = G(x)$  for all  $x$ , the two-sample K-S statistic is defined by:


$$D_{KS} = \max|\hat{F}_m(x) - \hat{G}_n(x)|,$$

where  $\hat{F}_m, \hat{G}_n$  are the two EDFs. The null hypothesis is rejected at confidence level  $\alpha$  if:

$$D_{KS} > c(\alpha) \sqrt{\frac{n+m}{n.m}},$$

where  $c(\alpha) = \sqrt{-0.5 \ln(\alpha/2)}$  in general.

## ORCID iDs

Bihan Banerjee  <https://orcid.org/0000-0001-8075-3819>  
 Mayank Narang  <https://orcid.org/0000-0002-0554-1151>  
 P. Manoj  <https://orcid.org/0000-0002-3530-304X>  
 Thomas Henning  <https://orcid.org/0000-0002-1493-300X>  
 Himanshu Tyagi  <https://orcid.org/0000-0002-9497-8856>  
 Arun Surya  <https://orcid.org/0000-0002-9967-0391>  
 Prasanta K. Nayak  <https://orcid.org/0000-0002-4638-1035>  
 Mihir Tripathi  <https://orcid.org/0009-0007-2723-0315>

## References

- Akeson, R. L., Chen, X., Ciardi, D., et al. 2013, *PASP*, **125**, 989  
 Armitage, P. J. 2020, *Astrophysics of Planet Formation* (2nd ed.; Cambridge: Cambridge Univ. Press)  
 Aumer, M., & Binney, J. J. 2009, *MNRAS*, **397**, 1286  
 Babusiaux, C., Fabricius, C., Khanna, S., et al. 2023, *A&A*, **674**, A32  
 Bailer-Jones, C. A. L., Rybizki, J., Fousneau, M., et al. 2021, *AJ*, **161**, 147  
 Barker, A. J. 2020, *MNRAS*, **498**, 2270  
 Barker, A. J., & Ogilvie, G. I. 2009, *MNRAS*, **395**, 2268  
 Barnes, S. A. 2009, in *IAU Symp. 258*, The Ages of Stars, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse (Cambridge: Cambridge Univ. Press), 345  
 Baruteau, C., Crida, A., Paardekooper, S. J., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 667  
 Batygin, K., Bodenheimer, P. H., & Laughlin, G. P. 2016, *ApJ*, **829**, 114  
 Bazot, M., Monteiro, M. J. P. F. G., & Straka, C. W. 2008, *JPhCS*, **118**, 012008  
 Biazzo, K., D'Orazi, V., Desidera, S., et al. 2022, *A&A*, **664**, A161  
 Binney, J., Dehnen, W., & Bertelli, G. 2000, *MNRAS*, **318**, 658  
 Bitsch, B., Lambrechts, M., & Johansen, A. 2015, *A&A*, **582**, A112  
 Bitsch, B., Trifonov, T., & Izidoro, A. 2020, *A&A*, **643**, A66  
 Blaylock-Squibbs, G. A., & Parker, R. J. 2023, *MNRAS*, **519**, 3643  
 Bodenheimer, P., & Pollack, J. B. 1986, *Icar*, **67**, 391  
 Bonomo, A. S., Desidera, S., Benatti, S., et al. 2017, *A&A*, **602**, A107  
 Boss, A. P. 2010, in *IAU Symp. 265*, Chemical Abundances in the Universe: Connecting First Stars to Planets, ed. K. Cunha, M. Spite, & B. Barbuy (Cambridge: Cambridge Univ. Press), 391  
 Buchhave, L. A., Bitsch, B., Johansen, A., et al. 2018, *ApJ*, **856**, 37  
 Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, **506**, 150  
 Butler, R. P., Marcy, G. W., Williams, E., Hauser, H., & Shirts, P. 1997, *ApJL*, **474**, L115  
 Carlberg, R. G., Dawson, P. C., Hsu, T., & Vandenberg, D. A. 1985, *ApJ*, **294**, 674  
 Chang, S. H., Gu, P. G., & Bodenheimer, P. H. 2010, *ApJ*, **708**, 1692  
 Chatterjee, S., Ford, E. B., Matsumura, S., & Rasio, F. A. 2008, *ApJ*, **686**, 580

- Chen, Di-Chang, Xie, Ji-Wei, Zhou, Ji-Lin, et al. 2023, *PNAS*, **120**, e2304179120
- Corder, G. W., & Foreman, D. I. 2009, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* (1st ed.; New York: Wiley)
- Cropper, M., Katz, D., Sartoretti, P., et al. 2018, *A&A*, **616**, A5
- D'Angelo, G., Kley, W., & Henning, T. 2003, *ApJ*, **586**, 540
- D'Angelo, G., & Lissauer, J. J. 2018, in *Handbook of Exoplanets*, ed. H. J. Deeg & J. A. Belmonte (Cham: Springer), 140
- Dawson, R. I., & Johnson, J. A. 2018, *ARA&A*, **56**, 175
- Dehnen, W., & Binney, J. J. 1998, *MNRAS*, **298**, 387
- Dressing, C. D., & Charbonneau, D. 2015, *ApJ*, **807**, 45
- Duffell, P. C., & Chiang, E. 2015, *ApJ*, **812**, 94
- Eberhardt, J., Hobson, M. J., Henning, T., et al. 2023, *AJ*, **166**, 271
- Fernandes, R. B., Mulders, G. D., Pascucci, I., Mordasini, C., & Emshenhuber, A. 2019, *ApJ*, **874**, 81
- Fischer, D. A., & Valenti, J. 2005, *ApJ*, **622**, 1102
- Ford, E. B., & Rasio, F. A. 2008, *ApJ*, **686**, 621
- Fressin, F., Torres, G., Charbonneau, D., et al. 2013, *ApJ*, **766**, 81
- Fulton, B. J., Rosenthal, L. J., Hirsch, L. A., et al. 2021, *ApJS*, **255**, 14
- Gaia Collaboration, Prusti, T., & Bruijn, J. H. J. 2016, *A&A*, **595**, A1
- Gaia Collaboration, Vallenari, A., & Brown, A. G. A. 2023, *A&A*, **674**, A1
- Gaudi, B. S., Meyer, M., & Christiansen, J. 2021, in *ExoFrontiers: Big Questions in Exoplanetary Science*, ed. N. Madhusudhan (Bristol: IOP), 2
- Ghosh, T., & Chatterjee, S. 2023, *MNRAS*, **527**, 79
- Goldreich, P., & Sari, R. 2003, *ApJ*, **585**, 1024
- Goldreich, P., & Soter, S. 1966, *Icar*, **5**, 375
- Goldreich, P., & Tremaine, S. 1980, *ApJ*, **241**, 425
- Gonzalez, G. 1997, *MNRAS*, **285**, 403
- Hamer, J. H., & Schlaufman, K. C. 2019, *AJ*, **158**, 190
- Hamers, A. S., Antonini, F., Lithwick, Y., Perets, H. B., & Portegies Zwart, S. F. 2017, *MNRAS*, **464**, 688
- Hardegree-Ullman, K. K., Cushing, M. C., Muirhead, P. S., & Christiansen, J. L. 2019, *AJ*, **158**, 75
- Harris, Charles R., Millman, K. Jarrod, Walt, Stéfan J., et al. 2020, *Natur*, **585**, 357
- Hodges, J. L. 1958, *ArM*, **3**, 469
- Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, *ApJS*, **201**, 15
- Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, *Sci*, **330**, 653
- Hsu, D. C., Ford, E. B., Ragozzine, D., & Ashby, K. 2019, *AJ*, **158**, 109
- Hunter, John D. 2007, *CSE*, **9**, 90
- Ida, S., & Lin, D. N. C. 2005, *PThPS*, **158**, 68
- Ikoma, M., Emori, H., & Nakazawa, K. 2001, *ApJ*, **553**, 999
- Jackson, B., Greenberg, R., & Barnes, R. 2008, *ApJ*, **678**, 1396
- Jackson, J. M., Dawson, R. I., Quarles, B., & Dong, J. 2023, *AJ*, **165**, 82
- Johansen, A., & Lambrechts, M. 2017, *AREPS*, **45**, 359
- Kornet, K., Bodenheimer, P., Różyczka, M., & Stepinski, T. F. 2005, *A&A*, **430**, 1133
- Kozai, Y. 1962, *AJ*, **67**, 591
- Kruskal, W. H., & Wallis, W. A. 1952, *JASA*, **47**, 583
- Kunimoto, M., & Bryson, S. 2021, *AJ*, **161**, 69
- Labadie-Bartz, J., Rodriguez, J. E., Stassun, K. G., et al. 2019, *ApJS*, **240**, 13
- Lai, D. 2012, *MNRAS*, **423**, 486
- Lee, E. J., & Chiang, E. 2017, *ApJ*, **842**, 40
- Lee, E. J., Chiang, E., & Ormel, C. W. 2014, *ApJ*, **797**, 95
- Lidov, M. L. 1962, *P&SS*, **9**, 719
- Lin, D. N. C., & Papaloizou, J. 1986, *ApJ*, **309**, 846
- Lubin, J., Wang, X.-Y., Rice, M., et al. 2023, *ApJL*, **959**, L5
- Maldonado, J., Villaver, E., & Eiroa, C. 2018, *A&A*, **612**, A93
- Mann, H. B., & Whitney, D. R. 1947, *Ann. Math. Stat.*, **18**, 50
- Manoj, P., & Bhatt, H. C. 2005, *A&A*, **429**, 525
- Mayor, M., & Queloz, D. 1995, *Natur*, **378**, 355
- Meusinger, H. 1991, *AN*, **312**, 231
- Miyazaki, S., & Masuda, K. 2023, *AJ*, **166**, 209
- Mordasini, C., Alibert, Y., Benz, W., & Naef, D. 2008, in *ASP Conf. Ser.* 398, *Extreme Solar Systems*, ed. D. Fischer et al. (San Francisco, CA: ASP), 235
- Mordasini, C., Alibert, Y., Benz, W., Klahr, H., & Henning, T. 2012, *A&A*, **541**, A97
- Mulders, G. D. 2018, in *Handbook of Exoplanets*, ed. H. Deeg & J. Belmonte (Cham: Springer)
- Mulders, G. D., Pascucci, I., & Apai, D. 2015, *ApJ*, **798**, 112
- Muñoz, D. J., Lai, D., & Liu, B. 2016, *MNRAS*, **460**, 1086
- Mustill, A. J., Lambrechts, M., & Davies, M. B. 2022, *A&A*, **658**, A199
- Narang, M. 2022, PhD thesis, Tata Institute of Fundamental Research, Mumbai
- Narang, M., Manoj, P., Furlan, E., et al. 2018, *AJ*, **156**, 221
- Narang, M., Oza, A. V., Hakim, K., et al. 2023, *MNRAS*, **522**, 1662
- NASA Exoplanet Archive 2022, *Planetary Systems Composite Parameters* NexSci-Caltech/IPAC, doi:10.26133/NEA13
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, *A&A*, **418**, 989
- Okuzumi, S., Tanaka, H., Kobayashi, H., & Wada, K. 2012, *ApJ*, **752**, 106
- Pecaut, M. J., & Mamajek, E. E. 2013, *ApJS*, **208**, 9
- Penev, K., Bouma, L. G., Winn, J. N., & Hartman, J. D. 2018, *AJ*, **155**, 165
- Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2018, *AJ*, **155**, 89
- Petrovich, C. 2015, *ApJ*, **799**, 27
- Piso, A.-M. A., & Youdin, A. N. 2014, *ApJ*, **786**, 21
- Piso, A.-M. A., Youdin, A. N., & Murray-Clay, R. A. 2015, *ApJ*, **800**, 82
- Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, *Icar*, **124**, 62
- Poon, S. T. S. 2021, PhD thesis, Queen Mary University, UK
- Powell, D., Murray-Clay, R., Pérez, L. M., Schlichting, H. E., & Rosenthal, M. 2019, *ApJ*, **878**, 116
- Rasio, F. A., & Ford, E. B. 1996, *Sci*, **274**, 954
- Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, *A&A*, **585**, A93
- Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2023, *A&A*, **674**, A29
- Reffert, S., Bergmann, C., Quirrenbach, A., Trifonov, T., & Künstler, A. 2015, *A&A*, **574**, A116
- Rosenthal, L. J., Howard, A. W., Knutson, H. A., & Fulton, B. J. 2024, *ApJS*, **270**, 1
- Santos, N. C., Adibekyan, V., Figueira, P., et al. 2017, *A&A*, **603**, A30
- Santos, N. C., Israelian, G., Mayor, M., Rebolo, R., & Udry, S. 2003, *A&A*, **398**, 363
- Santos, N. C., Mayor, M., & Israelian, G. 2006, *ISSIR*, **6**, 3
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, **403**, 1829
- Shara, M. M., Hurley, J. R., & Mardling, R. A. 2016, *ApJ*, **816**, 59
- Sharma, S., Bland-Hawthorn, J., Binney, J., et al. 2014, *ApJ*, **793**, 51
- Smirnov, N. 1948, *Ann. Math. Stat.*, **19**, 279
- Socrates, A., Katz, B., Dong, S., & Tremaine, S. 2012, *ApJ*, **750**, 106
- Soderblom, D. R. 2010, *ARA&A*, **48**, 581
- Swastik, C., Banyal, R. K., Narang, M., et al. 2021, *AJ*, **161**, 114
- Swastik, C., Banyal, R. K., Narang, M., et al. 2022, *AJ*, **164**, 60
- Swastik, C., Banyal, Ravinder K., Narang, Mayank, Unni, Athira, & Sivarani, T. 2024, *AJ*, **167**, 270
- Takeda, G., Ford, E. B., Sills, A., et al. 2007, *ApJS*, **168**, 297
- Taylor, M. B. 2005, in *ASP Conf. Ser.* 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Teyssandier, J., Lai, D., & Vick, M. 2019, *MNRAS*, **486**, 2265
- Udry, S., & Santos, N. C. 2007, *ARA&A*, **45**, 397
- Ujjwal, K., Kartha, S. S., Mathew, B., Manoj, P., & Narang, M. 2020, *AJ*, **159**, 166
- Unni, A., Narang, M., Sivarani, T., et al. 2022, *AJ*, **164**, 181
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wang, R., Luo, A. L., Chen, J.-J., et al. 2020, *ApJ*, **891**, 23
- Waskom, Michael L. 2021, *JOSS*, **6**, 3021
- Wielen, R. 1977, *A&A*, **60**, 263
- Wolthoff, V., Reffert, S., Quirrenbach, A., et al. 2022, *A&A*, **661**, A63
- Wyatt, M. C., Clarke, C. J., & Greaves, J. S. 2007, *MNRAS*, **380**, 1737
- Xiang-Gruess, M. 2016, *MNRAS*, **455**, 3086
- Yang, J.-Y., Xie, J.-W., & Zhou, J.-L. 2020, *AJ*, **159**, 164
- Yu, J., & Liu, C. 2018, *MNRAS*, **475**, 1093
- Zakamska, N. L., Pan, M., & Ford, E. B. 2011, *MNRAS*, **410**, 1895
- Zhu, W., & Dong, S. 2021, *ARA&A*, **59**, 291