

Published in final edited form as:

Nature. 2020 October 01; 586(7830): 528–532. doi:10.1038/s41586-020-2800-0.

Stellar clustering shapes the architecture of planetary systems

Andrew J. Winter^{1,2,✉}, J. M. Diederik Kruijssen¹, Steven N. Longmore³, Mélanie Chevance¹

¹Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Heidelberg, Germany

²School of Physics and Astronomy, University of Leicester, Leicester, UK

³Astrophysics Research Institute, Liverpool John Moores University, Liverpool, UK

Abstract

Planet formation is generally described in terms of a system containing the host star and a protoplanetary disk^{1–3}, of which the internal properties (for example, mass and metallicity) determine the properties of the resulting planetary system⁴. However, (proto)planetary systems are predicted^{5,6} and observed^{7,8} to be affected by the spatially clustered stellar formation environment, through either dynamical star–star interactions or external photoevaporation by nearby massive stars⁹. It is challenging to quantify how the architecture of planetary systems is affected by these environmental processes, because stellar groups spatially disperse within less than a billion years¹⁰, well below the ages of most known exoplanets. Here we identify old, co-moving stellar groups around exoplanet host stars in the astrometric data from the Gaia satellite^{11,12} and demonstrate that the architecture of planetary systems exhibits a strong dependence on local stellar clustering in position-velocity phase space. After controlling for host stellar age, mass, metallicity and distance from the star, we obtain highly significant differences (with *p* values of 10^{−5} to 10^{−2}) in planetary system properties between phase space overdensities (composed of a greater number of co-moving stars than unstructured space) and the field. The median semi-major axis and orbital period of planets in phase space overdensities are 0.087 astronomical units and 9.6 days, respectively, compared to 0.81 astronomical units and 154 days, respectively, for planets around field stars. ‘Hot Jupiters’ (massive, short-period exoplanets) predominantly exist in stellar phase space overdensities, strongly suggesting that their extreme orbits originate from environmental perturbations rather than internal migration^{13,14} or planet–planet scattering^{15,16}. Our findings reveal that stellar clustering is a key factor setting the architectures of planetary systems.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

[✉]Correspondence and requests for materials should be addressed to A.J.W. andrew.winter@uni-heidelberg.de.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author contributions A.J.W. led the study, developed the analysis method, and performed the analysis, with contributions from J.M.D.K. and S.N.L. A.J.W. and J.M.D.K. wrote the text, with contributions from S.N.L. and M.C. J.M.D.K. and M.C. developed the initial idea for the project. All authors contributed to aspects of the analysis and the interpretation of the results.

Competing interests The authors declare no competing interests.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

We measure the six-dimensional (6D; position-velocity) phase space densities of stars in the immediate vicinity of exoplanet host stars, using data from Gaia's second data release (DR2)^{11,12}, to quantify whether the environment in which planetary systems form and evolve affects planet properties such as orbit, mass and radius. Phase space overdensities can have different origins. At birth, stars are clustered in position-velocity space^{17,18}. The spatial overdensities in which stars form disperse dynamically on timescales of 10^1 – 10^3 Myr, depending on the overdensity's mass and gravitational boundedness, but the stars can remain clustered in velocity space as 'co-moving groups' over several billion years¹⁰. Velocity clustering can also be generated at a later age by galactic dynamics^{19,20}. Without additional constraints, it is not possible to establish the precise origins of phase space overdensities. By measuring whether exoplanet host stars reside in a phase space overdensity, we therefore assess their proximity in phase space to other stars at present. Any systematic trends of planetary properties with overdensity membership will trace an environmental impact either on planet formation (for example, by external photoevaporation of the protoplanetary disk) or evolution (for example, by later dynamical perturbation of the planetary system).

To determine whether exoplanet host stars reside in phase space overdensities, we first match each confirmed exoplanet in the NASA Exoplanet Archive²¹ to its host star in the Gaia DR2 data. Calculating a 6D phase space density requires radial velocity data, which is available in Gaia DR2 for the host stars of 1,525 out of 4,141 confirmed exoplanets (May 2020). We calculate the local 6D phase space densities of these host stars and their stellar neighbours using the 'Mahalanobis distance', which expresses separations in a heterogeneous, multi-dimensional phase space (here generated by combining positions and velocities). We summarize the procedure here and refer to the Methods for details. We first define a subset of up to 600 randomly drawn stars within 40 pc of (and in addition to) the exoplanet host. For each of these, we calculate the local relative phase space density by (i) measuring the Mahalanobis distances to all stars within 40 pc from that star, (ii) determining the 6D volume subtended by the Mahalanobis distance to its 20th-nearest neighbour, (iii) inverting the 6D volume to get the phase space density and (iv) dividing the phase space density by the median phase space density of all drawn stars. Because we use only relative phase space densities around the exoplanet host, the result is insensitive to the number of neighbours or the sample's completeness.

We first calculate the probability P_{null} that the phase space density distribution is drawn from a single lognormal probability density function (PDF; see Methods). For exoplanet hosts with $P_{\text{null}} < 0.05$ (1,493 out of 1,525), we distinguish 'low-density' and 'high-density' environments by performing a double-lognormal decomposition of the phase space density distribution around the host star (Fig. 1; see also Extended Data Fig. 1). We use this decomposition to define P_{low} and $P_{\text{high}} = 1 - P_{\text{low}}$, which are the probabilities that an exoplanet host is associated with an environment of low or high phase space density, respectively. If the clustered environment in which a star is born or evolves affects its planetary system, then planets orbiting stars in overdensities may exhibit different properties than those around true 'field stars', by which we refer to stars in the component with low phase space density.

We illustrate our method for distinguishing high- and low-density stars in Fig. 1 for two exoplanet hosts (HD 175541 and WASP-12) that are typical examples of a field star and of one occupying an overdensity. Both relative phase space density PDFs exhibit significant deviations from a single lognormal distribution (with $P_{\text{null}}=2.8 \times 10^{-5}$ and 6.9×10^{-9} , respectively). The overdensity is not visible in the spatial distribution of stars (Fig. 1c, d), but the velocity distributions do exhibit structure (Fig. 1e, f). The low-density host (HD 175541) is orbited by a $194M_{\oplus}$ (where M_{\oplus} is the mass of the Earth) planet with a period of 297.3 days (ref.²²). By contrast, the high-density host (WASP-12) is orbited by a ‘hot Jupiter’ of mass $448M_{\oplus}$ and period 1.09 days (ref.²³). These examples are representative: we find that hot Jupiters predominantly exist in high-density environments.

We investigate the statistical differences in properties of exoplanets orbiting stars in high- and low-density environments by splitting the sample using 1σ threshold probabilities of $P_{\text{low}} > 0.84$ and $P_{\text{high}} > 0.84$. We omit stars with insufficient neighbours (<400 ; see Methods) or ambiguous relative phase space densities ($0.16 < P_{\text{high}} < 0.84$). These threshold probabilities form a compromise between obtaining a large sample and minimizing the number of falsely categorised stars. Because exoplanet architectures correlate with the mass and age of the host star^{24,25}, we make further cuts in the stellar ages (1–4.5 Gyr) and stellar masses ($0.7M_{\odot}$ – $2M_{\odot}$, where M_{\odot} is the solar mass) to ensure that the low- and high-density subsamples have similar distributions in these properties. We provide a physical motivation for these limits in the Methods and demonstrate that our conclusions are unaffected by any of the above cuts in Extended Data Figs. 2 and 3. After making these cuts, we obtain 66 low-density hosts and 319 high-density hosts. These numbers do not necessarily imply that low-density exoplanet host stars are less common, because the sample is not complete (see Methods).

To assess the impact of stellar clustering on the observed planet population, we first compare the distributions of planet masses and semi-major axes in the low- and high-density samples (Fig. 2). The most prominent difference between both environments is the abundance of hot Jupiters, defined by masses $M_p > 50M_{\oplus}$ and semi-major axes $a_p < 0.2$ AU. We find that hot Jupiters are rare around field stars (Fig. 2a), constituting just $13.1 \pm 4.9\%$ of the detected planets, whereas they represent $30.4 \pm 3.5\%$ of the planets in phase space overdensities (Fig. 2b). Of all the hot Jupiters with unambiguous environment classifications ($P_{\text{low}} > 0.84$ or $P_{\text{high}} > 0.84$), $92.4 \pm 0.7\%$ are orbiting host stars in phase space overdensities.

Figure 2 suggests that the extreme orbits of hot Jupiters originate from environmental perturbations, either due to the destructive effect of external photoevaporation on the protoplanetary disks where they formed, or by dynamical interactions with nearby stars. Chemical abundance studies indicate that hot Jupiters orbiting Sun-like stars formed at larger separations than their current semi-major axes²⁶. Although this observation does not directly probe the influence of external photoevaporation, it suggests that dynamical interactions with nearby stars induce migration and produce hot Jupiters.

An externally induced origin for hot Jupiters would have important implications for the initial properties of unperturbed planetary systems, because the existence of hot Jupiters has been used to suggest that planet formation in isolation is a fundamentally disordered and

chaotic process²⁷. With the exception of two low-mass outliers and eight hot Jupiters (which could belong to a multiple stellar system or a past overdensity), detected planets around field stars are approximately distributed as $M_p \propto a_p^{1.5}$, with a dispersion of about 0.4 dex (Fig. 2a and discussion in the Methods). Exoplanet surveys are incomplete towards low planet masses and large semi-major axes, so this relationship might not persist in a complete sample. In either case, Fig. 2 suggests that the processes driving the formation and evolution of planetary systems in isolation may not represent the dominant formation channel of hot Jupiters. This changes the target outcome for models describing the initial architecture of planetary systems forming in isolation.

We perform a systematic statistical analysis to quantify how the properties of planets and their hosts differ between our low- and high-density samples. We additionally assess whether these differences may be caused by any systematic biases in host star properties. Figure 3 shows normalized cumulative distribution functions of the planet semi-major axis, period, eccentricity, mass, radius, and density (Fig. 3a–f), as well as the host star mass, metallicity, age, and distance (Fig. 3g–j). For each property, we determine the probability p_{KS} that the low- and high-density samples are drawn from the same distribution by performing a two-tailed Kolmogorov–Smirnov test.

The most significant dependence on host phase space density is found for the orbital semi-major axis and period, with $p_{KS} = 6.8 \times 10^{-5}$ and 4.8×10^{-5} , respectively (Fig. 3a, b). Field stars have a dearth of planets at small semi-major axes ($a_p < 0.1$ astronomical units, AU) and short periods (< 20 days), resulting in median values of 0.81 AU and 154 days, compared to 0.087 AU and 9.6 days for planets in overdensities. The small semi-major axes and short periods of planets orbiting stars in overdensities could result from radiative or dynamical perturbations by stellar neighbours, either during or after planet formation. Additionally, planets in overdensities have significantly (Kolmogorov–Smirnov test $p_{KS} = 1.2 \times 10^{-3}$; Fig. 3c) smaller eccentricities (median $e = 0.062$) compared to those orbiting field stars (median $e = 0.16$). This could be due to tidal circularization²⁸, because planets in overdensities have shorter periods, or due to different growth, migration and encounter histories (also see Methods). Finally, the overall distributions of planet properties themselves exhibit little dependence on host phase space density. Planets with host stars in high-density environments do have lower masses than those in low-density environments ($p_{KS} = 1.1 \times 10^{-2}$; Fig. 3d; see also Extended Data Fig. 4). Planet radii exhibit a similar trend at minimal significance ($p_{KS} = 7.1 \times 10^{-2}$; Fig. 3e), such that the planet densities do not exhibit a significant environmental dependence at all (Fig. 3f).

The relation between planetary properties and the host stellar phase space density does not result from an underlying bias or variation of the host stellar properties. We find that the distributions of host stellar mass (Fig. 3g), metallicity (Fig. 3h) and age (Fig. 3i) do not differ significantly between field stars and hosts in overdensities. However, the median host in the field is a factor of two closer to the Sun than than the median in overdensities (Fig. 3j). To determine whether this reflects a bias of the spatial distribution of low- and high-density hosts (which could either be physical or a selection bias of our method), we construct a control sample by drawing a star at random within a 40-pc radius of each exoplanet host star. We then redefine P_{low} and P_{high} based on the phase space density of this random neighbour

instead of the host and use these to split the sample into ‘control set’ cumulative distribution functions at low and high densities. We repeat this 100 times and include the resulting cumulative distribution functions and their corresponding values of p_{KS} in Fig. 3a–f.

The pairs of control samples in all panels of Fig. 3 are statistically indistinguishable, with $p_{KS} > 0.27$ for all six planet properties, confirming that the differences between low and high phase space densities do not result from a spatial bias. This means that the difference between the distance distributions of low- and high-density hosts (Fig. 3j) results from covariance with exoplanet properties rather than the other way around. For instance, hot Jupiters can be detected out to larger distances, and preferentially reside in overdensities. When restricting the sample to exoplanet hosts within 300 pc, where the distance distributions are similar for low- and high-density hosts, we find that the differences in exoplanet architectures persist (Extended Data Fig. 5). We also demonstrate that neither the host star kinematics (Extended Data Fig. 6) nor the host star mass, metallicity and age (Extended Data Fig. 7) depend on the exoplanet detection method, and differences in exoplanet architectures persist when splitting the sample by discovery method (Extended Data Fig. 8). Finally, the differences in planet properties between low- and high-density environments increase when controlling for the differences in stellar host properties (compare Fig. 3 and Extended Data Fig. 3), which argues against a kinematic or detection bias. Therefore, our findings result from physical differences in the environments of exoplanet host stars.

Our sample is dominated by relatively massive planets (with a median mass of about $200M_{\oplus}$, or nearly one Jupiter mass) as a result of current detection limits. Future observatories are required to determine how stellar clustering affects low-mass planets at large orbital separations. We expect our results to extend to the low-mass planet population, because the dynamical stability and architecture of planetary systems is often dominated by the orbital properties of their most massive members²⁹. Given that the environment affects planetary orbits and masses, it is plausible that the atmospheric composition and chemistry of planets may be affected too³⁰. The key question is which physical mechanisms drive the differences in exoplanet properties between high- and low-density environments, and at what evolutionary stage they operate. External photoevaporation can rapidly truncate protoplanetary disks, reduce their masses and curtail planet formation early⁵. However, planets may also be scattered by stellar encounters until long after they formed⁶. A combination of both effects may be required to explain the differences reported here⁹ (see Methods).

Our results show that stellar clustering is a key factor setting the architectures of planetary systems. Although it has been mostly overlooked in models and observational surveys of the planet population, the environment represents a fundamental axis along which exoplanetary and atmospheric properties may vary, with possible implications for planetary habitability and the likelihood of life in the Universe. Here we considered relatively young planetary systems with ages of 1–4.5 Gyr to detect phase space overdensities with Gaia. However, star formation is likely to have been more clustered in the past¹⁷, so that the impact of the environment on older planetary systems may have been even greater. As a result, planetary systems forming today may not be representative precursors of the observed population of exoplanetary systems. Future work should also target the evolved planet population to

quantify the impact of environmental processes. To enable these and other future efforts, we provide our ambient phase space density classification for all known exoplanets in the Supplementary Information.

Methods

The goal of our analysis is to determine whether the observed properties of exoplanetary systems depend on the degree of stellar clustering in the environment of the host star. The environmental factor most likely to affect exoplanet properties is the proximity to other stars during planet formation or evolution, primarily through external photoevaporation, chemical enrichment and/or dynamical encounters. Young stellar systems are clustered in both kinematic and spatial coordinates, and are therefore easily identified as ‘overdensities’ in phase space, and this early environment can therefore be directly quantified. However, most known exoplanets are hosted by stars much older than about 1 Gyr. By this time, most stellar systems have spatially dispersed through dynamical interactions, making it difficult to distinguish them from field stars and infer their natal environment by conventional methods.

Here we have applied a statistical method to overcome this problem, in which we identify local, 6D (velocity and spatial) phase space overdensities relative to the field star population in a given neighbourhood of the Galaxy. For an old or initially low-density stellar population that does not currently exhibit spatial or kinematic substructure, the distribution of phase space densities is expected to follow a lognormal distribution. Identifying the (unstructured) field star population therefore provides a reference population against which to identify phase space perturbations (overdensities). These overdensities may either originate from Galactic perturbations, or represent a relic of the initial stellar clustering at the time of formation. In the latter case, if the formation environment plays a role in planet formation, stars in phase space overdensities may host exoplanets that differ substantially from those orbiting field stars. By comparing properties of exoplanets in regions of high and low stellar phase space density, we aimed to determine whether the environment affects the formation and evolution of planetary systems.

Observational data

To obtain the most up-to-date properties of exoplanets and their host stars, we downloaded the Composite Planet Data catalogue from the NASA Exoplanet Archive²¹ (May 2020). We used the Gaia DR2^{11,12} catalogue to determine the position and velocity phase space information for exoplanet host stars and their surrounding stellar neighbours. Out of all 4,141 exoplanet host stars in the catalogue, we conducted our analysis on the 1520 hosts with full 6D phase space information in Gaia DR2. Typical Gaia DR2 uncertainties in position (or parallax), proper motion and velocity for these stars are 0.04 mas, 0.06 mas yr⁻¹ and 0.3 km s⁻¹, respectively. Even for the highest-density neighbourhoods analysed here, we ended up identifying phase space structures separated by about 10 pc and $\gtrsim 3$ km s⁻¹. These phase space structures are typically much larger than the astrometric uncertainties. In the rare situation that one dimension does have uncertainties comparable to the scale of the phase space structures (this mostly corresponds to parallax uncertainties at the largest distances in our sample), this disperses any overdensity along this dimension, such that in

the extreme case it is effectively excluded from the identification of substructure in our algorithm (see below). Therefore, the astrometric uncertainties do not affect our results.

Density metric

To quantify the phase space density of stars surrounding exoplanet hosts, we require a metric for distance in 6D that takes into account the mixed nature of distance and velocity units and additionally avoids geometrical effects arising from projected positions and velocities on the plane of the sky. We therefore re-project the spatial and velocity coordinates of all stars into a 6D Cartesian system and use the covariance matrix, C , of the phase space coordinates of all stars within 80 pc of the host star (we motivate this distance cut below) to calculate the Mahalanobis distance (d_M) between any two points (defined by the vectors \mathbf{x} and \mathbf{y}) in 6D phase space as

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T C^{-1} (\mathbf{x} - \mathbf{y})}. \quad (1)$$

By normalizing the vectors using the covariance matrix, their elements become unitless, so that the Mahalanobis distance in the resulting, transformed system space can be consistently defined across all six dimensions. The use of the covariance matrix accounts for deviations from sphericity.

We next define the size of the region over which the Mahalanobis distances between stars are calculated, which we refer to as a ‘neighbourhood’. We require a region large enough to contain sufficient stars to obtain a statistically representative phase space density distribution, yet small enough to (i) eliminate systematic gradients in the phase space distribution across the neighbourhood (for example, under the influence of galactic dynamics) and (ii) be justified computationally, given that the time needed to calculate the distances between all pairs scales as the search radius to the sixth power. To determine how many stellar neighbours are needed to characterize the phase space density distribution, we replicate our analysis using synthetic stellar populations (see below for details). This experiment shows that overdensities and the field are robustly delineated for phase space distributions with a sample size of $S = 600$ stars, but we find that they are unreliable for $S < 400$. Using the Gaia DR2 data, we find that nearly all (1,404 of 1,525) exoplanet host stars have neighbourhoods with $S > 400$ within a radius of about 40 pc, and most have $S \gg 600$. A radius of 40 pc is much smaller than the typical features introduced by galactic dynamics, such as spiral arms and resonances^{19,20,31}, and also represents a computationally viable sample of stars. Therefore, we define the ‘neighbourhood’ of a given exoplanet host star to refer to a spherical region of radius 40 pc centred on the host coordinates in Gaia DR2 (because this radius cut is applied around each star in the neighbourhood, the covariance matrix must be calculated using neighbours out to twice that distance from the exoplanet host star). This choice also has the advantage that the diameter of the neighbourhood is smaller than the typical separation of independent star-forming regions (≈ 100 pc, refs.^{32,33}), such that young regions are unlikely to contaminate the stellar phase space distributions. We exclude host stars that have fewer than 400 neighbouring stars with 6D phase space information within their 40-pc neighbourhood. Finally, we find that constructing the phase

space distributions using more than $S = 600$ stars within the neighbourhood does not much change our results. We therefore apply a ceiling of 600 neighbours, selected at random from the neighbourhood, such that the phase space density distributions are calculated using $S = 400\text{--}600$ for all host stars.

For each exoplanet host and for each of its $S \approx 600$ stellar neighbours, we calculate the Mahalanobis distances to every other star within 40 pc that has 6D astrometric data. We calculate the local phase space density of each host and its neighbours by using the Mahalanobis distances to their M th nearest neighbours ($d_{M,N}$). The resulting ‘Mahalanobis density’ is defined as

$$\rho_{M,N} = N d_{M,N}^{-D}, \quad (2)$$

where $D = 6$ is the number of dimensions, and N is the number of neighbours used to calculate the corresponding density. In other words, the phase space density is N divided by the 6D Mahalanobis volume required to contain the N nearest neighbours. The number of neighbours N should be chosen such that it reaches a compromise between minimising Poisson noise and achieving sufficient density contrast to identify overdensities. Following previous work quantifying phase space substructure^{34,35}, we use $N = 20$, but emphasize that our conclusions are not sensitive to changes in N . We then normalize $\rho_{M,N}$ to compare relative phase space densities, $\tilde{\rho}_{M,N}$, across regions:

$$\tilde{\rho}_{M,N} = \frac{\rho_{M,N}}{\rho_{M,N,\text{med}}}, \quad (3)$$

where $\rho_{M,N,\text{med}}$ is the median $\rho_{M,N}$ of all S stars in the neighbourhood of each individual exoplanet host star.

Density distribution

Having defined the phase space densities for the subset of S stars in a neighbourhood around each exoplanet host, we separate them into subgroups of high and low phase space density. We do this by decomposing the distribution of phase space densities. To illustrate the concept behind the method, we first consider a neighbourhood without any spatial substructure and with stellar velocities that are well described by a single velocity dispersion σ_v . The resulting distribution of $\tilde{\rho}_{M,N}$ will be lognormal. If we now introduce a subset of stars with a smaller velocity dispersion than this ‘field population’, as would be expected for a co-moving group, this will generate a second lognormal distribution at the high-density end of the $\tilde{\rho}_{M,N}$ distribution (which may manifest itself as a slight excess, depending on the relative numbers of stars).

This principle is demonstrated for synthetic data in Extended Data Fig. 1. We randomly draw positions and velocities from two independent distributions. One is a ‘background’ population, with an isotropic spatial density. The velocity vector \mathbf{v}_j of each star j is defined by drawing its components $v_{j,a} = \mathbf{v}_j \cdot \mathbf{e}_a$ (where \mathbf{e}_a are orthogonal unit vectors for $a = 1, 2, 3$) from a normal distribution with a dispersion σ_v . We then draw a ‘perturbed’ set with the same isotropic spatial density distribution, scaled by a multiplicative factor $\delta\rho$ relative to the

background distribution. The velocity distribution of these stars again follows a normal distribution, with a dispersion scaled by a multiplicative factor $\delta\rho_v$ relative to the background distribution. Since the Mahalanobis distance is calculated by normalizing to the covariance matrix, the absolute values of the velocity dispersions and spatial densities are not relevant. As motivated above, we perform the phase space density calculation for 600 stars, so we must allow at least this many within the volume (in arbitrary units). We define the volume by requiring approximately 1,000 stars in the background distribution (as a result, the total number of stars within the same volume that belong to the perturbed set depends on $\delta\rho$). The resulting histograms of the relative phase space densities of the two populations are shown in Extended Data Fig. 1.

To split stars into high- and low-phase space density groups, we must first determine whether the $\tilde{\rho}_{M,N}$ distribution within a neighbourhood exhibits a significant deviation from a single lognormal. We use the Python implementation for Gaussian mixture modelling, GAUSSIAN-MIXTURE in SCIKIT-LEARN³⁶. We first fit a single, lognormal PDF to the distribution of phase space densities in a given neighbourhood. We then use a Kolmogorov–Smirnov (KS) test to calculate the probability P_{null} that the observed $\tilde{\rho}_{M,N}$ distribution is drawn from the best-fitting lognormal PDF. If P_{null} is low, then the $\tilde{\rho}_{M,N}$ distribution is not well described by a single lognormal. We take $P_{\text{null}} = 0.05$ as a threshold above which we do not attempt to decompose high-density and low-density groups, but the results are insensitive to small variations of P_{null} . This requirement excludes neighbourhoods exclusively composed of field stars, or where phase space perturbations are not detectable against the background. We quote the values of P_{null} for the synthetic stellar populations in Extended Data Fig. 1 and for some examples of real exoplanet host stars in Extended Data Fig. 9. In Extended Data Fig. 9, we also show the best-fitting single lognormal distribution as a red line, from which we derive P_{null} .

For neighbourhoods around exoplanet host stars where the phase space density distribution deviates from a single lognormal ($P_{\text{null}} < 0.05$), we decompose the population into low- and high-density components. The low-density component must correspond to the unstructured ‘ground state’ field star population, which we find to be always well described by a lognormal phase space density distribution. We find that the additional (perturbed) component is also well described by a lognormal functional form. Therefore, we again use GAUSSIANMIXTURE modelling to fit two independent lognormal functions to the phase space density distributions. The population that exhibits enhanced phase space densities (consistent with substructure) may be composed of multiple co-moving groups. Despite this potential heterogeneous nature, we find that the high-density component is always clearly distinguishable from the lower phase space density population and it is not necessary to distinguish individual co-moving groups for the following analysis.

From the best-fitting, double-lognormal PDF (composed of PDFs p_1 and p_2), we estimate the probability that a star at a given phase space density is a component of the high- or low-density population. We define p_1 and p_2 such that they correspond to the low- and high-density population, respectively:

$$\frac{\int p_1(\tilde{\rho}) \tilde{\rho} d\tilde{\rho}}{\int p_1(\tilde{\rho}) d\tilde{\rho}} < \frac{\int p_2(\tilde{\rho}) \tilde{\rho} d\tilde{\rho}}{\int p_2(\tilde{\rho}) d\tilde{\rho}}. \quad (4)$$

The probabilities of a star being a member of the low- and high-density components are then

$$P_{\text{low}}(\tilde{\rho}) = \frac{p_1}{p_1 + p_2} \text{ and } P_{\text{high}}(\tilde{\rho}) = \frac{p_2}{p_1 + p_2}, \quad (5)$$

such that $P_{\text{high}} = 1 - P_{\text{low}}$. Before fitting, we remove stars outside of the 2σ range in phase space densities. This cut has no influence on the fit, except in cases where individual outliers would otherwise lead to spurious fitting results. We also remove stars with phase space densities $\tilde{\rho}_{M, N} > 50$, because we find empirically that such high relative phase space densities are associated with gravitationally bound clusters. For stars in such environments, we define $P_{\text{high}} = 1$.

Choice of threshold probability

Throughout our analysis, we adopt a threshold probability of $P_{\text{th}} = 0.84$ to delineate stars into low and high phase space densities. This is a compromise between obtaining a sufficiently large sample size and minimising the number of stars that are falsely categorized. We show in Extended Data Fig. 2 how the median orbital period, eccentricity, and planet mass in low- and high-density environments depend on the choice of P_{th} . We find that differences in the medians between the two samples persist across a wide range of P_{th} . We conclude that the differences between the distributions of exoplanet properties that we identify are not sensitive to this choice.

Robustness and nature of low- and high-density phase space structures

The phase space density classification of stars is not always unambiguous. This occurs for two reasons. First, we find that stars in the field have a broader phase space density distribution than stars in overdensities. This means that a larger fraction of stars in overdensities fall within the range of phase space densities spanned by field stars than vice versa (see for example, the purple overlapping bars in Extended Data Fig. 1). This asymmetry means that field stars can often be identified with greater confidence than those in overdensities. Second, a star belonging to either of the components (field or overdensity) can have a large number of nearest neighbours that actually belong to the other component. As a result, these contaminants contribute to the inferred phase space density of that star. Low-density contaminants neighbouring a star in an overdensity do not alter the phase space density of that star, because they are statistically less numerous than the high-density neighbours. By contrast, high-density contaminants neighbouring a field star can drive up the local phase space density. This results in enhanced phase space densities of a subset of stars in the field. In turn, this affects the decomposition of the phase space density distribution into low- and high-density components by boosting the number of stars in the overdensity (see Extended Data Fig. 1g). For the affected stars, this translates into a corresponding overestimate of P_{high} . Since our aim is to separate out two populations, not to assign individual stars accurately to specific stellar groups, potentially overestimating P_{high}

is not a problem in a statistical sense. We are still able to obtain one sample which preferentially contains field stars and another that preferentially contains stars in overdensities. We minimize the above effects by requiring $P_{\text{high}} > 0.84$ for membership of an overdensity (see below).

Finally, the overdensities that we identify should not be interpreted as monolithic, co-moving groups. In a given neighbourhood, the presence of multiple co-moving groups can yield deviations from Gaussianity in the local stellar velocity distribution³⁷. We interpret stars in an overdensity as phase space neighbours and plausible members of such groups. It is likely that at least some of the stars in an overdensity were born together³⁸. By contrast, stars occupying an environment of low phase space density are the least likely to have neighbours with which they were born. In this way, our division reflects the fact that kinematic substructure may persist after the spatial substructure (that is, cluster) has dispersed. Future studies aiming to assign individual stars accurately to specific co-moving groups should include information such as stellar ages and chemical abundances, in addition to the 6D astrometric data used here.

Persistence of kinematic structure

Bound stellar groups (clusters) typically retain their spatial structure for $\lesssim 1$ Gyr (refs. 10,39,40). It is less clear what the dissolution timescale is of their local kinematic substructure, which represents a ‘memory’ of the formation environment. Many unbound co-moving groups, that are correlated both spatially and kinematically, occupy the solar neighbourhood^{41–43}. Such groups may be composed of stars that are co-natal⁴⁴, or may be the result of resonant perturbation due to the galactic potential^{19,20,45,46}. The dependence of planetary system architecture on the local phase space density that we identify in this work suggests that at least some degree of kinematic substructure has a co-natal origin (although it does not preclude the role of galactic perturbations). In support of the hypothesis that stars may retain nearest neighbours from their formation environment in 6D phase space, simulations that follow star formation and the subsequent orbital evolution of the stars within the host galaxy indicate that formation neighbours persist after the dispersal of the bound cluster⁴⁷. In these simulations, most of the resulting co-moving neighbours are younger than 1 Gyr, but beyond that time, a roughly constant number of pairs of such neighbours persist for several billions of years. These long-lived phase space overdensities originate mostly from the most massive clusters. This finding is observationally supported by the fact that co-moving stellar pairs (and ‘networks’ of such pairs, that is, groups) are common even at separations of tens of parsecs⁴⁸ and that the vast majority of these pairs exhibit similar metallicities³⁸. Observationally, dynamical heating mechanisms appear to affect stars in the Milky Way on timescales of about 4.5 Gyr (ref.³⁷), which therefore represents an upper limit on the time for which co-moving pairs (groups) are likely to persist. This is confirmed by the fact that we observe a pronounced drop of the age distribution of exoplanet host stars in overdensities at that age (Extended Data Fig. 10). This motivates our use of a maximum age of 4.5 Gyr in our analysis.

Planetary system formation timescale

In addition to the above upper limit on the ages of systems we wish to compare, we can define a similar lower limit set by the time over which an isolated exoplanetary system forms and reaches a stable configuration. A protoplanetary disk of dust and gas around a stellar host largely disperses within 5–10 Myr (ref.⁴⁹), curtailing the accretion of gas onto cores, which is required for gas giant formation. However, this does not necessarily mark the end of the early assembly of planetary systems. Debris disks⁵⁰ composed of dust and planetesimals may be the site of continued giant impacts that reshape planetary systems over longer timescales²⁵. For example, the Earth itself may have taken around 100 Myr to reach its present-day mass⁵¹. Debris disks have been observed around stars with a wide range of ages. They are most common for stars younger than 120 Myr and become rare for stars older than about 1 Gyr (ref.⁵²).

Ultimately, our choice of a lower limit on the stellar host age when comparing the architectures of exoplanetary systems is mainly motivated by achieving a similar distribution of ages for stars in overdensities and the field. In this context, a sensible lower limit in age is one for which the planetary systems have reached a stable state and open clusters have spatially dissolved. Because the timescales for both of these processes are estimated at $\lesssim 1$ Gyr, we exclude exoplanet hosts younger than 1 Gyr in our analysis.

Stellar age estimates

To enable making the above age cuts to our sample of exoplanet host stars, we use the stellar age estimates quoted in the composite table of the NASA Exoplanet Archive (see Supplementary Table 1). Many of these ages have considerable associated uncertainties. We investigate how these uncertainties may influence the impact of our applied stellar age cuts on the resulting sample of host stars. To do so, we perform a Monte Carlo reassignment of each of the stellar ages by randomly drawing from a normal distribution with a median equal to the measured age and a standard deviation equal to the quoted uncertainty. For host stars without any quoted uncertainty, we randomly draw a relative uncertainty from the host stars that do have associated uncertainties. We repeat our Monte Carlo procedure 200 times and show the result in Extended Data Fig. 10. The distribution of stellar ages in overdensities ($P_{\text{high}} > 0.84$) and the field ($P_{\text{low}} > 0.84$) remains similar across all realizations. Most importantly, all realizations show the same drop of the fraction of exoplanets hosted by stars in overdensities for ages $\gtrsim 5$ Gyr, which is expected due to dynamical heating (and therefore the dispersal of kinematic substructure). Across all realizations, the number of host stars within our adopted age range of 1–4.5 Gyr is 276 ± 9 for overdensities and 61 ± 5 for the field. As discussed in the main text, we perform our analysis on 322 host stars in overdensities and 66 in the field. This 8–17% difference in sample size is too small to reasonably affect the systematic trends that we identify—all potential contaminant host stars would be required to reside at the same end of the distributions of exoplanet properties to systematically skew these distributions.

Host-star property distributions

Our fiducial analysis is restricted to exoplanet host stars with ages 1–4.5 Gyr and masses $0.7M_{\odot}$ – $2M_{\odot}$. For comparison to Fig. 3, we present the cumulative distribution functions

for all the exoplanet and host properties in Extended Data Fig. 3, but this time without applying any cuts in host stellar mass or age. We still restrict the sample to host stars that (i) have sufficient neighbours to characterize the phase space distribution ($S > 400$), (ii) have a bimodal phase space distribution ($P_{\text{null}} < 0.05$), and (iii) have known ages and masses. This results in a sample of 1,077 exoplanets, or 784 independent exoplanet hosts. When considering host star properties, we count each star only once, even if it has multiple planets, to ensure that the measurements are independent. This choice has a minimal effect on the overall distributions.

Several of the differences in exoplanet properties between low- and high-density environments identified in the analysis of our fiducial sample (Fig. 3) also stand out in Extended Data Fig. 3, particularly in the distributions of semi-major axis and orbital period. However, when considering the entire sample, the properties of the host stars also differ considerably between low- and high-density regions. For instance, the host star age distributions differ strongly. As explained above, there is an overabundance of low-density hosts at ages $\gtrsim 4.5$ Gyr due to the dispersal of overdensities, and an overabundance of high-density hosts at ages $\lesssim 1$ Gyr due to the persistence of initial stellar clustering.

The metallicity distributions in Extended Data Fig. 3 also exhibit clear differences between high- and low-density host stars. Hosts at low densities tend to have lower metallicities. This may be due to the covariance between stellar age and metallicity. It is also possible that stars in high-density regions have enhanced metallicity due to their proximity to massive stars during formation. In principle, we are unable to differentiate between these two scenarios from the samples presented here. However, we note that the Milky Way has chemically enriched at a rate of approximately 0.05 dex per Gyr for the past approximately 8 Gyr (ref. 53). In Extended Data Fig. 3i, the median ages of host stars residing in overdensities and the field are about 3 Gyr and 6.5 Gyr, respectively. The metallicity offset in panel h is consistent with this age difference, given the enrichment history of the Galactic disk.

Finally, exoplanet host stars in low-density environments are preferentially lower in mass than those in high-density environments. Again, this may be affected by covariance with age, as stellar evolution limits the mass of the most massive stars. Additional effects may be mass segregation in young stellar populations, which leads to the preferential ejection of low-mass stars⁵⁴, or a more efficient disruption of planetary systems around low-mass stars in high-density environments. The adopted stellar mass range in our fiducial sample (see below) eliminates the above excess of low-mass stars in low-density environments across the sample.

Because the host star properties (which are observed to be correlated with exoplanet properties^{24,55–57}) differ strongly between low- and high-density environments across the full sample shown in Extended Data Fig. 3, it is unclear to what extent this contributes to the differences in exoplanet properties. Throughout our analysis, we therefore restrict the sample to exoplanet host stars with ages 1–4.5 Gyr and masses $0.7M_{\odot}$ – $2M_{\odot}$. Figure 3 shows that these cuts result in similar host star properties between low- and high-density environments. In turn, this allows us to conclude that exoplanet properties are significantly different in environments of high phase space density compared to those in the field. The

relatively narrow stellar mass range of $0.7M_{\odot}$ – $2M_{\odot}$ also implies that the planetary semi-major axes and orbital periods are highly correlated (through Kepler’s law) and follow similar environmental trends (see Fig. 3).

Checking for spatial bias

We have eliminated possible sources of systematic biases in exoplanetary properties that result from covariances with the host star age, mass and metallicity. However, the distance distributions (or, more generally, the spatial distributions) also differ between the low- and high-density subsamples, both before and after applying our age and mass cuts (Fig. 3j and Extended Data Fig. 3j). These differences could introduce selection effects in the observed exoplanet sample, which could bias the planetary properties (for example, mass and orbital period). Conversely, if the architectures of planetary systems themselves differ between low- and high-density environments, then the fact that it is easier to detect massive, close-in exoplanets naturally generates environmental differences in the distance (or spatial) distribution of planets. The symmetry of this problem makes it non-trivial to determine whether the differences in exoplanet properties are due to the difference in spatial distributions, or vice versa. We therefore test whether we would obtain the same results when randomly generating a sample of host stars that are similarly distributed in space.

To establish whether the spatial distribution of the exoplanet host stars alone could be responsible for the observed relation between exoplanet properties and the ambient stellar phase space density, we construct a suite of control experiments. Each of these experiments represents a different realization of the full set of host star phase space densities. This is achieved by drawing a star at random from each host star neighbourhood and assigning its phase space density to the corresponding host star. We then repeat our analysis of the distributions of exoplanet properties for each control experiment. Because we choose one neighbour per neighbourhood, each control realization has the same size as the set of exoplanet host stars. By construction, it also has a spatial distribution similar to the set of exoplanet hosts, because we draw a neighbour from within a 40-pc radius of each host star. For each control experiment, we define low-density ($P_{\text{low}} > 0.84$) and high-density ($P_{\text{high}} > 0.84$) subsamples of the exoplanet host stars as before, this time using the phase space densities of the randomly drawn neighbours. We generate a total of 100 control experiments by repeating the above procedure 100 times.

The results of the above test are shown by faint lines in Fig. 3 and Extended Data Fig. 3. Across all panels, the exoplanet properties in low- and high-density environments differ considerably less in the control experiments than they do in the real measurement. This means that our results cannot be attributed to biases caused by the spatial distribution of host stars. In grey, we list the median p -values obtained by applying a two-tailed Kolmogorov–Smirnov test to each control experiment, with error bars indicating 16th and 84th percentiles. None of the differences are statistically significant (that is, $\log p_{\text{KS}} > -1.3$ in all cases). Nonetheless, some weak systematic trends can be identified visually. For instance, even in the control experiments, exoplanets in overdensities have slightly smaller semi-major axes. Such differences are expected for the control experiments carried out here if there exist spatial (or distance) trends both in exoplanetary properties and in the relative numbers of

neighbours in the low- and high-density components of the phase space density distribution, because these relative numbers determine the probability of assigning the host star to a low- or high-density environment in the control experiments. For instance, it is easier to detect close-in planets out to larger distances; if the fraction of neighbours in overdensities also increases with distance, this naturally produces the trend shown by the control experiments in Fig. 3a and Extended Data Fig. 3a. However, the resulting trends are not statistically significant, such that we can conclusively rule out the hypothesis that the differences in exoplanet properties are due to the difference in spatial distributions.

The above tests demonstrate that spatial bias is not responsible for our findings, but we can additionally verify whether differences persist if we restrict our samples such that they have similar distance distributions. Extended Data Fig. 5 shows the cumulative distribution functions of our fiducial sample (Fig. 3), restricted to host stars within 300 pc of the Sun. Over this distance range, the distributions of distances in the two samples are similar ($\log p_{KS} = -0.46$), and significant differences in the semi-major axis ($\log p_{KS} = -1.87$), period ($\log p_{KS} = -2.02$) and eccentricity ($\log p_{KS} = -1.80$) distributions remain (albeit at a lower statistical significance owing to the smaller sample size). We conclude that differences in the distance distributions between low and high phase space densities do not drive the differences in planetary architectures.

Other possible sources of bias

The most obvious remaining, alternative interpretation of our results is that they could be caused by a potential bias in the selection of targets between different exoplanet surveys. In particular, we find a dearth of exoplanets at small separations from the exoplanet host stars in low phase space densities. The majority of short period exoplanets are discovered by transit surveys, so a bias in the stars targeted by such surveys could be responsible for our findings. Because we find no significant correlation between high-density and low-density exoplanet properties in the control experiments (Fig. 3 and Extended Data Fig. 3), which are chosen from the neighbourhood of each exoplanet host star, any such selection biases cannot be related to the spatial distribution of survey targets. Therefore, any bias that could be responsible for our positive result must be due to a kinematic selection effect in the targets of transit surveys. It is unclear what the origin for such a bias would be, but we investigate the possibility as follows.

To understand whether a kinematic survey bias exists, we compare the magnitudes of the proper motions and the radial velocities of exoplanet hosts discovered by radial velocity and transit surveys. To make this comparison, we must also control for the distance to the source, because it is correlated with the proper motion and radial velocity. We achieve this by splitting the sample of exoplanet host stars into two sets based on the discovery method (transit and radial velocity). We then construct a sample from each set by removing elements such that both samples have the same distance distribution. The resulting distributions of proper motions and radial velocities are shown in Extended Data Fig. 6. We find that the host star kinematics exhibit no significant differences between exoplanets discovered by radial velocity or transit surveys. The absence of significant kinematic selection effects favours a physical interpretation of our results.

We additionally verify whether the properties of the stellar hosts differ between radial velocity and transit surveys. We show the result of this analysis in Extended Data Fig. 7. For all physical properties of exoplanet host stars we find the same environmental trends in subsamples divided by discovery method as we do for the full sample in Extended Data Fig. 3. We conclude that the differences in exoplanet host star properties are also physical in origin, and not the result of target selection biases in stellar metallicity, age or mass between surveys.

The result of splitting the exoplanet architectures by discovery method (transit and radial velocity) is shown in Extended Data Fig. 8. Despite the drastically reduced range of semi-major axes, differences between exoplanet architectures in overdensities and the field persist when splitting the sample by discovery method. In particular, the fraction of hot Jupiters in overdensities is enhanced relative to the field by about a factor of two in either case (16% versus 8% for radial velocity surveys, and 40% versus 23% for transit surveys). In addition, for both discovery methods, exoplanets in the field cluster around the $M_p \propto a_p^{1.5}$ trend that we identified for the full sample (Fig. 2). The fact that differences in exoplanet properties between overdensities and the field persist when controlling for discovery method indicates that survey biases cannot be responsible for our findings.

Finally, dynamical interactions in stellar multiples may perturb planetary systems^{58–61}. If the multiplicity fraction were to differ between stars in overdensities and the field, this could potentially be responsible for our results. However, based on the number of stars in each system quoted in the NASA Exoplanet Archive, we find equal multiplicity fractions (around 20%) in both of our fiducial samples. It remains possible that unresolved, unconfirmed or undetected binaries^{62,63} exist at different ratios within the two samples, but this would not change the empirical result that exoplanet architectures vary with environment. If it was found that the multiplicity fraction does differ between low and high stellar phase space densities, it would require an explanation of which physical mechanism drives this difference.

Bimodal exoplanet properties

A number of exoplanet properties have bimodal distributions (see Fig. 3), including their radii and masses. For such properties, it is instructive to consider differences in the distributions split into ‘low’ and ‘high’ components for overdensities and the field. The downside of this exercise is that it reduces both the dynamic range and the number of data points. Extended Data Fig. 4 shows the resulting distributions for the semi-major axis, planet mass and planet radius. The semi-major axis distribution across all stellar hosts is bimodal owing to an excess of hot Jupiters detected at small separations. However, contrary to the significant difference in semi-major axis distributions between low- and high-density environments shown in Fig. 3, the dependence on the phase space density is much smaller after dividing the sample into close-in and far-out planets, particularly for the latter population. The phase space density predominantly controls the relative sizes of these two populations. This suggests that the environment affects the semi-major axis distribution through a stochastic process that either leads to a pronounced reorganization of the planetary system (for example, by driving outer planets inward) or does not affect the semi-major axis

distribution at all. Both external photoevaporation and dynamical perturbations could act stochastically in principle. However, external photoevaporation can act on protoplanetary disks at a larger distance from an irradiation source than the encounter distance at which the dynamical interaction with a passing star can disrupt planetary systems. This suggests that the latter mechanism would be more stochastic and may be responsible for driving the migration of exoplanets to smaller semi-major axes that we have identified in this work.

Across all panels in Extended Data Fig. 4, we find the most significant dependence on phase space density at the low end of the mass distribution, suggesting that the environment most strongly affects the formation or evolution of terrestrial planets. In particular, low mass ($M_p < 50M_\oplus$) exoplanets are less massive when orbiting stars in overdensities. This is consistent with the results obtained for the unsplit sample; Fig. 3 shows that only 10% of planets around field stars have masses $M_p < 5M_\oplus$, whereas this is 24% for planets around hosts in overdensities. This is mirrored by a tentative, similar trend in planet radii. External photoevaporation due to irradiation of the circumstellar material by neighbouring massive stars could be responsible for these trends⁶⁴.

Eccentricity distribution

Figure 3 indicates that exoplanets found in overdensities typically have lower eccentricities than those in the field. Overdensities also exhibit an excess of hot Jupiters, of which the orbits are known to be susceptible to tidal circularization²⁸. We explore the possibility that the differences in eccentricity are driven by an enhanced incidence of hot Jupiters. We consider the median eccentricities of our fiducial sample, split by planet mass and semi-major axis (see Fig. 2). Next to being less common, the hot Jupiters orbiting field stars have marginally higher median eccentricities ($e = 0.05$) than those in overdensities ($e = 0.01$). If field hot Jupiters originate from stellar multiple interactions (see section ‘Hot Jupiters in the field’), this enhanced eccentricity could potentially indicate ongoing gravitational interactions within a stellar multiple system. However, this hypothesis remains speculative given that we identify only eight hot Jupiters around field stars.

Lower-mass exoplanets ($<50M_\oplus$) orbiting field stars also have a greater median eccentricity ($e = 0.17$) than those in overdensities ($e = 0.03$). This trend does not exist for ‘cool Jupiters’, with masses $M_p > 50M_\oplus$ and semi-major axes $a_p \lesssim 0.2$ AU (where dynamical encounters between stars would have the greatest influence). For these planets, the median eccentricity in overdensities is $e = 0.20$, whereas it is $e = 0.16$ in the field. This comparison shows that the difference in eccentricities between overdensities and the field is not restricted to hot Jupiters, but also extends to sub-Jupiter masses. This means that the reported eccentricity difference does not simply trace the difference in hot Jupiter incidence.

Hot Jupiters in the field

We report a correlation between the incidence of hot Jupiters and the phase space density around the host star, suggesting that the interaction with the environment is an important driver of hot Jupiter formation. Despite the clear association of hot Jupiters with phase space overdensities, we also find eight examples in the field. If the extreme orbits of hot Jupiters are due to perturbation by encounters with neighbouring stars, this process is stochastic and

could also result from stellar multiplicity^{56,65,66}. This could explain why hot Jupiters are found both in overdensities and the field. To investigate this, we here consider the eight examples we find in the field (see Fig. 2). These hot Jupiters are hosted by HAT-P-7 (ref.⁶⁷), HAT-P-12 (ref.⁶⁸), HD 68988 (ref.⁶⁹), HD 118203 (ref.⁷⁰), HIP 91258 (ref.⁷¹), TrES-3 (ref.⁷²), WASP-89 (ref.⁷³), and WASP-98 (ref.⁷⁴). Several of these systems are affected by large age or membership uncertainties. Specifically, HD 118203 is only marginally defined as a member of the low-density phase space component ($P_{\text{low}} = 0.87$). HIP 91258 and HAT-P-12 have uncertain ages consistent with being older than 4.5 Gyr and may therefore have been part of an overdensity in the past that has since dispersed.

In line with our hypothesis, we find that several field hot Jupiters may indeed be (or have been) part of a multiple system. HAT-P-7, HD 68988, HIP 91258 and WASP-98 may have a (sub-)stellar companion^{63,71,75,76}. For other hot Jupiters, the evidence is more circumstantial. An important difference between a multiple interaction and a stellar flyby is that the former can occur many times within a single system and therefore may redistribute planets differently than (possibly hyperbolic) encounters with stellar neighbours do. Generally speaking, we find that field hot Jupiters have a considerably broader distribution of masses and semi-major axes than those in overdensities (see Fig. 2). The hot Jupiter orbiting TrES-3 has an extremely short orbital period of just 1.31 days (semi-major axis of 0.023 AU), the one orbiting WASP-89 is particularly massive ($1,800M_{\oplus}$), whereas the one orbiting HAT-P-12 has a very low mass ($(67 \pm 4)M_{\oplus}$). Of course, multiple interactions may also affect planetary systems in overdensities. However, the difference between the distributions of hot Jupiters in Fig. 2 suggests the existence of more than one mode of dynamically redistributing planets, and in such a way that the balance between these modes differs between field stars and those in overdensities. In summary, it is possible that some fraction of these hot Jupiter host stars in the field originated in stellar multiples (that since may have dynamically decayed). Future studies of these systems may reveal further peculiarities.

Relation to other work

We now discuss the method adopted here in relation to previous methods and also discuss the physical implications of our results in relation to other works in the literature.

Analysis method—Our approach differs from other works in that the aim here is not to categorize substructure or infer the properties of individual co-moving groups, but to search for stellar phase space density perturbations. We make use of the Mahalanobis distance (equation (1)) to establish a density metric similar to those applied in cluster-finding algorithms^{34,35,77–80}. However, because these other approaches often aim to define and characterize specific groups, several of them require parameters (such as a density threshold) that need to be tuned to the problem at hand. Most importantly, all of these approaches implicitly assume that the ‘default’ state of a body is not to be a member of a group, and subsequently identify neighbours to build up structure according to a specific definition. Because we do not aim to associate stars to specific groups, we are free to employ a symmetric and probabilistic approach to delineating low- and high-density environments, which uses a continuous metric (the Mahalanobis distance) to quantify a membership

probability. Crucially, this approach does not require an implicit definition of moving groups, clusters or sub-clusters.

External effects on planetary systems— Figures 2 and 3 show clear differences in the architectures of exoplanetary systems between host stars in high-density and low-density environments. Across the sample of currently known exoplanets, stars in high-density environments are more likely to host short-period ($\lesssim 10$ days) and low-mass ($\lesssim 10M_{\oplus}$) planets, with an overabundance of hot Jupiters (about 30%). Conversely, exoplanets found around field stars follow a rough power-law trend between mass and semi-major axis ($M_p \propto a_p^{1.5}$), with a low fraction of hot Jupiters (about 10%). Although the best-fitting relation might just represent the upper envelope of a more complete sample of exoplanets, it is intriguing that planetary systems in overdensities intrinsically do not seem to follow such a relationship. Fitting a power law to the overdensity sample (again excluding the hot Jupiters) simply yields a (steeper) relationship between two unconnected regions of parameter space (low-mass, short-orbital-period planets and high-mass, long-orbital-period planets); within each of these subsets there is no clear correlation. The fact that the power-law fit provides a better description of the low-density sample than the high-density one favours a physical origin for the relation over selection biases. This would suggest that the planet population in overdensities evolved away from the trend observed for field stars. However, future exoplanet surveys (for example, with the Nancy Grace Roman Space Telescope⁸¹) are necessary to confirm or refute this hypothesis.

The physical mechanisms by which the birth environment of a host star might influence the architecture of planetary systems can be broadly divided into processes acting during planet formation and processes acting during the evolution of planetary systems. During planet formation in the first few million years of the system, the environment might affect the protoplanetary disk such that the properties of the resulting planetary system are changed too. After the dispersal of the protoplanetary disk, ambient stellar clustering may continue to affect the evolution of the planetary system by interactions with neighbouring stars.

The properties of protoplanetary disks (within which planets are born) are correlated with the properties of the star formation environment^{7,8,82–85}. In regions of high stellar density, protoplanetary disks can be truncated by encounters with nearby stars^{86–88}. However, extremely high stellar densities ($\gg 10^4$ stars pc⁻³) are required for dynamical encounters to have a substantial impact on the disk within a timescale similar to the disk lifetime. At high stellar densities, protoplanetary disks are also subjected to external irradiation by neighbouring massive stars. In local star-forming regions, models predict that the impact of the environment on overall disk evolution is dominated by external irradiation^{5,64,89–93}. In massive star-forming regions, the local far-ultraviolet flux experienced by protoplanetary disks induces mass loss due to external photoevaporation. Approximately half of the stars in the solar neighbourhood are born in environments experiencing sufficient flux to induce substantial mass loss and reduce the dispersal timescale, and this fraction increases further with the large-scale surface density of star formation within a galaxy^{9,94}. There are several observations that confirm the photoevaporative ‘depletion’ of disks, such as the ‘proplyds’ (bright ionization fronts surrounding protoplanetary disks) in the Orion Nebula cluster⁹⁵, the

fraction of surviving disks as a function of position in Pismis 24 (ref.⁸²) and Cygnus OB2^{83,96} and the radial gradient of disk dust masses in σ Orionis⁸.

It remains unclear how the effects of external photoevaporation influence the resultant planetary systems. External photoevaporation is most efficient at large disk radii, of $\gtrsim 10$ AU (ref.⁹²), such that exoplanets at smaller separations may not be influenced directly. However, premature disk dispersal^{82,83} may stop the growth of inner planets and their rapid migration by planet–disk interaction^{97–101}, which could also affect the orbital eccentricity^{102,103}. If exoplanets orbiting field stars are allowed to grow (because they are not externally irradiated in their birth environment), rapidly migrate, and ultimately accrete onto the host star, this could explain the dearth of short-period (low-mass) exoplanets. Remaining exoplanets around field stars would be more ‘loosely packed’ with respect to the compact exoplanetary systems in overdensities, and could therefore be subject to eccentricity excitation by more-massive outer companions¹⁰⁴. Finally, the environment may affect disk and planet properties through chemical enrichment. Recent work has highlighted that protoplanets may be heated by short-lived radionuclides, which probably originate in massive Wolf-Rayet stars and must be deposited quickly into nascent planetary systems owing to their short half-lives¹⁰⁵. This heating process sets the bulk water content and influences the formation of terrestrial planets¹⁰⁶.

After the dispersal of the protoplanetary disk and the emergence of a planetary system, the system can be subjected to gravitational perturbations by close encounters with neighbouring stars. From a theoretical perspective, it is well understood how such interactions may dynamically destabilize planetary systems, leading to scattering and the ejection of planets^{66,107,108}. In addition to the curtailing of planet growth and migration by external photoevaporation (discussed above), this dynamical process probably represents a second mode for close-in planet formation in overdensities that operates on longer (approximately 1 Gyr) timescales. Dynamical encounters can induce scattering and high-eccentricity planetary migration, after which the planetary orbits can be tidally circularized^{28,29,109}, and in some cases the planets can be photoevaporated by the central star¹¹⁰. Exoplanets that remain at wider separations may also have been dynamically perturbed, as has been suggested to explain the highly eccentric orbit of Pr 0211c in Praesepe (with orbital eccentricity $e \approx 0.7$, semi-major axis $a_p \approx 5.5$ AU, and mass $M_p \sin i \approx 7.8 M_J$, where M_J is the mass of Jupiter; refs.^{111,112}), as well as the peculiar orbits of trans-Neptunian objects orbiting the Sun¹¹³.

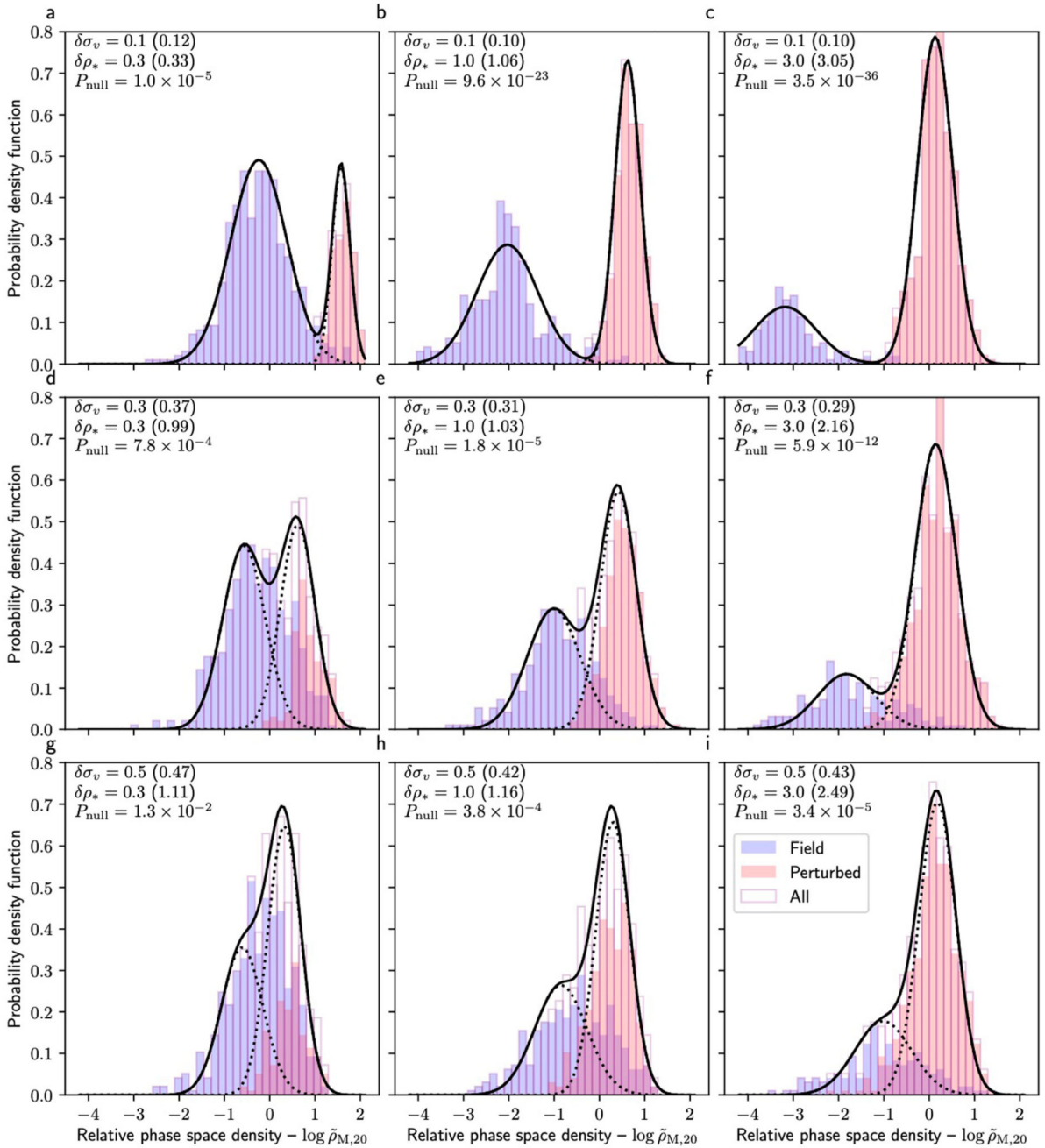
It remains unclear whether encounters in short-lived open clusters or associations would result in a greatly enhanced fraction of planets with semi-major axes of $a_p \lesssim 0.2$ AU. Numerical simulations of planet populations with initial semi-major axes of $a_p \approx 1$ AU show that stellar number densities of about 4×10^4 stars pc⁻³, typical of the cores of globular clusters⁵⁴, are required to scatter planets to the close-in orbits of hot Jupiters. At lower densities, encounters are too rare to drive inward migration to such small semi-major axes. At higher densities, they often eject the planet altogether¹¹⁴. However, for initial semi-major axes of $a_p \gg 1$ AU, stellar encounters can substantially alter exoplanet architectures even in clusters of moderate densities (about 100 stars pc⁻³)^{66,115}. Recent simulations have tried to address this problem by taking a population synthesis approach¹¹⁶. These authors

find that encounters can only marginally alter the exoplanet semi-major axes from the observed distribution. However, we have demonstrated in this work that many host stars that have previously been attributed to the field actually inhabit phase space overdensities. As a result, the observed distribution of semi-major axes does not necessarily represent the initial conditions for numerical experiments investigating the impact of dynamical perturbations, but may in part represent the target outcome. Future numerical simulations adopting an initial distribution of planet properties as in Fig. 2a could provide more representative insight into the true impact of dynamical perturbations on the architecture of planetary systems.

In general, any mechanism invoked to explain the differences in properties of exoplanets between high-density environments and the field needs to address (i) the overabundance of hot Jupiters in overdensities and the deviation of these systems from the power-law trend between mass and semi-major axis observed in the field (Fig. 2 and Fig. 3a), (ii) the decreased eccentricities of planets in overdensities (Fig. 3c), and (iii) the decrease of planet masses in overdensities (Fig. 3d). It is possible that a combination of external photoevaporation and short-lived radionuclide deposition (decreasing planet masses), followed by planet scattering due to stellar encounters (changing planet orbits) is responsible for the observed differences. The role and relative importance of these mechanisms represent important topics for future studies.

Origin of the Solar System—In addition to our analysis of exoplanetary systems, we have also characterized the kinematic environment of the Solar System, following the same methodology as for our exoplanet sample. We find that the Sun occupies an overdensity with moderate-to-high confidence ($P_{\text{high}} = 0.89$), which supports the hypothesis that the Sun was born in a high-mass star forming region^{117–119}. The Solar System planets are therefore included for comparison to exoplanetary systems occupying phase space overdensities in Fig. 2b. However, the incompleteness of exoplanet samples for Solar-System-like planets means that it is at present not possible to make a quantitative comparison.

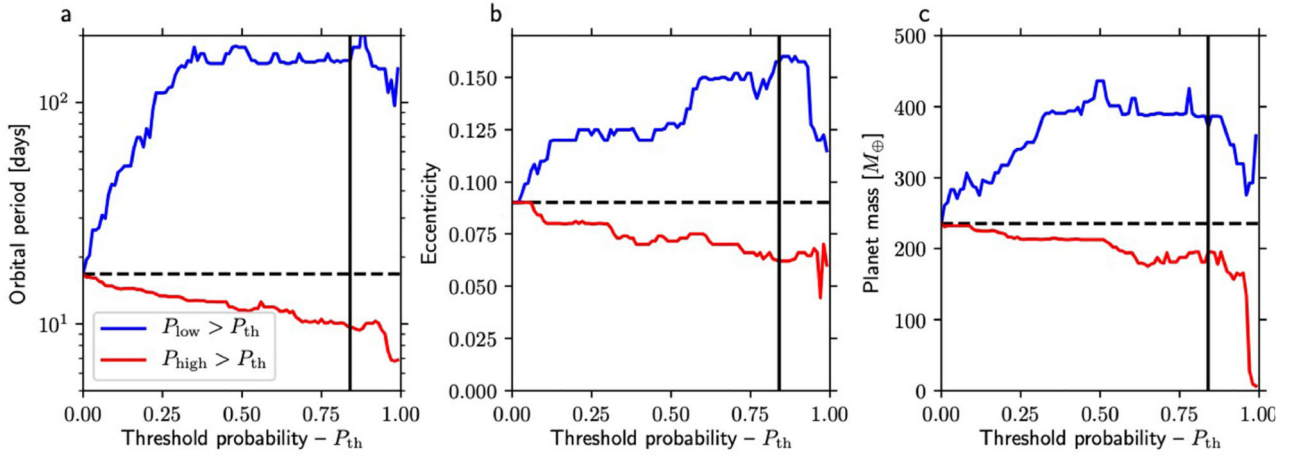
Extended Data



Extended Data Fig. 1. Probability density functions of the relative phase space density for synthetic stellar populations.

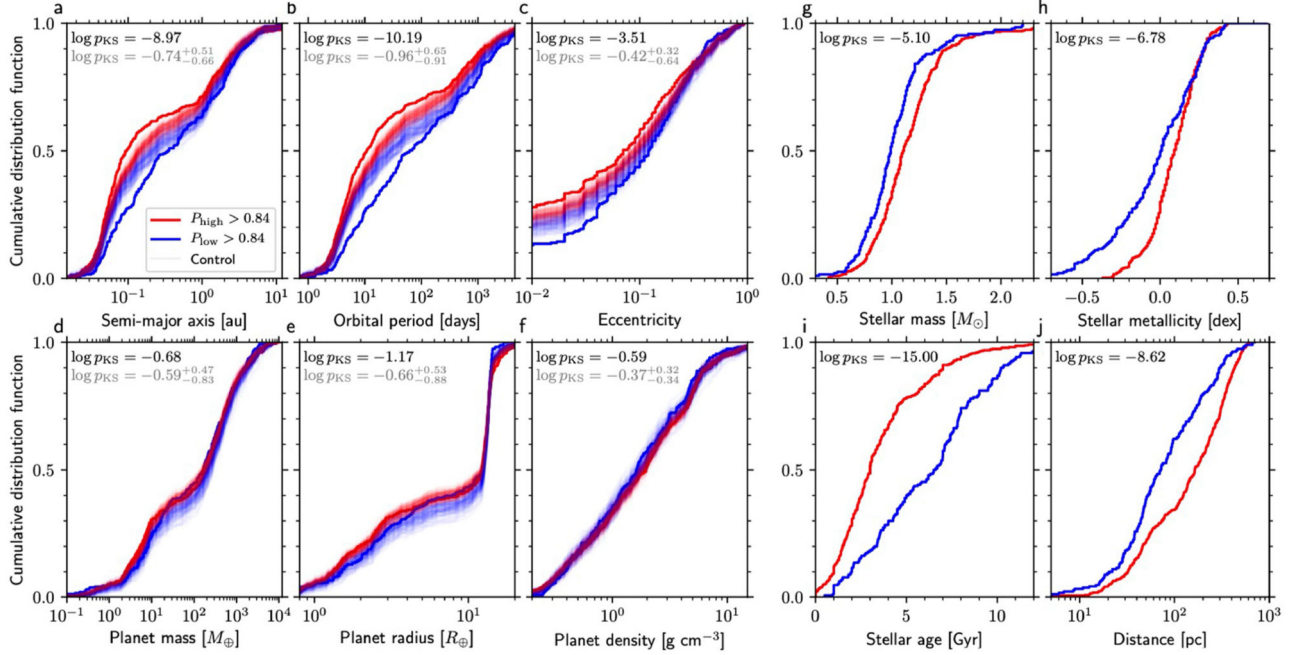
Blue histograms represent the distribution of $\tilde{\rho}_{M,20}$ for a background ('field') population, while red histograms represent a population of stars with a spatial density perturbed by a multiplicative factor Sp . (increasing from left to right, **a, d, g-c, f, i**) and with a velocity dispersion perturbed by a multiplicative factor $S\sigma_v$ (increasing from top to bottom; **a, b, c-g, h, i**). Outlined purple histograms show the sum of the perturbed and background

populations. The solid black line represents a double-lognormal fit to this combined phase space density distribution, with both lognormal components marked by dotted lines. The multiplicative factors by which the density and velocity dispersion are perturbed (numbers in brackets are the values of $\delta\rho$ and $\delta\sigma_v$ inferred from the phase space density decomposition), as well as the probability that the distribution can be described by a single lognormal (P_{null}) are shown.



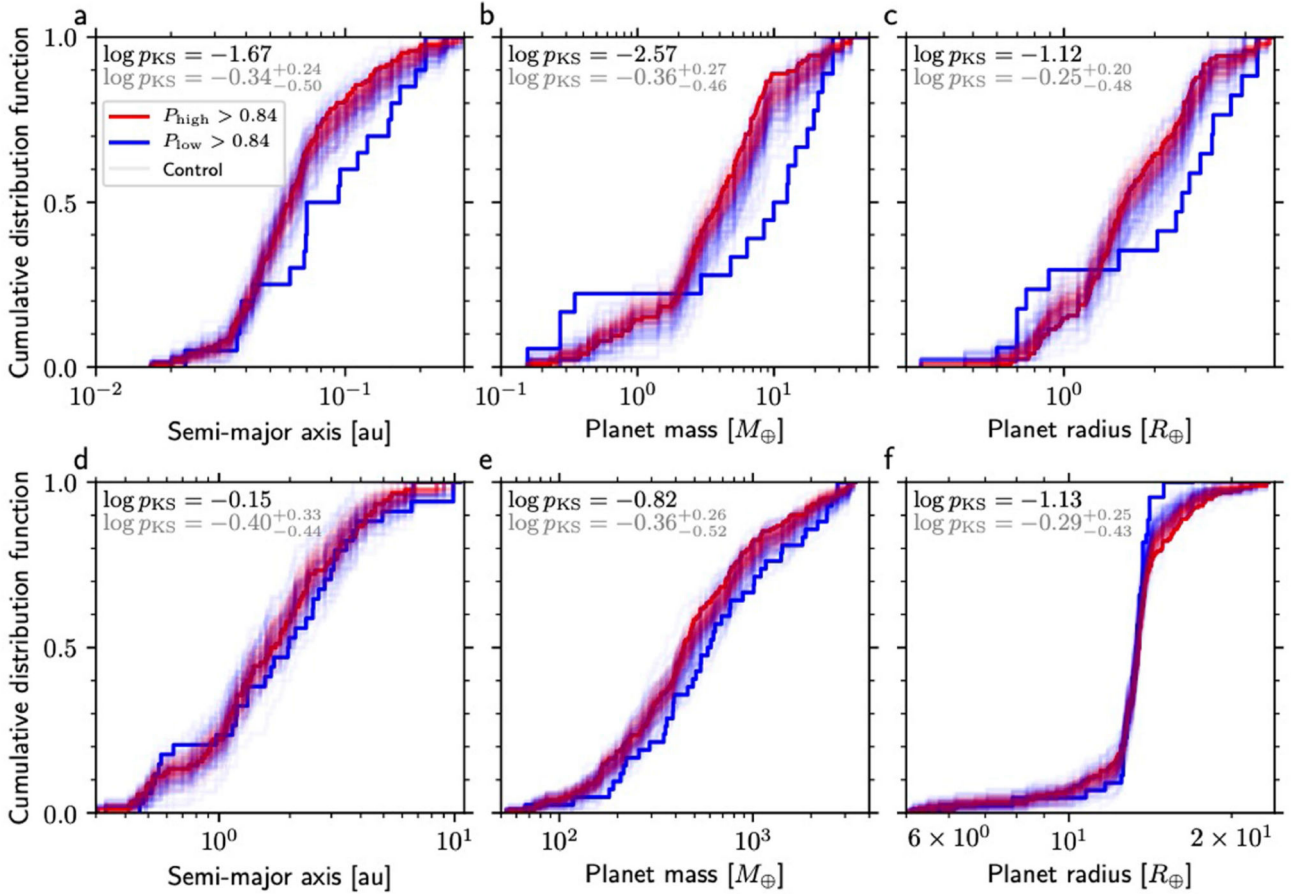
Extended Data Fig. 2. Effect of the choice of threshold probability on the median exoplanet properties in environments with low and high phase space density.

The panels show the median orbital period (a), orbital eccentricity (b), and planet mass (c), for the same exoplanet host star sample as in Fig. 3. Exoplanets orbiting field stars ($P_{\text{low}} > P_{\text{th}}$) are shown in blue, and exoplanets orbiting stars within overdensities ($P_{\text{high}} > P_{\text{th}}$) are shown in red. The median of the full sample is shown as a dashed black line, and the chosen $P_{\text{th}} = 0.84$ (adopted for our main results) is shown as a vertical black line.



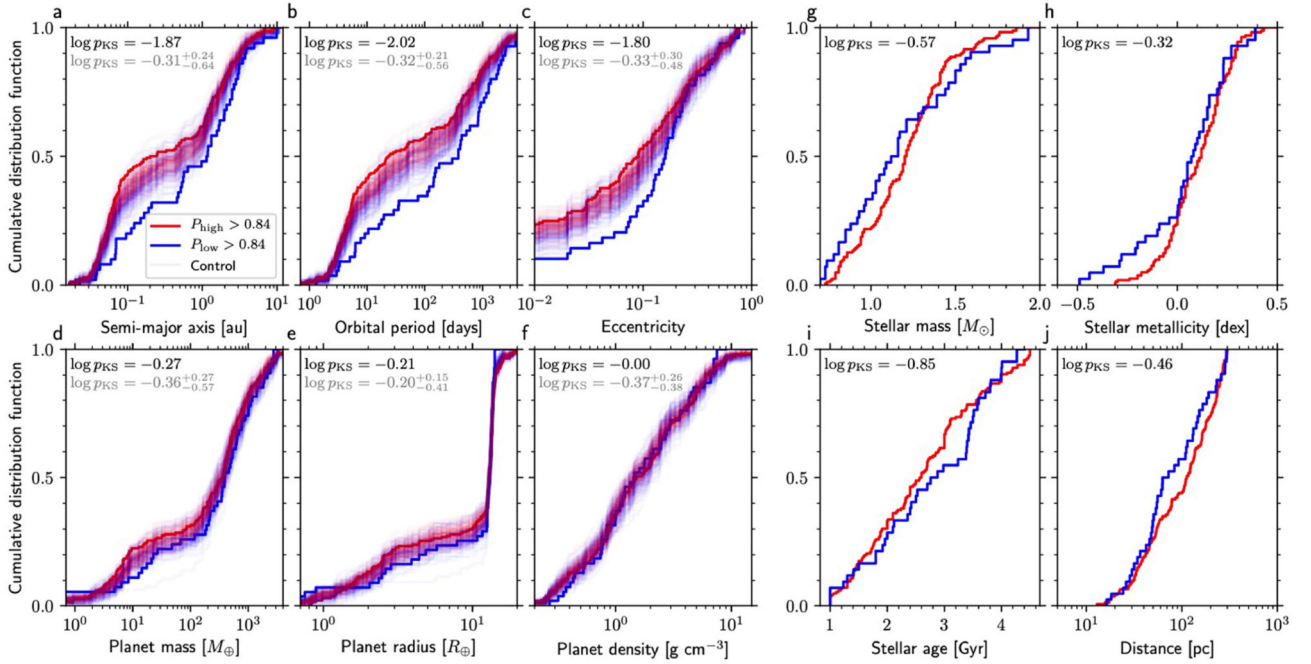
Extended Data Fig. 3. Normalized cumulative distribution functions of planet and host star properties.

The samples are divided into low (blue) and high (red) host star phase space densities, without applying any cuts in host star age or mass (unlike in Fig. 3). The panels are the same as in Fig. 3 (a-f for exoplanet properties, g-j for stellar host properties). The faint lines represent 100 Monte Carlo control experiments, constructed by drawing a star at random from within 40 pc of each exoplanet host and using the phase space density of that star instead. The logarithms of p values obtained from a two-tailed Kolmogorov–Smirnov test for the exoplanet hosts (black) and for the median of all control experiments (grey; including 16th–84th percentile uncertainties) are shown.



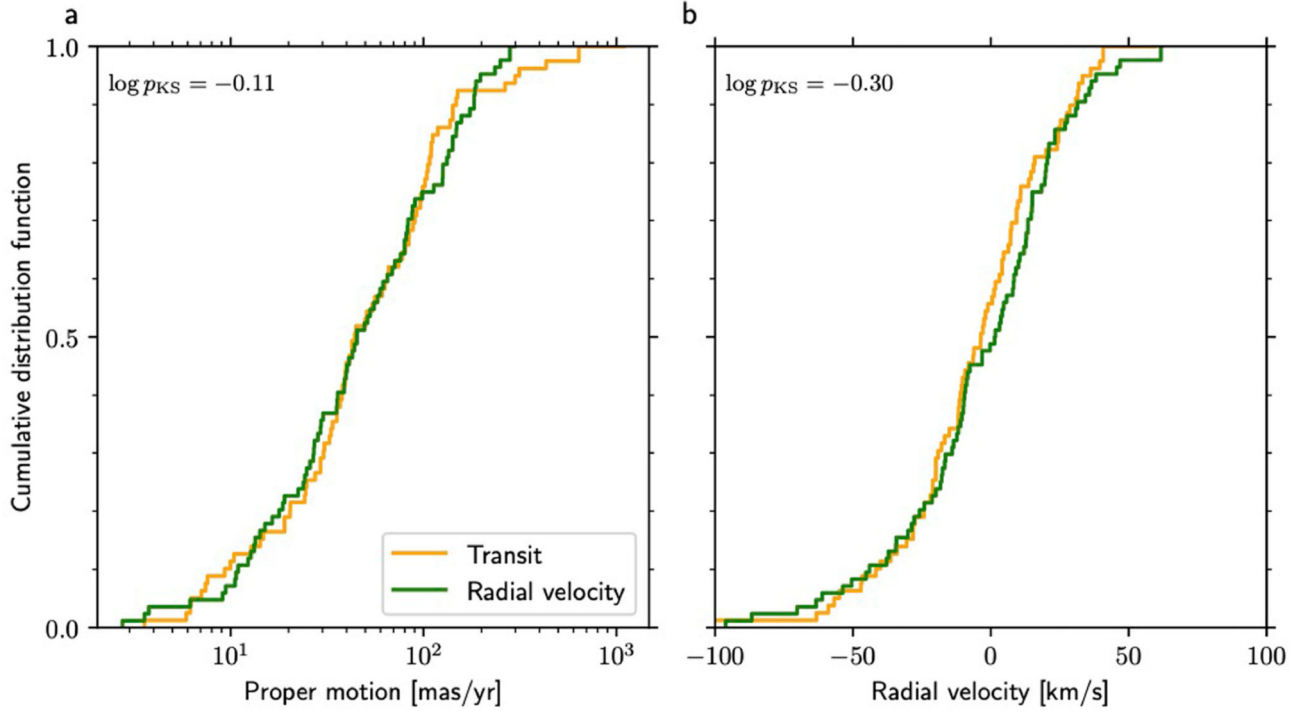
Extended Data Fig. 4. Normalized cumulative distribution functions of exoplanet properties that exhibit bimodal distributions.

The samples are divided into low (blue) and high (red) host star phase space densities. The sample is split across the top and bottom rows by semi-major axes (**a**, <0.3 AU; **d**, >0.3 AU), planet masses (**b**, $<50 M_{\oplus}$; **e**, $>50 M_{\oplus}$), and radii (**c**, $<5 R_{\oplus}$; **f**, $>5 R_{\oplus}$). The distributions are shown for the same exoplanet host sample as in Fig. 3. The faint lines represent 100 Monte Carlo control experiments, constructed by drawing a star at random from within 40 pc of each exoplanet host and using the phase space density of that star instead. The logarithms of p values obtained from a two-tailed Kolmogorov–Smirnov test for the exoplanet hosts (black) and for the median of all control experiments (grey; including 16th–84th percentile uncertainties) are shown.



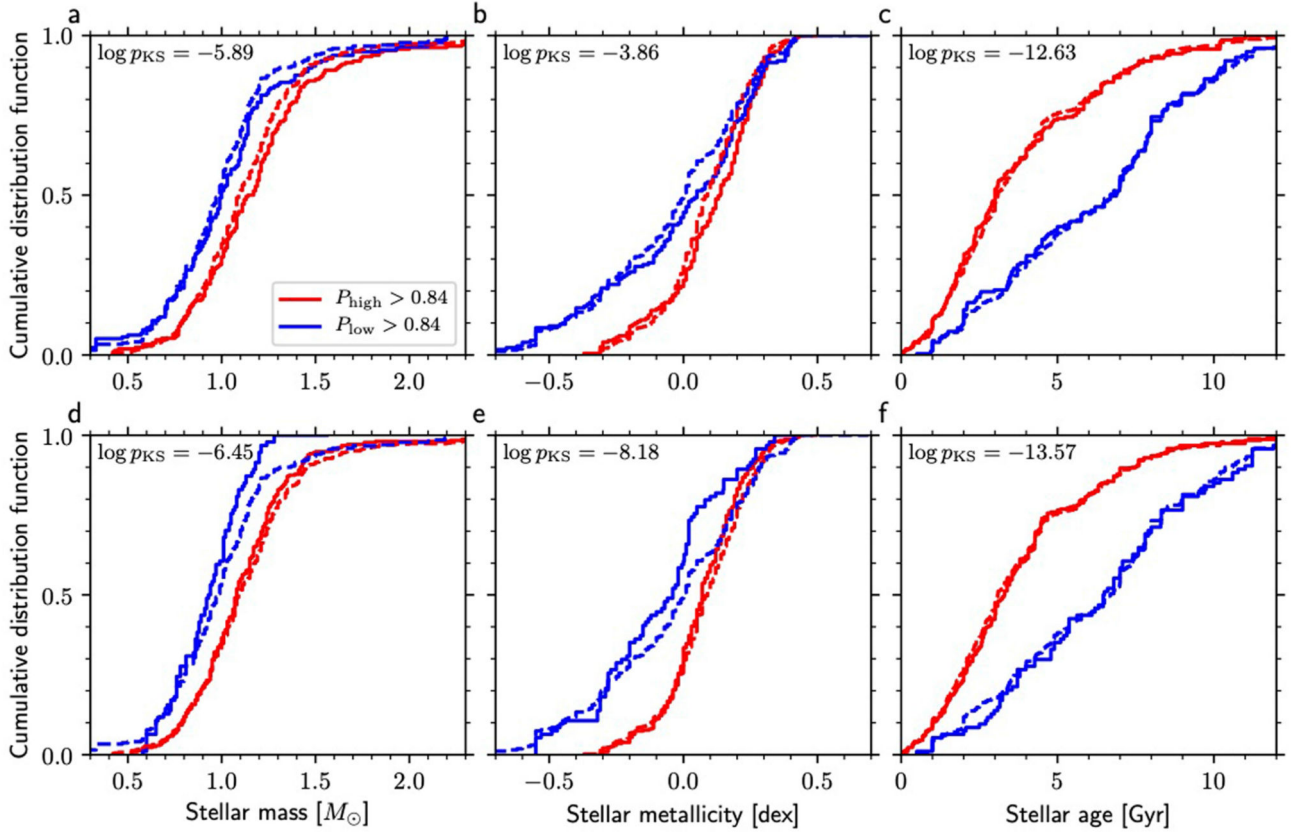
Extended Data Fig. 5. Normalized cumulative distribution functions of planet and host star properties in our fiducial sample, limiting the sample to systems within 300 pc of the Sun (unlike in Fig. 3).

The samples are divided into low (blue) and high (red) host star phase space densities. The panels are the same as in Fig. 3 (a–f for exoplanet properties, g–j for stellar host properties). The faint lines represent 100 Monte Carlo control experiments, constructed by drawing a star at random from within 40 pc of each exoplanet host and using the phase space density of that star instead. The logarithms of p values obtained from a two-tailed Kolmogorov–Smirnov test for the exoplanet hosts (black) and for the median of all control experiments (grey; including 16th–84th percentile uncertainties) are shown.



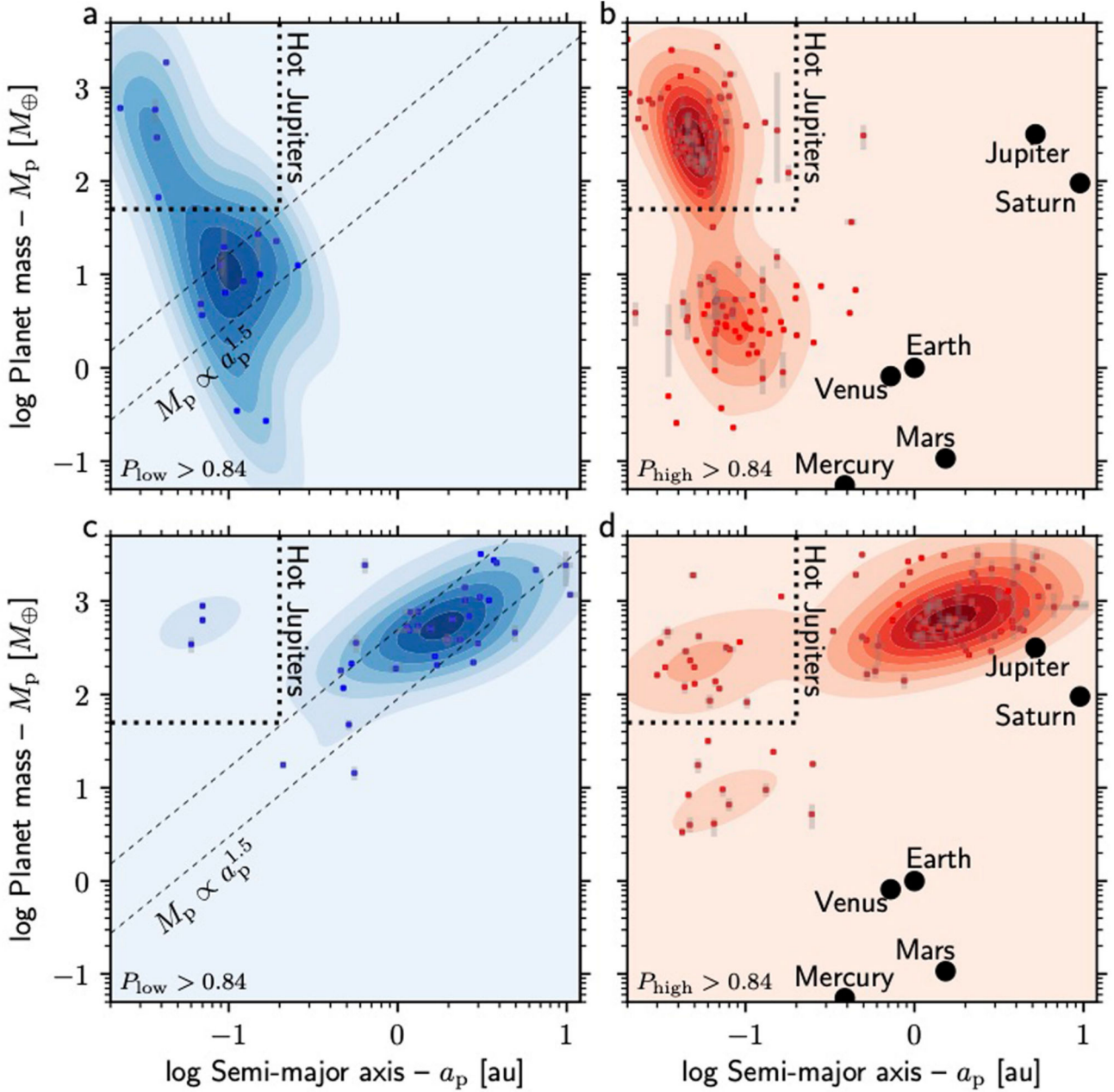
Extended Data Fig. 6. Normalized cumulative distribution functions of the kinematic properties of the host stars.

Panel **a** shows the distribution of absolute proper motions, whereas panel **b** shows the same for radial velocities. The distributions are shown for all exoplanet host stars that have age and mass estimates. The sample is split by exoplanet discovery method (radial velocity in green, transit in orange) and both subsamples have the same distance distribution by construction (see Methods). The logarithms of p values obtained from a two-tailed Kolmogorov–Smirnov test between the two survey types are shown.

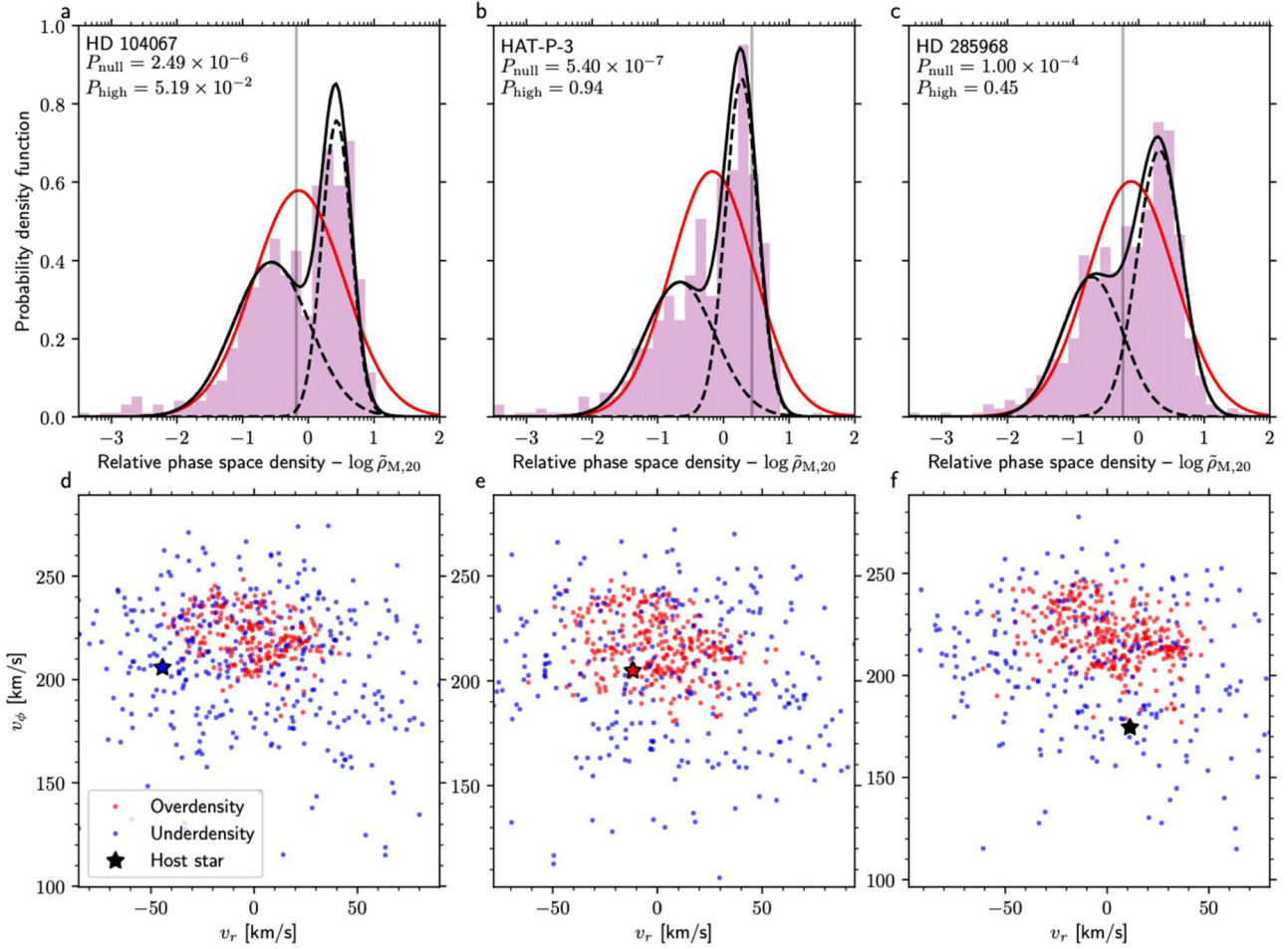


Extended Data Fig. 7. Normalized cumulative distribution functions of host star properties in the complete sample of Extended Data Fig. 3.

The sample is divided into exoplanets discovered by radial velocity (**a-c**) and transit (**d-f**) surveys. Red lines indicate exoplanet host stars that occupy a phase space overdensity, whereas blue lines represent host stars in the field. For reference, the distributions of the entire host star sample (including all detection methods) from Extended Data Fig. 3 are shown as dashed lines. The logarithms of p values obtained from a two-tailed Kolmogorov–Smirnov test are shown.

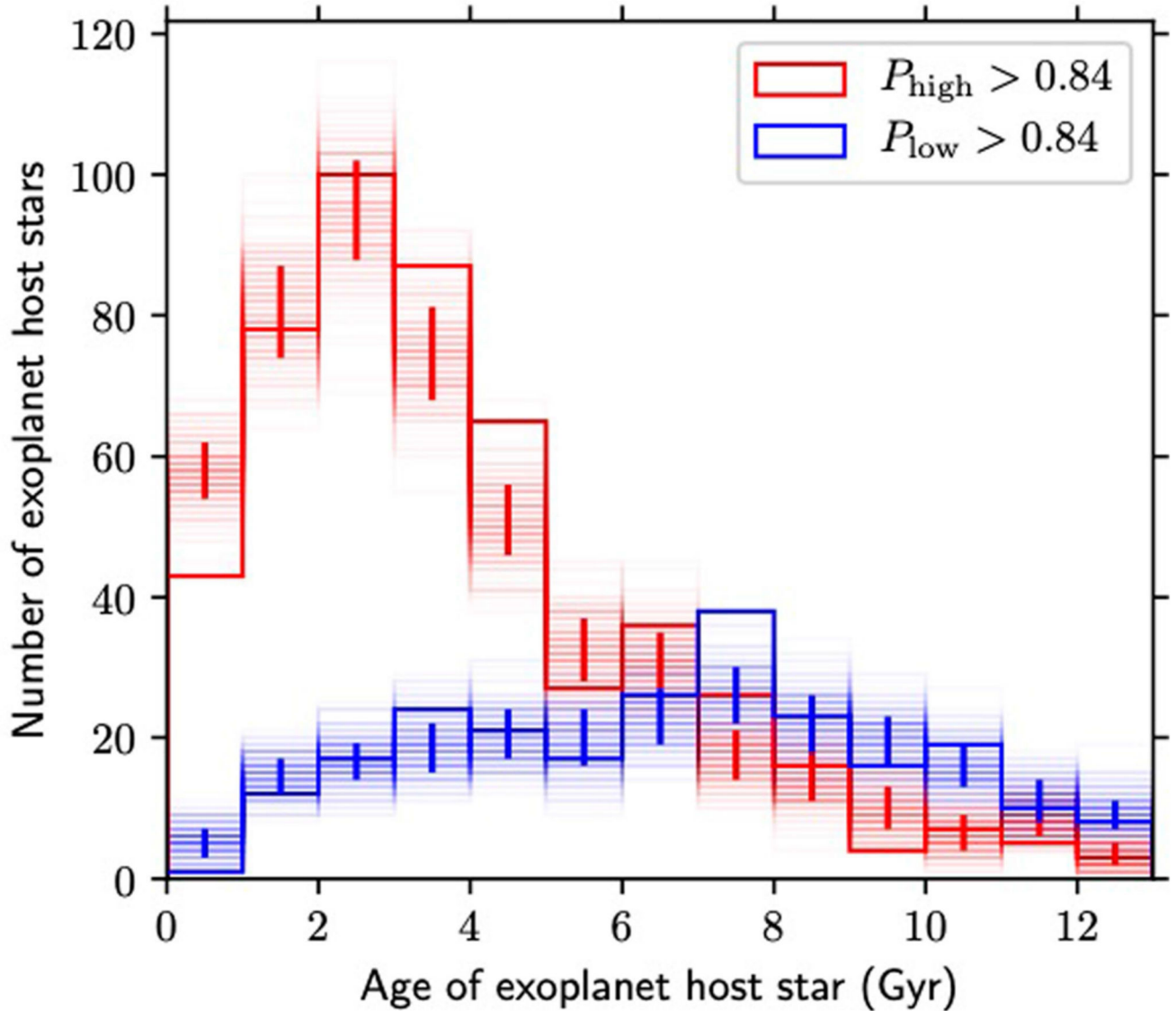


Extended Data Fig. 8. Distributions of exoplanet semi-major axes and masses split by ambient stellar phase space density for different planet discovery methods. Columns indicate low (**a, c**; $P_{\text{low}} > 0.84$) and high (**b, d**; $P_{\text{high}} > 0.84$) phase space densities (as in Fig. 2), split into rows of exoplanets discovered by transit (**a, b**) and radial velocity (**c, d**) surveys. Data points with grey error bars (indicating 1σ uncertainties) show individual planets and contours show a two-dimensional Gaussian kernel density estimate. The dashed black lines in **a** and **c** follow $M_p \propto a_p^{1.5}$ and illustrate the 1σ scatter around an orthogonal distance regression to all planets orbiting field stars that are not hot Jupiters (see Fig. 2a). For reference, **b** and **d** includes the Solar System ($P_{\text{high}} = 0.89$) planets within $a_p < 10$ AU.



Extended Data Fig. 9. Phase space distributions of stars near the three exoplanet host stars HD 104067, HAT-P-3 and HD 285968.

Panels **a-c** show the phase space density distributions (purple histograms), together with the best-fitting double-lognormal function (black solid line) and the individual lognormal components (black dashed lines) obtained by Gaussian mixture modelling. Keys list the probability that the density distribution is described by a single lognormal (red line) as P_{null} , and the probability that each exoplanet host is associated with a phase space overdensity as P_{high} . Panels **d-f** show the azimuthal (V_ϕ) and radial (V_r) components of the stellar velocities in galactocentric coordinates. Stars in overdensities are shown in red, whereas field stars are shown in blue. To divide the stars into a low- and high-density population, we apply a Monte Carlo procedure that randomly assigns stars based on their individual probabilities of belonging to either of the two components (equation (5)). The host star velocity is shown as a star symbol. These three host stars illustrate cases of a highly significant ($P_{\text{high}} = 0.05$) low phase space density (HD 104067), a highly significant ($P_{\text{high}} = 0.94$) phase space overdensity (HAT-P-3) and an ambiguous ($P_{\text{high}} = 0.45$) phase space density (HD 285968).



Extended Data Fig. 10. Age distributions of exoplanet host stars with masses $0.7M_{\odot}$ - $2M_{\odot}$. The red histogram shows stars in overdensities ($P_{\text{high}} > 0.84$) and the blue histogram shows field stars ($P_{\text{low}} > 0.84$). The faint lines represent the results of performing 200 Monte Carlo realizations of the ages, drawn from normal distributions defined by the measured ages and their uncertainties. The error bars show the 16th-84th percentile range of the resulting age distributions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

A.J.W. thanks R. Alexander for discussions. A.J.W. acknowledges funding from the Alexander von Humboldt Stiftung in the form of a Postdoctoral Research Fellowship and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 681601). J.M.D.K. and M.C. acknowledge funding from the German Research Foundation (DFG) in the form of an Emmy Noether Research Group (grant no. KR4801/1-1) and the DFG Sachbeihilfe (grant no. KR4801/2-1). J.M.D.K. acknowledges funding from the ERC under the European Union's Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (grant agreement no. 714907). This research made use of data from the European Space Agency mission Gaia (<http://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <http://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. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

Data availability

The Gaia data used in this work are publicly available through the Gaia archive (<https://gea.esac.esa.int/archive/>). The exoplanetary catalogue used in this work is publicly available through the NASA Exoplanet Archive (<https://exoplanetarchive.ipac.caltech.edu/>). Results of the calculations performed as part of this work are either available in the Supplementary Information, or from the authors upon request. A table containing the planet properties, host star properties, and the phase space decomposition is publicly available at <https://github.com/ajw278/astrophasesplit> with file name planetdata (2).csv.

Code availability

The code used for the phase space decomposition is publicly available at <https://github.com/ajw278/astrophasesplit>.

References

1. Armitage PJ. Dynamics of protoplanetary disks. *Annu Rev Astron Astrophys.* 2011; 49:195–236.
2. Williams JP, Cieza LA. Protoplanetary disks and their evolution. *Annu Rev Astron Astrophys.* 2011; 49:67–117.
3. Winn JN, Fabrycky DC. The occurrence and architecture of exoplanetary systems. *Annu Rev Astron Astrophys.* 2015; 53:409–447.
4. Mordasini C, Alibert Y, Benz W, Klahr H, Henning T. Extrasolar planet population synthesis. IV. Correlations with disk metallicity, mass, and lifetime. *Astron Astrophys.* 2012; 541:A97.
5. Adams FC, Hollenbach D, Laughlin G, Gorti U. Photoevaporation of circumstellar disks due to external far-ultraviolet radiation in stellar aggregates. *Astrophys J.* 2004; 611:360–379.
6. Cai MX, Portegies Zwart S, van Elteren A. The signatures of the parental cluster on field planetary systems. *Mon Not R Astron Soc.* 2018; 474:5114–5121.
7. de Juan Ovelar M, et al. Can habitable planets form in clustered environments? *Astron Astrophys.* 2012; 546:L1.
8. Ansdell M, et al. An ALMA survey of protoplanetary disks in the σ Orionis cluster. *Astron J.* 2017; 153:240.
9. Winter AJ, Kruijssen JMD, Chevance M, Keller BW, Longmore SN. Prevalent externally driven protoplanetary disc dispersal as a function of the galactic environment. *Mon Not R Astron Soc.* 2020; 491:903–922.
10. Krumholz MR, McKee CF, Bland-Hawthorn J. Star clusters across cosmic time. *Annu Rev Astron Astrophys.* 2019; 57:227–303.
11. Gaia Collaboration. et al. The *Gaia* mission. *Astron Astrophys.* 2016; 595:A1.

12. Gaia Collaboration. et al. *Gaia* Data Release 2. Summary of the contents and survey properties. *Astron Astrophys.* 2018; 616:A1.
13. Batygin K. A primordial origin for misalignments between stellar spin axes and planetary orbits. *Nature.* 2012; 491:418–420. [PubMed: 23151584]
14. Baruteau, C, , et al. Planet-disk interactions and early evolution of planetary systems Protostars and Planets VI. Beuther, H, Klessen, RS, Dullemond, CP, Henning, T, editors. 2014. 667
15. TriAUD AHMJ, et al. Spin-orbit angle measurements for six southern transiting planets. New insights into the dynamical origins of hot Jupiters. *Astron Astrophys.* 2010; 524:A25.
16. Albrecht S, et al. Obliquities of hot Jupiter host stars: evidence for tidal interactions and primordial misalignments. *Astrophys J.* 2012; 757:18.
17. Kruijssen JMD. On the fraction of star formation occurring in bound stellar clusters. *Mon Not R Astron Soc.* 2012; 426:3008–3040.
18. Hopkins PF. Why do stars form in clusters? An analytic model for stellar correlation functions. *Mon Not R Astron Soc.* 2013; 428:1950–1957.
19. Quillen AC, et al. Spiral arm crossings inferred from ridges in Gaia stellar velocity distributions. *Mon Not R Astron Soc.* 2018; 480:3132–3139.
20. Fragkoudi F, et al. On the ridges, undulations, and streams in Gaia DR2: linking the topography of phase space to the orbital structure of an N-body bar. *Mon Not R Astron Soc.* 2019; 488:3324–3339.
21. Composite Planet Data Table. NASA Exoplanet Archive; 2019.
22. Johnson JA, et al. Retired A stars and their companions: exoplanets orbiting three intermediate-mass subgiants. *Astrophys J.* 2007; 665:785–793.
23. Hebb L, et al. WASP-12b: the hottest transiting extrasolar planet yet discovered. *Astrophys J.* 2009; 693:1920–1928.
24. Johnson JA, Aller KM, Howard AW, Crepp JR. Giant planet occurrence in the stellar mass-metallicity plane. *Publ Astron Soc Pacif.* 2010; 122:905.
25. Wyatt MC, Jackson AP. Insights into planet formation from debris disks. II. Giant impacts in extrasolar planetary systems. *Space Sci Rev.* 2016; 205:231–265.
26. Madhusudhan N, Amin MA, Kennedy GM. Toward chemical constraints on hot Jupiter migration. *Astrophys J Lett.* 2014; 794:L12.
27. Dawson RI, Johnson JA. Origins of Hot Jupiters. *Annu Rev Astron Astrophys.* 2018; 56:175–221.
28. Jackson B, Greenberg R, Barnes R. Tidal evolution of close-in extrasolar planets. *Astrophys J.* 2008; 678:1396–1406.
29. Rasio FA, Ford EB. Dynamical instabilities and the formation of extrasolar planetary systems. *Science.* 1996; 274:954–956. [PubMed: 8875930]
30. Morbidelli A, Lunine JI, O'Brien DP, Raymond SN, Walsh KJ. Building terrestrial planets. *Annu Rev Earth Planet Sci.* 2012; 40:251–275.
31. Minchev I, Boily C, Siebert A, Bienayme O. Low-velocity streams in the solar neighbourhood caused by the Galactic bar. *Mon Not R Astron Soc.* 2010; 407:2122–2130.
32. Kruijssen JMD, et al. Fast and inefficient star formation due to short-lived molecular clouds and rapid feedback. *Nature.* 2019; 569:519–522. [PubMed: 31118525]
33. Chevance M, et al. The lifecycle of molecular clouds in nearby star-forming disc galaxies. *Mon Not R Astron Soc.* 2020; 493:2872–2909.
34. Eisenstein DJ, Hut P. HOP: a new group-finding algorithm for N-body simulations. *Astrophys J.* 1998; 498:137–142.
35. Maciejewski M, Colombi S, Springel V, Alard C, Bouchet FR. Phase-space structures. II. Hierarchical structure finder. *Mon Not R Astron Soc.* 2009; 396:1329–1348.
36. Pedregosa F, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011; 12:2825–2830.
37. Seabroke GM, Gilmore G. Revisiting the relations: Galactic thin disc age-velocity dispersion relation. *Mon Not R Astron Soc.* 2007; 380:1348–1368.
38. Kamdar H, et al. Stars that move together were born together. *Astrophys J Lett.* 2019; 884:L42.

39. Pfeffer J, et al. Young star cluster populations in the E-MOSAICS simulations. *Mon Not R Astron Soc.* 2019; 490:1714–1733.
40. Adamo A, et al. Star clusters near and far; tracing star formation across cosmic time. *Space Sci Rev.* 2020; 216:69.
41. Antoja T, Figueras F, Fernández D, Torra J. Origin and evolution of moving groups. I. Characterization in the observational kinematic-age-metallicity space. *Astron Astrophys.* 2008; 490:135–150.
42. Fürnkranz V, Meingast S, Alves J. Extended stellar systems in the solar neighborhood. III. Like ships in the night: the Coma Berenices neighbor moving group. *Astron Astrophys.* 2019; 624:L11.
43. Meingast S, Alves J, Fürnkranz V. Extended stellar systems in the solar neighborhood. II. Discovery of a nearby 120° stellar stream in Gaia DR2. *Astron Astrophys.* 2019; 622:L13.
44. Price-Jones N, et al. Strong chemical tagging with APOGEE: 21 candidate star clusters that have dissolved across the Milky Way disc. *Mon Not R Astron Soc.* 2020; 496:5101–5115.
45. Famaey B, Siebert A, Jorissen A. On the age heterogeneity of the Pleiades, Hyades, and Sirius moving groups. *Astron Astrophys.* 2008; 483:453–459.
46. Lépine JRD, Michtchenko TA, Barros DA, Vieira RSS. The dynamical origin of the local arm and the Sun's trapped orbit. *Astrophys J.* 2017; 843:48.
47. Kamdar H, et al. A dynamical model for clustered star formation in the Galactic disk. *Astrophys J.* 2019; 884:173.
48. Oh S, Price-Whelan AM, Hogg DW, Morton TD, Spergel DN. Comoving stars in Gaia DR1: an abundance of very wide separation comoving pairs. *Astron J.* 2017; 153:257.
49. Haisch J, Karl E, Lada EA, Lada CJ. Disk frequencies and lifetimes in young clusters. *Astrophys J Lett.* 2001; 553:L153–L156.
50. Holland WS, et al. Submillimetre images of dusty debris around nearby stars. *Nature.* 1998; 392:788–791.
51. Halliday AN. A young Moon-forming giant impact at 70–110 million years accompanied by late-stage mixing, core formation and degassing of the Earth. *Phil Trans R Soc Lond A.* 2008; 366:4163–4181.
52. Kennedy GM, Wyatt MC. The bright end of the exo-Zodi luminosity function: disc evolution and implications for exo-Earth detectability. *Mon Not R Astron Soc.* 2013; 433:2334–2356.
53. Snaith O, et al. Reconstructing the star formation history of the Milky Way disc(s) from chemical abundances. *Astron Astrophys.* 2015; 578:A87.
54. Portegies Zwart SF, McMillan SLW, Gieles M. Young massive star clusters. *Annu Rev Astron Astrophys.* 2010; 48:431–493.
55. Fischer DA, Valenti J. The planet-metallicity correlation. *Astrophys J.* 2005; 622:1102–1117.
56. Winn JN, Fabrycky D, Albrecht S, Johnson JA. Hot stars with hot Jupiters have high obliquities. *Astrophys J Lett.* 2010; 718:L145–L149.
57. Reffert S, Bergmann C, Quirrenbach A, Trifonov T, Künstler A. Precise radial velocities of giant stars. VII. Occurrence rate of giant extrasolar planets as a function of mass and metallicity. *Astron Astrophys.* 2015; 574:A116.
58. Kaib NA, Raymond SN, Duncan M. Planetary system disruption by Galactic perturbations to wide binary stars. *Nature.* 2013; 493:381–384. [PubMed: 23292514]
59. Veras D, Georgakarakos N, Dobbs-Dixon I, Gänsicke BT. Binary star influence on post-main-sequence multi-planet stability. *Mon Not R Astron Soc.* 2017; 465:2053–2059.
60. Kervella P, Arenou F, Mignard F, Thévenin F. Stellar and substellar companions of nearby stars from Gaia DR2. Binarity from proper motion anomaly. *Astron Astrophys.* 2019; 623:A72.
61. Fontanive C, et al. A high binary fraction for the most massive close-in giant planets and brown dwarf desert members. *Mon Not R Astron Soc.* 2019; 485:4967–4996.
62. Evans DF. Evidence for unresolved exoplanet-hosting binaries in Gaia DR2. *Res Not Am Astron Soc.* 2018; 2:20.
63. Belokurov V, et al. Unresolved stellar companions with Gaia DR2 astrometry. *Mon Not R Astron Soc.* 2020; 496:1922–1940.

64. Johnstone D, Hollenbach D, Bally J. Photoevaporation of disks and clumps by nearby massive stars: application to disk destruction in the Orion nebula. *Astrophys J.* 1998; 499:758–776.
65. Knutson HA, et al. Friends of hot Jupiters. I. A radial velocity search for massive, long-period companions to close-in gas giant planets. *Astrophys J.* 2014; 785:126.
66. Davies, MB, , et al. The long-term dynamical evolution of planetary systems. *Protostars and Planets VI.* University of Arizona Press; 2014.
67. Pál A, et al. HAT-P-7b: an extremely hot massive planet transiting a bright star in the Kepler field. *Astrophys J.* 2008; 680:1450–1456.
68. Hartman JD, et al. HAT-P-12b: a low-density sub-Saturn mass planet transiting a metal-poor K dwarf. *Astrophys J.* 2009; 706:785–796.
69. Vogt SS, et al. Ten low-mass companions from the Keck Precision Velocity Survey. *Astrophys J.* 2002; 568:352–362.
70. Da Silva R, et al. Elodie metallicity-biased search for transiting Hot Jupiters. I. Two hot Jupiters orbiting the slightly evolved stars HD 118203 and HD 149143. *Astron Astrophys.* 2006; 446:717–722.
71. Moutou C, et al. The SOPHIE search for northern extrasolar planets. VI. Three new hot Jupiters in multi-planet extrasolar systems. *Astron Astrophys.* 2014; 563:A22.
72. O'Donovan FT, et al. TrES-3: a nearby, massive, transiting hot Jupiter in a 31 hour orbit. *Astrophys J Lett.* 2007; 663:L37–L40.
73. Hellier C, et al. Three WASP-South transiting exoplanets: WASP-74b, WASP-83b, and WASP-89b. *Astron J.* 2015; 150:18.
74. Hellier C, et al. Transiting hot Jupiters from WASP-South, Euler and TRAPPIST: WASP-95b to WASP-101b. *Mon Not R Astron Soc.* 2014; 440:1982–1992.
75. Winn JN, et al. HAT-P-7: a retrograde or polar orbit, and a third body. *Astrophys J Lett.* 2009; 703:L99–L103.
76. Mugrauer M. Search for stellar companions of exoplanet host stars by exploring the second ESA-Gaia data release. *Mon Not R Astron Soc.* 2019; 490:5088–5102.
77. Casertano S, Hut P. Core radius and density measurements in N-body experiments. Connections with theoretical and observational definitions. *Astrophys J.* 1985; 298:80–94.
78. Ester, M; Kriegel, H-P; Sander, J; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc 2nd Int Conf on Knowledge Discovery and Data Mining (KDD'96)*; AAAI Press; 1996. 226–231.
79. Sharma S, Johnston KV. A group finding algorithm for multidimensional data sets. *Astrophys J.* 2009; 703:1061–1077.
80. Myeong GC, Evans NW, Belokurov V, Amorisco NC, Koposov SE. Halo substructure in the SDSS-Gaia catalogue: streams and clumps. *Mon Not R Astron Soc.* 2018; 475:1537–1548.
81. Spergel, D; , et al. Wide-field infrared survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 report. 2015. Preprint at <https://arxiv.org/abs/1503.03757>
82. Fang M, et al. Star formation and disk properties in Pismis 24. *Astron Astrophys.* 2012; 539:A119.
83. Guarcello, MG; , et al. Photoevaporation and close encounters: how the environment around Cygnus OB2 affects the evolution of protoplanetary disks. 2016. Preprint at <https://arxiv.org/abs/1605.01773>
84. Winter AJ, Ansdell M, Haworth TJ, Kruijssen JMD. Testing viscous disc theory using the balance between stellar accretion and external photoevaporation of protoplanetary discs. *Mon Not R Astron Soc.* 2020; 497:L40–L45.
85. van Terwisga SE, Hacar A, van Dishoeck EF. Disk masses in the Orion Molecular Cloud-2: distinguishing time and environment. *Astron Astrophys.* 2019; 628:A85.
86. Clarke CJ, Pringle JE. Accretion disc response to a stellar fly-by. *Mon Not R Astron Soc.* 1993; 261:190–202.
87. Ostriker EC. Capture and induced disk accretion in young star encounters. *Astrophys J.* 1994; 424:292.
88. Bate MR. On the diversity and statistical properties of protostellar discs. *Mon Not R Astron Soc.* 2018; 475:5618–5658.

89. Anderson KR, Adams FC, Calvet N. Viscous evolution and photoevaporation of circumstellar disks due to external far ultraviolet radiation fields. *Astrophys J.* 2013; 774:9.
90. Facchini S, Clarke CJ, Bisbas TG. External photoevaporation of protoplanetary discs in sparse stellar groups: the impact of dust growth. *Mon Not R Astron Soc.* 2016; 457:3593–3610.
91. Winter AJ, et al. Protoplanetary disc truncation mechanisms in stellar clusters: comparing external photoevaporation and tidal encounters. *Mon Not R Astron Soc.* 2018; 478:2700–2722.
92. Haworth TJ, Clarke CJ, Rahman W, Winter AJ, Facchini S. The FRIED grid of mass-loss rates for externally irradiated protoplanetary discs. *Mon Not R Astron Soc.* 2018; 481:452–466.
93. Concha-Ramírez F, Wilhelm MJC, Portegies Zwart S, Haworth TJ. External photoevaporation of circumstellar discs constrains the time-scale for planet formation. *Mon Not R Astron Soc.* 2019; 490:5678–5690.
94. Fatuzzo M, Adams FC. UV radiation fields produced by young embedded star clusters. *Astrophys J.* 2008; 675:1361–1374.
95. O'dell CR, Wen Z. Postrefurbishment mission Hubble Space Telescope images of the core of the Orion Nebula: Proplyds, Herbig-Haro Objects, and measurements of a circumstellar disk. *Astrophys J.* 1994; 436:194.
96. Winter AJ, Clarke CJ, Rosotti GP. External photoevaporation of protoplanetary discs in Cygnus OB2: linking discs to star formation dynamical history. *Mon Not R Astron Soc.* 2019; 485:1489–1507.
97. Tanaka H, Takeuchi T, Ward WR. Three-dimensional interaction between a planet and an isothermal gaseous disk. I. Corotation and Lindblad torques and planet migration. *Astrophys J.* 2002; 565:1257–1274.
98. Kennedy GM, Kenyon SJ. Planet formation around stars of various masses: hot super-Earths. *Astrophys J.* 2008; 682:1264–1276.
99. Ida S, Lin DNC. Toward a deterministic model of planetary formation. VI. Dynamical interaction and coagulation of multiple rocky embryos and super-Earth systems around solar-type stars. *Astrophys J.* 2010; 719:810–830.
100. Johansen A, Lambrechts M. Forming planets via pebble accretion. *Annu Rev Earth Planet Sci.* 2017; 45:359–387.
101. Lambrechts M, et al. Formation of planetary systems by pebble accretion and migration. How the radial pebble flux determines a terrestrial-planet or super-Earth growth mode. *Astron Astrophys.* 2019; 627:A83.
102. Ward WR. Protoplanet migration by nebula tides. *Icarus.* 1997; 126:261–281.
103. Ragusa E, et al. Eccentricity evolution during planet-disc interaction. *Mon Not R Astron Soc.* 2018; 474:4460–4476.
104. Pu B, Lai D. Eccentricities and inclinations of multiplanet systems with external perturbers. *Mon Not R Astron Soc.* 2018; 478:197–217.
105. Dwarkadas VV, Dauphas N, Meyer B, Boyajian P, Bojazi M. Triggered star formation inside the shell of a Wolf-Rayet bubble as the origin of the Solar System. *Astrophys J.* 2017; 851:147. [PubMed: 30905969]
106. Lichtenberg T, et al. A water budget dichotomy of rocky protoplanets from ^{26}Al -heating. *Nat Astron.* 2019; 3:307–313.
107. Malmberg D, Davies MB, Heggie DC. The effects of fly-bys on planetary systems. *Mon Not R Astron Soc.* 2011; 411:859–877.
108. van Elteren A, Portegies Zwart S, Pelupessy I, Cai MX, McMillan SLW. Survivability of planetary systems in young and dense star clusters. *Astron Astrophys.* 2019; 624:A120.
109. Chatterjee S, Ford EB, Matsumura S, Rasio FA. Dynamical outcomes of planet-planet scattering. *Astrophys J.* 2008; 686:580–602.
110. Owen JE, Lai D. Photoevaporation and high-eccentricity migration created the sub-Jovian desert. *Mon Not R Astron Soc.* 2018; 479:5012–5021.
111. Malavolta L, et al. The GAPS programme with HARPS-N at TNG. XI. Pr 0211 in M 44: the first multi-planet system in an open cluster. *Astron Astrophys.* 2016; 588:A118.

112. Pfalzner S, Bhandare A, Vincke K. Did a stellar fly-by shape the planetary system around Pr 0211 in the cluster M44? *Astron Astrophys.* 2018; 610:A33.
113. Pfalzner S, Bhandare A, Vincke K, Lacerda P. Outer Solar System possibly shaped by a stellar fly-by. *Astrophys J.* 2018; 863:45.
114. Hamers AS, Tremaine S. Hot Jupiters driven by high-eccentricity migration in globular clusters. *Astron J.* 2017; 154:272.
115. Li D, Mustill AJ, Davies MB. Fly-by encounters between two planetary systems I: Solar System analogues. *Mon Not R Astron Soc.* 2019; 488:1366–1376.
116. Fujii MS, Hori Y. Survival rates of planets in open clusters: the Pleiades, Hyades, and Praesepe clusters. *Astron Astrophys.* 2019; 624:A110.
117. Adams FC. The birth environment of the Solar System. *Annu Rev Astron Astrophys.* 2010; 48:47–85.
118. Gounelle M, Meynet G. Solar system genealogy revealed by extinct short-lived radionuclides in meteorites. *Astron Astrophys.* 2012; 545:A4.
119. Batygin K, Adams FC, Batygin YK, Petigura EA. Dynamics of planetary systems within star clusters: aspects of the Solar System's early evolution. *Astron J.* 2020; 159:101.

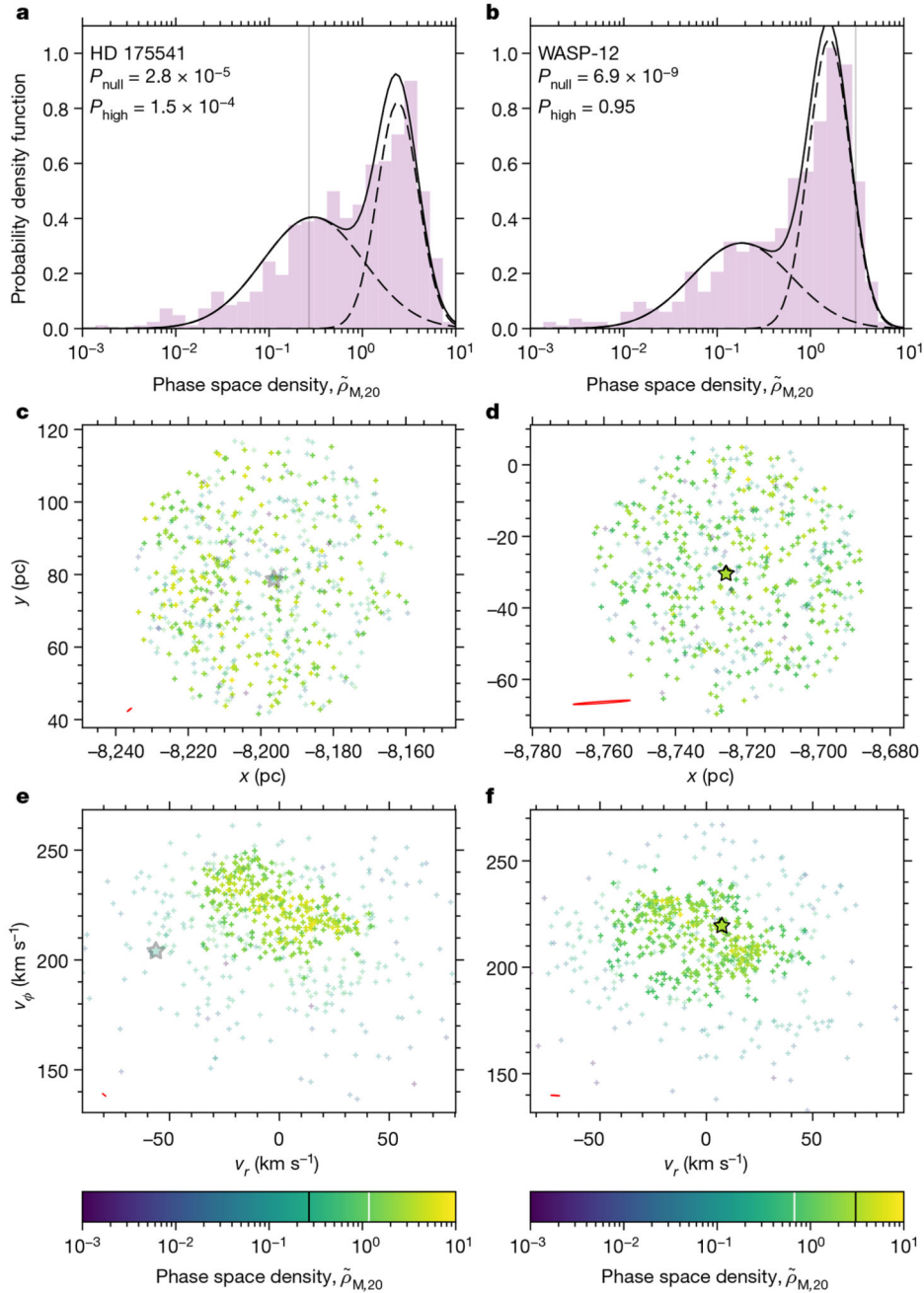


Fig. 1. Spatial and kinematic distributions of stars within 40 pc of two exoplanet host stars. The examples shown are HD 175541 (low phase space density; left column) and WASP-12 (phase space overdensity; right column). **a, b**, Histograms of the distribution of phase space densities (shaded area), with the best-fitting double-lognormal decomposition (black lines) and the relative phase space density of the host star (vertical grey line). The probability that the distribution follows a single lognormal function (P_{null}) and that the host star is associated with the overdensity (P_{high}) are shown. **c-f**, Projected spatial and kinematic distribution of stars in galactocentric coordinates. Data points are coloured by relative phase

space density. The black line on the colour scale marks the host, and the white line indicates where $P_{\text{high}} = P_{\text{low}} = 0.5$. Stars with $P_{\text{high}} < 0.5$ are shown as transparent points. The host is indicated with a star. Red ellipses indicate typical (1σ) astrometric errors.

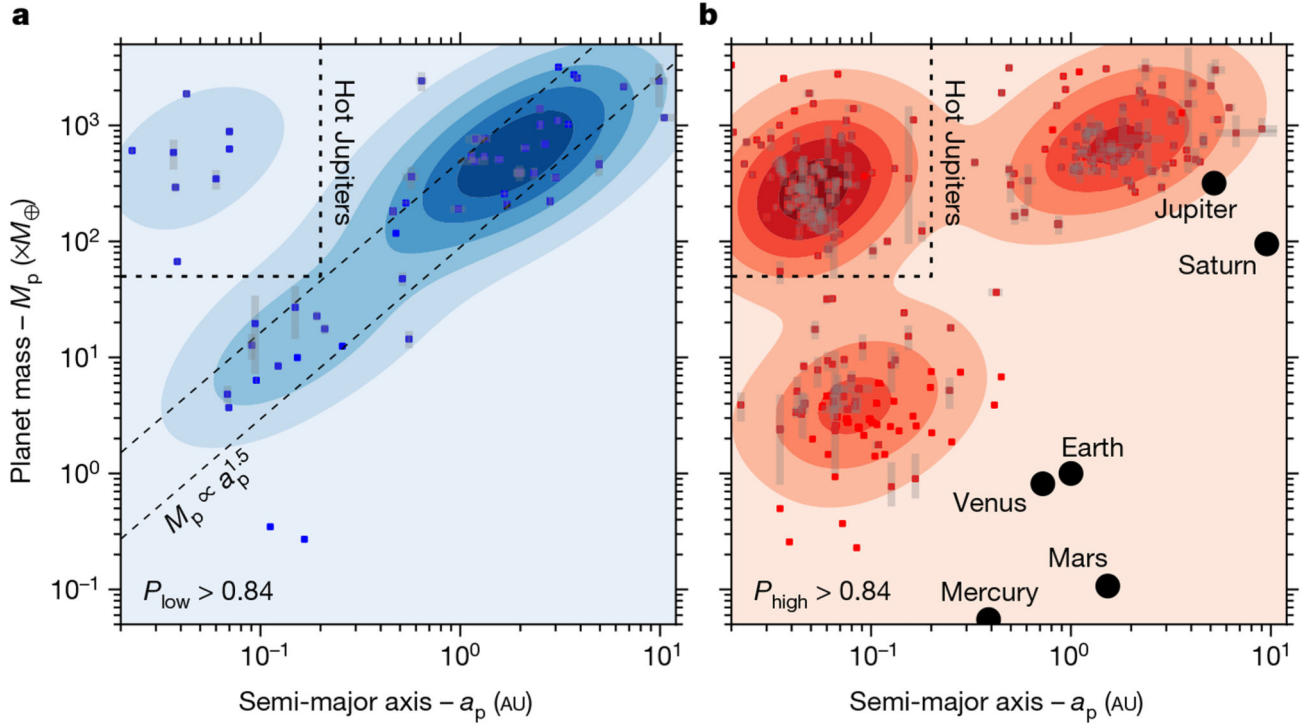


Fig. 2. Distributions of exoplanet semi-major axes and masses split by ambient stellar phase space density.

a, Low phase space densities ($P_{\text{low}} > 0.84$). **b**, High phase space densities ($P_{\text{high}} > 0.84$).

Data points with grey error bars (indicating 1σ uncertainties) show individual planets and contours show a two-dimensional Gaussian kernel density estimate. The dashed black lines in **a** follow $M_p \propto a_p^{1.5}$ and illustrate the 1σ scatter around an orthogonal distance regression to all planets orbiting field stars that are not ‘hot Jupiters’ (massive, close-in planets). Hot Jupiters fall outside this range and are mostly found in overdensities (**b**), suggesting that their extreme orbits originate from environmental perturbations. For reference, **b** includes the Solar System ($P_{\text{high}} = 0.89$) planets within $a_p < 10$ AU.

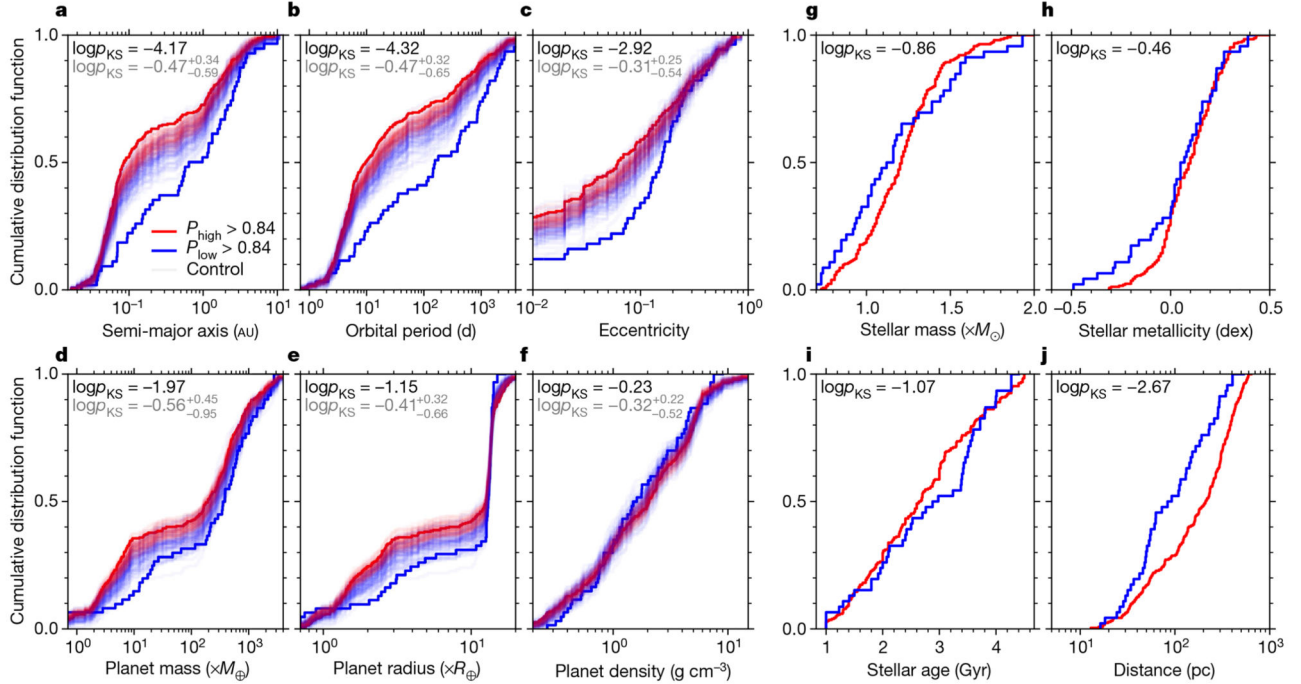


Fig. 3. Normalized cumulative distribution functions of planet and host star properties, split by ambient stellar phase space density.

Blue and red lines show low and high phase space densities, respectively. **a-f**, Exoplanet properties. **g-j**, Stellar host properties. The opaque lines show the observed distributions for the planets and host star properties. The faint lines represent 100 Monte Carlo control experiments, constructed by drawing a star at random from within 40 pc of each exoplanet host and using the phase space density of that star instead. The logarithm of p values obtained from a two-tailed Kolmogorov–Smirnov test for the exoplanet hosts (black) and for the median of all control experiments (grey; including 16th–84th percentile uncertainties) are shown. Differences between the low- and high-density samples are highly statistically significant for the orbital semi-major axis (**a**; $p_{KS} = 6.8 \times 10^{-5}$) and period (**b**; $p_{KS} = 4.8 \times 10^{-5}$), moderately significant for orbital eccentricity (**c**; $p_{KS} = 1.2 \times 10^{-3}$), and marginally significant for planet mass (**d**; $p_{KS} = 1.1 \times 10^{-2}$). These do not result from differences in host stellar mass (**g**), metallicity (**h**), age (**i**), or distance (**j** and control experiments in **a-f**).