PNAS RESEARCH ARTICLE | ASTRONOMY

# Planets larger than Neptune have elevated eccentricities

Gregory J. Gilbert[a,1], Erik A. Petigura[a], and Paige M. Entrican[a]

Edited by Neta Bahcall, Princeton University, Princeton, NJ; received March 13, 2024; accepted January 16, 2025

**NASA's Kepler mission identified over 4,000 extrasolar planets that transit (cross in front of) their host stars. This sample has revealed detailed features in the demographics of planet sizes and orbital spacings. However, knowledge of their orbital shapes—a key tracer of planetary formation and evolution—remains far more limited. We present measurements of eccentricities for 1,646 Kepler planets, 92% of which are smaller than Neptune. For all planet sizes, the eccentricity distribution peaks at $e = 0$ and falls monotonically toward zero at $e = 1$. As planet size increases, mean population eccentricity rises from $\langle e \rangle = 0.05 \pm 0.01$ for small planets to $\langle e \rangle = 0.20 \pm 0.03$ for planets larger than ∼3.5 Earth-radii $R_\oplus$. The overall planet occurrence rate and planet-metallicity correlation also change abruptly at this size. Taken together, these patterns indicate distinct formation channels for planets above and below ∼3.5 $R_\oplus$. We also find size-dependent associations between eccentricity, host star metallicity, and orbital period. While smaller planets generally have low eccentricities, there are hints of a noteworthy exception: eccentricities are slightly elevated in the "radius valley," a narrow band of low occurrence rate density which separates rocky "super-Earths" (1.0 to 1.5 $R_\oplus$) from gas-rich "sub-Neptunes" (2.0 to 3.0 $R_\oplus$). We detect this feature at 2.1$\sigma$ significance. Planets in single- and multitransiting systems exhibit the same size–eccentricity relationship, suggesting they are drawn from the same parent population.**

exoplanets | orbital eccentricities | planetary dynamics | transits

## Significance

The eccentricity (ellipticity) of a planet's orbit is a relic of its formation history. We measured eccentricities of 1,646 planets with sizes ranging from 0.5 to 16 Earth-radii ($R_\oplus$). On average, large planets (4 to 16 $R_\oplus$) are four times more eccentric than small planets (0.5 to 4 $R_\oplus$), pointing to distinct formation channels for these two size groups. Small planets typically form on nearly circular orbits and experience minimal perturbations, while large planets are more likely to experience eccentricity excitation. Small planets are bifurcated into at least two groups, super-Earths (1.0 to 1.5 $R_\oplus$) and sub-Neptunes (2.0 to 3.0 $R_\oplus$), with few planets in between. The planets that fall between these two populations may also have elevated eccentricities, pointing to dynamically exotic formation histories.

The eight planets of the Solar System have nearly circular orbits, a fact that provided early motivation for the nebular hypothesis of planet formation (1). Their small but nonzero eccentricities (ellipticities) have a mean value $\langle e \rangle = 0.06$, which has been interpreted as the result of moderate orbital migration and resonance crossing (2). The Solar System planets may be grouped according to three size classes: terrestrials ($R_p$ = 0.5 to 1.0 Earth-radii or $R_\oplus$), ice-giants (∼4 $R_\oplus$), and gas-giants (∼10 $R_\oplus$). In contrast, NASA's Kepler mission showed that extrasolar planets span a continuum of sizes and that small planets, with sizes between 1 and 4 $R_\oplus$ and orbital periods less than a year, are ubiquitous, and occur at the rate of ∼1 per star (3). Small planets outnumber large planets by an order of magnitude (4). Moreover, small planets are bifurcated into two distinct groups: rocky super-Earths (1.0 to 1.5 $R_\oplus$) and gas-rich sub-Neptunes (2 to 3 $R_\oplus$) with few planets in between (5). While the size distribution of close-in small planets has been characterized to high detail, their eccentricities remain far more uncertain.

Giant exoplanets are known to possess a wide range of eccentricities, from nearly circular ($e = 0$) to highly elliptical (the current record-holder is HD20782b with $e$ = 0.97, 6). Hot Jupiters, with orbital periods $P < 10$ d and sizes ranging from 1 to 2× Jupiter, tend to have small eccentricities, likely the result of tidal circularization (7–9). In contrast, longer-period Doppler-detected Jovians have a wide range of eccentricities with a mean value of $\langle e \rangle \approx 0.3$ (10). Doppler measurements of small planet eccentricities are often prohibitively expensive since their Doppler amplitudes are much smaller.

Transit-based analyses have provided some insight into the eccentricities of small planets. In these studies, the uncertainties on individual planet's eccentricities are large ($\sigma_e \sim 0.3$ is a typical value; e.g., ref. 11). Fortunately, the census of transiting planets is large enough that many imprecise measurements may be combined to yield insights into the population-level distribution of eccentricities. Prior studies have found that planets in single-transiting systems are more eccentric than planets in multis: $e_{\rm single} \sim 0.3$ and $e_{\rm multi} \sim 0.05$ (11–16). These studies fall into two categories: 1) analyses that used transit durations (12–14), and 2) analyses that modeled the full transit profile (11, 15, 16). Duration-based studies are computationally efficient, but lead to a loss of information. Transit profiling involves more human and computational effort but to date has been restricted to samples of ∼100 planets. Our present work extends the latter method to over 1,000 planets, enabling a more detailed exploration of eccentricity, in particular its relation to other star and planet properties.

Author affiliations: [a]University of California, Los Angeles, CA 90095-1547

Author contributions: G.J.G. and E.A.P. designed research; G.J.G., E.A.P., and P.M.E. performed research; G.J.G., E.A.P., and P.M.E. analyzed data; and G.J.G. and E.A.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: gjgilbert@astro.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405295122/-/DCSupplemental.

Published March 3, 2025.

The basic paradigm of planet formation by core-nucleated accretion has been remarkably successful at explaining the observed population of close-in exoplanets (17). However, when, where, and how planet cores grow; accrete and lose gaseous envelopes; and dynamically interact with their siblings remains poorly understood. Many proposed processes imprint themselves on the extant distributions of exoplanet sizes, eccentricities, and other orbital characteristics. For example, planet–planet scattering (18) and Lidov-Kozai oscillations (19) excite eccentricity, whereas tidal damping (8), inelastic mergers (20), and disk-driven migration (21) quench it. Each mechanism predicts distinct relationships between eccentricity $e$, planet radius $R_p$, host star metallicity [Fe/H] (i.e., stellar inventory of elements heavier than helium), and orbital period $P$. The goal of this paper is to uncover these relationships.

## Methods

We begin with a high-level summary of our methods here before expanding upon them in the sections below. We first select a sample of 1,209 single Sun-like stars hosing 1,646 planets (Fig. 1). Kepler gathered some 40,000 photometric measurements of each star over its four-year mission. Each transit may be described with five transit shape parameters. We derive these parameters with particular care to preserve covariances and uncertainties, thereby compressing the high-dimensional photometric dataset into a low-dimensional catalog of transit shape measurements. Next, we combine transit shape and stellar density measurements to derive eccentricities of individual planets. Finally, we characterize the eccentricity distributions of various subsamples of planets with a hierarchical model. The flow from photometry to eccentricities is shown in Fig. 2.

**Defining the Star/Planet Sample.** From 2009 to 2013, the Kepler spacecraft continuously monitored ~150, 000 Sun-like stars, yielding a sample of over 4,000 exoplanet candidates with high reliability (false positive rate below 2% for $P < 100$ d; 22) and well-characterized completeness (23). This enabled the measurement of exoplanet occurrence rates as a function of planet size and period (see ref. 24 and references therein). The Kepler team measured orbital period to exquisite precision, and subsequent efforts have measured planetary sizes to ~3 to 10% for the bulk of the sample (e.g., refs. 5 and 25).

Starting with the final Kepler catalog of planet candidates (22), we eliminated any objects marked as false positives, leaving 4,078 planets orbiting 3,087 stars. We cross-matched these stars against a single homogeneous stellar properties catalog based on *Gaia* astrometric and photometric measurements (26) and selected stars with the following properties: radius $R_\star = 0.7$ to $1.4\ R_\odot$, effective temperature $T_{\rm eff} = 4{,}700$ to 6,500 K, and surface gravity $\log g > 4.0$. These criteria correspond to main sequence stars with masses $M_\star = 0.7$ to $1.3\ M_\odot$, or mid-K to late-F spectral type. Finally, we eliminated any stars with greater than 5% flux contamination from nearby sources, as identified by 27, as well as any stars with a Gaia renormalized unit weigh error RUWE > 1.4; such objects have a high probability of being unresolved binaries with separations between 0.1 and 1.0 arcsec and $G$-band contrasts of 3 magnitudes or less (28). Finally, we eliminated all stars with a fractional uncertainty in radius that exceeded 20%.

Throughout this work, we considered only planets between $P = 1$ to 100 d and $R_p = 0.5$ to $16\ R_\oplus$ (corresponding roughly to Mars- to Jupiter-size objects). All together, these restrictions reduced our sample to 1,646 planets and 1,209 stars (Fig. 1). Roughly half of the planets (848) belong to single-transiting systems, while the other half (798) belong to multitransiting systems.

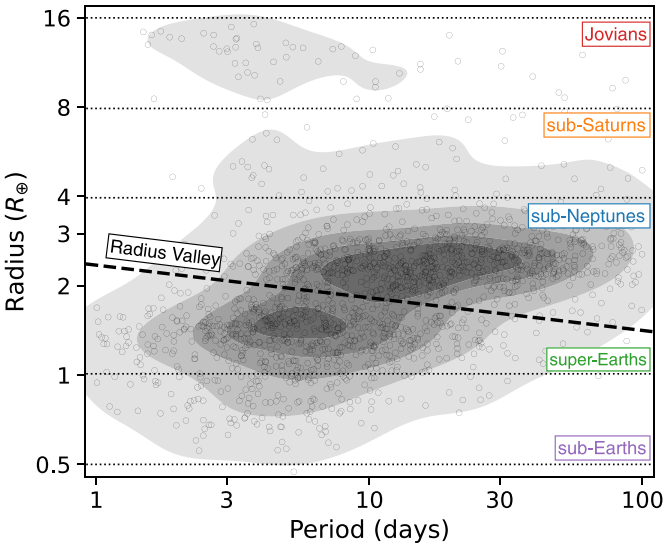


**Fig. 1.** Our sample consists of 1,646 planets with orbital periods $P = 1 - 100$ d and sizes $R_p = 0.5 - 16 R_\oplus$. Each circle represents a single planet, with isodensity contours drawn to guide the eye. Planets may be broadly classified based on size as sub-Earths (purple), super-Earths (green), sub-Neptunes (blue), sub-Saturns (orange), and Jovians (red). Dashed lines denote our operational definition for each group, with consistent colors adopted throughout this manuscript. The two most populous groups—super-Earths and sub-Neptunes—are separated by a "radius valley" whose center is given by $R_p = 1.84 (P/10\ {\rm d})^{-0.11}\ R_\oplus$(29). Planets below the radius valley are mostly rocky, whereas those above the valley possess cores with sufficient gravity to maintain massive gas envelopes.

**Inferring the Transit Model from Photometry.** We modeled the transit profiles using `ALDERAAN`, an open-source pipeline we developed for this and related projects.* `ALDERAAN`takes in raw Kepler photometric observations to produce posterior samples of period $P$, transit epoch $t_0$, planet-to-star radius ratio $R_p/R_\star$, impact parameter $b$, and first-to-fourth contact transit duration $T$. It also models transit timing variations (TTVs, i.e., deviations from a linear ephemeris) if necessary.

The first step of our procedure was to detrend (prewhiten) the photometric lightcurves. To maintain homogeneity in our dataset, we used long cadence (30 min) photometry for all targets. We began with the presearch data conditioning simple aperture photometry (30), hosted at the Mikulski Archive for Space Telescopes. We accessed the data via the NASA Exoplanet Archive on 1 October 2024 (31). We removed residual instrumental noise and astrophysical stellar variability using a Gaussian Process (GP) regression. `ALDERAAN`automatically detected gaps in time (i.e., missing data) or jumps in flux (i.e., discontinuities in brightness values) and treated each disjoint section of data independently. To avoid fitting out the transit profile with the GP model, we excluded three times the total transit duration from the regression. Transit durations were determined by querying the DR25 catalog (22) and TTVs (where applicable) were determined by querying a uniform TTV catalog (32).

Following detrending, our next step was to produce a self-consistent estimate of the transit parameters $\{P, R_p/R_\star, b, T\}$ and TTVs. Accurate determination of TTVs is important because

*`ALDERAAN`stands for "Automated Lightcurve Detrending, Exoplanet Recovery, and Analysis of Autocorrelated Noise." The full pipeline can be accessed at https://www.github.com/gjgilbert/alderaan.

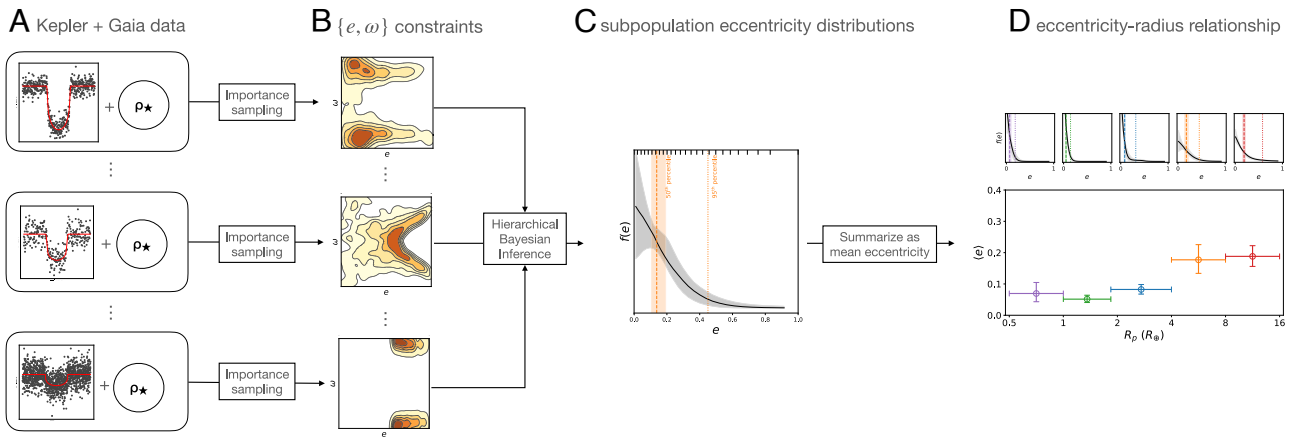


**Fig. 2.** Schematic description of our methods. (*A*) We began with 1,646 photometric lightcurves observed by Kepler. We fit each lightcurve using a five-parameter transit model to produce posterior samples of period $P$, transit epoch $t_0$, planet-to-star-radius ratio $R_p/R_\star$, impact parameter $b$, and transit duration $T$. (*B*) By comparing our transit-derived $T$ to the value predicted for a circular, center-crossing transit $T_0$ and applying the importance sampling scheme developed in refs. 41 and 45, we derived joint posterior samples for eccentricity $e$ and argument of pericenter $\omega$. (*C*) After dividing planets into various size bins, we fit a hierarchical Bayesian model (49) to determine each subpopulation's eccentricity distribution $f(e)$. We modeled the distribution nonparametrically to infer the functional form of the distribution and found a self-similar shape across planet sizes. We used this shape to construct an empirical distribution template. (*D*) Finally, we recalculated the hierarchical model using our empirical template, allowing us to efficiently "slice-and-dice" the population by planet radius and other variables. To summarize each subpopulation, we computed mean eccentricity $\langle e \rangle$ and its uncertainty .

unaccounted for TTVs will smear out transit ingress and egress and bias eccentricity measurements (15, 33). The two issues are coupled: Accurate determination of the transit shape is needed to infer accurate TTVs, and vice versa. We employed an iterative solution: First, we adopted a nominal TTV model, and then, we determined the transit shape, refined the TTV model, and repeated. An overview is as follows, with further details in *SI Appendix*.

We first fixed the transit times to the values in ref. 32 and calculated the *maximum* a posteriori values of the transit parameters, producing a first estimate for the transit shape. Stellar limb darkening coefficients were held constant at values derived from Gaia measurements (26, 34) and stellar atmosphere models (35, 36). We then held the transit parameters fixed and measured each transit time individually by cross-correlating the transit template across a grid of transit center time offsets $t_c$, identifying the best-fit transit time from the $\chi^2$ minimum. For low signal-to-noise transits, individual $t_c$ measurements are noisy, so we applied a regularization scheme. First, we clipped $5\sigma$ outliers, after which we tested the following suite of models: a Matern-3/2 GP, a single-component sinusoid, and polynomials of first, second, and third degree. We selected the model favored by the Akaike Information Criterion (37).

To produce our final set of detrended lightcurves, we repeated the detrending procedure above using our self-consistent TTVs. For this second iteration, we set the width of the GP interpolation window for each known transit to $\Delta t = T + 3\sigma_{\rm TTV} + 30$ min, where $\sigma_{\rm TTV}$ is the RMSD of our individually measured transit times compared to our fiducial regularized model. This interpolation window was usually much narrower than the window used during the initial detrending, which produced a better estimate of the true variability in and near transit.

To extract transit parameters, we fit each planet using a five-parameter transit model $\{c_0, c_1, R_p/R_\star, b, T\}$, where $c_0$ and $c_1$ were linear perturbations to the fiducial TTV model. Priors on all parameters were set to broad, uninformative normal $(c_0, c_1)$, uniform $(b)$, or log-uniform $(R_p/R_\star, T)$ distributions. The model also included two quadratic stellar limb darkening coefficients $\{q_1, q_2\}$, following ref. 38 to ensure a physical profile. During sampling, we converted $q_1, q_2$ into the standard physical limb darkening coefficients $\{u_1, u_2\}$, where $q_1 = (u_1 + u_2)^2$ and $q_2 = 0.5u_1(u_1 + u_2)^{-1}$. We then applied Gaussian priors on $\{u_1, u_2\}$ with SD $\sigma(u) = 0.1$ and mean values informed by stellar properties; each nominal mean was determined using Gaia stellar parameters (26, 34) and PHOENIX stellar atmosphere models (35, 36). In multiplanet systems, we fit all planets simultaneously which involved a total of $5 \times N + 2$ free parameters.

Accurately determining the credible set of transit shape parameters that are consistent with the observed data requires care. In general, the posterior probability distributions exhibit thin, curving covariances between $b$, $R_p/R_\star$, and $T$. These covariances are especially strong for grazing and near-grazing transits ($b \gtrsim 0.7$) and present a challenge for standard Monte Carlo samplers (39). Previous analyses of Kepler photometry have struggled to accurately characterize the impact parameter (39–41). Prior work has demonstrated that nested sampling is well suited to this posterior topology (42–45).

To ensure our modeling proceeded as expected, we systematically inspected each of the 1,646 lightcurve fits. To facilitate the process, we built a custom visualization tool that displays the following data products: the phase-folded light curve, credible model draws, and model residuals; the individual transits; time series of TTVs; and posterior corner plots. We verified the detrending step by confirming that the model residuals in- and out-of-transit were white, Gaussian, and homoscedastic. We ensured the TTVs models were not over- or underfit by direct inspection of the TTV timeseries and by checking the folded transit model for excess noise during ingress/egress (a sign of unaccounted for TTVs). We also inspected the 2 d joint posteriors, and confirmed that the covariances between $b$, $T$, and $R_p/R_\star$ met expectations for each transits S/N and morphology (e.g., $b$ for V-shaped transits should extend beyond 1). Of the 1,646 planets, only 89 (5%) were flagged for manual intervention. In these cases, we refit the transit model using an alternative TTV model selection routine. We found that our inferences of population eccentricity were virtually identical whether we adopted pre- or postintervention posterior chains for these 89 planets.

**Extracting Eccentricity Constraints from the Transit Model.** Although our model did not directly encode eccentricity, we may compare the observed duration $T$ to the predicted duration $T_0$ for a transit on a circular ($e = 0$), center-crossing ($b = 0$) orbit around a star with density $\rho_\star$ to constrain eccentricity on a planet-by-planet basis (33, 46, 47). The two durations are related via:

$$\frac{T}{T_0} \approx \sqrt{1-b^2}\left(\frac{\sqrt{1-e^2}}{1+e\sin\omega}\right). \quad [1]$$

The exact expression is given in *SI Appendix*. Physically, the $\{e, \omega\}$ term arises from the change in orbital velocity from nonzero eccentricity, while the $\sqrt{1-b^2}$ term arises from the shortened transit chord for nonzero impact parameter.

If impact parameter is measured well, then a significant difference between $T$ and $T_0$ implies a nonzero eccentricity. For exceptionally high signal-to-noise transits this is possible because ingress/egress timescale $\tau$ encodes the impact parameter. However, for the majority of Kepler light-curves, constraints on $b$ are broad. Thus, most of the information about $b$, $e$, and $\omega$ is encoded in $T$ and the inversion is underconstrained. Nevertheless, some information can yet be extracted from even imprecise measurements of $\tau$, which is why the "full lightcurve profiling" method which considers joint samples $\{b, \rho_\star, e, \omega\}$ outperforms the simpler "duration-only" method.

We converted our transit-derived posterior samples of $\{P, R_p/R_\star, b, T\}$ to joint samples of $\{P, R_p, b, \rho_\star, e, \omega\}$ following the conceptual framework outlined above and implementing an importance sampling routine described in refs. 41 and 45. A brief summary of the procedure is as follows. First, we drew random samples of $\{e, \omega\}$ from uniform distributions $e \sim \mathcal{U}(0, 1)$, $\omega \sim \mathcal{U}(0, 2\pi)$ and combined these with observed samples $\{P, R_p/R_\star, b, T\}$. For each sample, we computed physical stellar density $\rho_{\star,\mathrm{samp}}$ required to produce the observed $T$ given $\{P, R_p/R_\star, b, e, \omega\}$ (see ref. 48 and *SI Appendix*). We drew an identical number of stellar density $\rho_\star$ samples based on the Gaia constraints. Then, we computed the log-likelihood of each set of $\{\rho_{\star,\mathrm{samp}}, \rho_\star\}$:

$$\log \mathcal{L}_i = -\frac{1}{2}\left(\frac{\rho_{\star,\mathrm{samp},i} - \rho_{\star,i}}{\sigma_{\rho_{\star,i}}}\right)^2. \quad [2]$$

We converted each sample's $\mathcal{L}_i$ into an importance weight

$$w_i = \frac{\mathcal{L}_i}{\sum_i \mathcal{L}_i}. \quad [3]$$

Finally, we converted $R_p/R_\star \rightarrow R_p$ using the Gaia-derived $R_\star$. To simplify downstream calculations, converted our weighted samples to unweighted samples of $\{P, R_p, b, \rho_\star, e, \omega\}$ for each planet.

The posterior samples generated by the above procedure are equivalent to samples which would have been generated by directly fitting an eccentric transit model to the lightcurves using Gaia-informed Gaussian priors on $\rho_\star$ and uninformative priors on all other parameters (45). The resulting posterior distributions on $e$ have a typical spread $\sigma(e) \sim 0.3$ and are asymmetric and heavy-tailed (see *SI Appendix* for some examples).

To ensure we included the highest quality measurements, we removed 46 planets which had greater than 20% fractional uncertainty on $R_p$. We also removed 63 planets with more than 5% of samples met the criterion $b > 1 - R_p/R_\star$, as this region of parameter space contains pathological covariances which frustrate accurate sampling (39). We also removed 4 planets where less than 5% of the importance weights $w_i$ exceeded machine precision; this condition demonstrates a fundamental inconsistency between Gaia stellar values and our derived transit parameters and could indicate inaccurate star and/or planet properties. We experimented with these selection criteria, finding our results were insensitive to exact cutoffs on $R_p$ precision, grazing $b$ fraction, or importance weight efficiency.

**Inferring the Distribution of Exoplanet Eccentricities.** Care is required to characterize population eccentricity distribution, given the varied and non-Gaussian nature of the eccentricity constraints of individual objects. We accomplished this using an approximate hierarchical Bayesian model, following the formalism first described by Hogg et al. (49). This technique has previously been used to infer exoplanet eccentricity distributions by Van Eylen et al. (11), Sagear and Ballard (16), and Bowler et al. (50). Here, we present a brief overview of the method, with mathematical details enumerated in *SI Appendix*.

Given a population of $N$ planets each with $K$ samples from its own eccentricity posterior distribution, $e_{nk}$, the combined likelihood for the population distribution $f(e)$ is

$$\mathcal{L}_\alpha = \prod_{n=1}^{N} \frac{1}{K}\sum_{k=1}^{K} \frac{f_\alpha(e_{nk})}{p_0(e_{nk})}. \quad [4]$$

Here, $p_0(e)$ is the uninformative "interim" prior on $e$ applied during transit modeling and $f_\alpha(e)$ is the informative "updated" population distribution we wish to infer; the subscript $\alpha$ denotes the vector of hyperparameters describing $f(e)$. Conceptually, the ratio $f_\alpha/p_0$ captures the degree to which our informative eccentricity population model improves upon an uninformative prior, marginalized over all covariant quantities. While the hyperpriors on the interim distribution $p_0(e)$ are fixed, the hyperpriors on the target distribution $f_\alpha(e)$ are themselves open to inference. Determining $\{\alpha\}$ for some assumed functional form $f_\alpha(e)$ enables inference of the optimal population model.

For our specific application to eccentricity, we modified $\mathcal{L}_\alpha$ to account for geometric biases which cause preferential detection of exoplanets on eccentric orbits (51–53). The full marginalized likelihood accounting for detection biases is

$$\mathcal{L}_\alpha = \prod_{n=1}^{N} \frac{1}{K}\sum_{k=1}^{K} f_\alpha(e_{nk})\left(\frac{1-e_{nk}^2}{1+e_{nk}\sin\omega_{nk}}\right), \quad [5]$$

where $p_0 = 1$ because we sampled $\{e, \omega\}$ from a uniform interim prior. Omitting this geometric term would result in the measurement of transiting planet eccentricity distribution rather than the intrinsic eccentricity distribution. Prior works have adopted both conventions (see, e.g., refs. 11 and 16) but the latter permits a straightforward comparison to Doppler studies. Since most eccentricities for real planets are, in fact, low, the inclusion of the detection efficiency term has only a small effect on $f_\alpha(e)$.

The true functional form of $f_\alpha(e)$ is presently unknown, although the half-normal, Rayleigh, and beta distributions are popular parametric models in the literature (10, 11, 13, 14, 16). The first two have a single free parameter, while the third has two free parameters. Imposing any of these functions at the outset runs the risk of choosing a model that cannot accurately describe the data. Therefore, we adopt a more flexible approach and use a regularized histogram instead (54, 55). The benefit of this method

is that we make no assumption about the form of the distribution (unimodal, symmetric, etc.) other than that it must be smooth. We fit the distribution $f_\alpha(e)$ using a 25-bin histogram (resolution $\Delta e \sim 0.04$), with distribution "smoothness" enforced via a Gaussian process prior on the bin heights (see Fig. 2 for a visual representation of the histogram bin positions). In this context, our hyperparameters $\{\alpha\}$ are the 25 logarithmic histogram bin heights, plus two parameters specifying the smoothing function.

To investigate whether and how the shape of the eccentricity distribution changes as a function of planet size, we divided the planet sample into five physically motivated subpopulations in $R_p$ and derived $f_\alpha(e)$ independently for each (see *SI Appendix* for details). We find that all subpopulations exhibit the same qualitative shape for $f(e)$: a mode at $e = 0$ and a monotonic, quasi-exponential descent toward zero at $e = 1$ (Fig. 2). Compared to the commonly used parametric distributions in the literature, our flexible distribution most resembles the beta distribution. This is consistent with the sample of Doppler measurements of giant planets, whose eccentricities are well described by a beta distribution (10). The Rayleigh distribution and half-normal distribution are poor fits at the low and high ends of the eccentricity range respectively (see *SI Appendix* for discussion).

Given that the different subpopulations were described by similar $f(e)$ shapes, we adopted an empirical template for $f(e)$:

$$\ln f_E(e; \nu, h) = \nu\,[\ln f_0(he)] + (1-\nu)\,[\ln f_0(0)]\,. \qquad [6]$$

Here, $f_0(e)$ is derived by fitting a hierarchical model to our full planet sample. The empirical $f_E(e)$ has two free parameters: $\nu$ sets the tail weight, and $h$ sets the central width of the distribution. These transformations do not necessarily preserve normalization, so after transforming $f_E(e)$ we numerically renormalized the distribution. Moving forward, we assume that $f(e)$ is self-similar across subpopulations and fit models by determining these two hyperparameters $\{\nu, h\}$.

**Characterizing Mean Eccentricity for Subpopulations.** Now that we have determined the shape of the eccentricity distribution, we refit various subpopulations in order to explore the covariance between $e$ and other quantities of interest, in particular $P$, $R_p$, and [Fe/H]. For each subpopulation, we summarized the distribution eccentricity as $\langle e\rangle \pm \sigma_{\langle e\rangle}$, where $\langle e\rangle$ is a median-of-means. More specifically, for each subpopulation, we have generated a set of samples of $f(e)$. For each $\{\nu, h\}_i$ we generated a $f(e; \nu, h)$ and computed its mean $\langle e\rangle_i$. We report the median of the $\langle e\rangle_i$ along with their 16th and 84th percentiles to convey uncertainties. An illustration of this data compression process is shown in Fig. 2.

## Analysis and Results

Figs. 3–5 summarize the interrelationships between eccentricity and planetary radius, orbital period, and stellar metallicity. We discuss each relationship below.

**The Relationship between Eccentricity and Planet Radius.** To investigate the $\langle e\rangle - R_p$ relationship, we split the sample in the following way. First, we separated single- and multitransiting systems (henceforth referred to as "singles" and "multis"), as previous works have consistently found significant differences between these two populations (11, 13, 14, 16). We also divided the populations into narrower bins of width in log-radius of 20% for planets between 1 and $4R_\oplus$ and 30 or 40% for smaller/larger

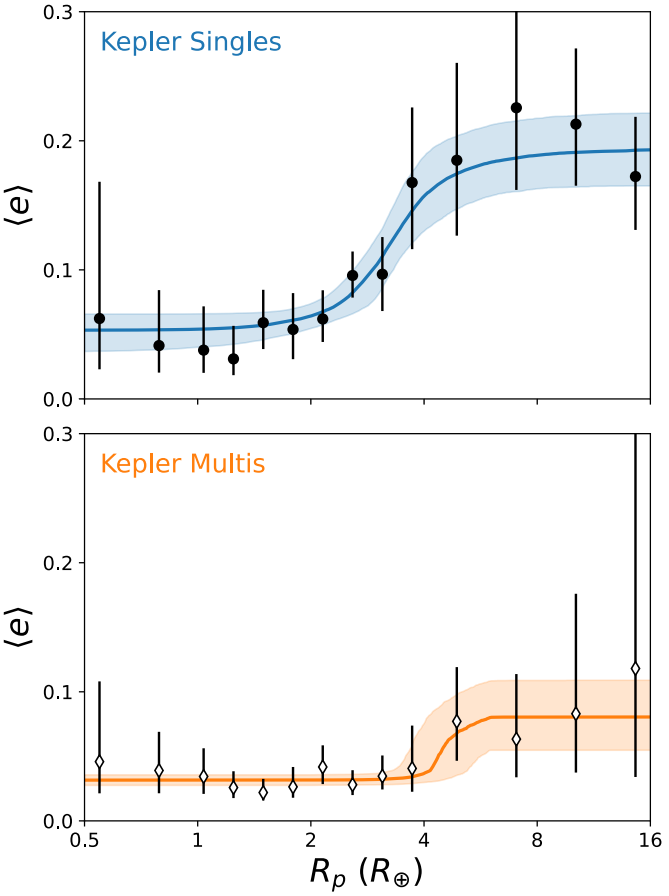


**Fig. 3.** Mean eccentricity $\langle e\rangle$ as a function of planet radius $R_p$ for single-transiting systems (*Top*) and multitransiting systems (*Bottom*). Singles and multis exhibit a similar $\langle e\rangle$-$R_p$ relationship, with planets larger than Neptune having about three times the eccentricity of smaller planets. Singles are 2.4 ± 0.9 times more eccentric than multis across the full range of planet sizes.

planets (where lower numbers motivated coarser binning; see Table 1).

Because most planets had fractional radius uncertainties $\sigma(R_p)\sim 10\%$ that were comparable to bin size, we followed (56) and modified the hierarchical Bayesian likelihood (Eq. **5**) to account for their size uncertainties:

$$\mathcal{L}_\alpha = \prod_{n=1}^{N}\left[\frac{1}{K}\sum_{k=1}^{K} f_\alpha(e_{nk})\left(\frac{1-e_{nk}^2}{1+e_{nk}\sin\omega_{nk}}\right)\right]^{w_n}. \qquad [7]$$

We computed the weights $w_n$ as follows: first, we defined each bin as a Gaussian

$$g(R) = \frac{1}{\sigma_R\sqrt{2\pi}}\exp\left(-(R-R_0)^2/2\sigma_R^2\right), \qquad [8]$$

where $R_0$ is the bin center, and $\sigma_R$ is the bin width. We then calculated each planet's weight by summing over all $K$ samples as

$$w_n = \frac{1}{K}\sum_{k=1}^{K} g(R_{nk}). \qquad [9]$$

We then normalized the weights so that the expected value $\langle w\rangle = 1$.

We found that fits using the weighted likelihood (Eq. **7**) were consistent with the unweighted likelihood (Eq. **5**). We elected to

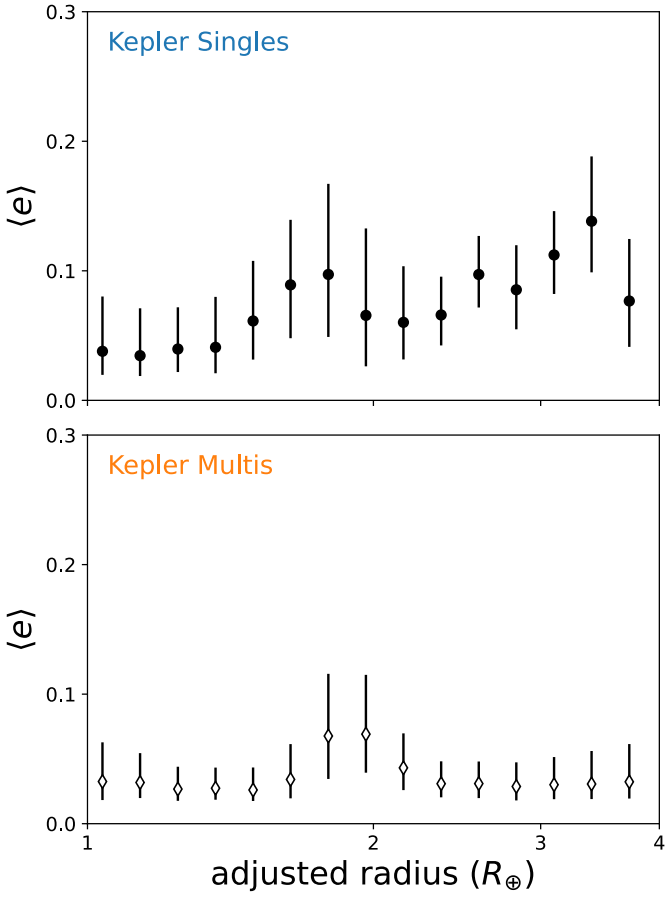


**Fig. 4.** Mean eccentricity ⟨*e*⟩ as a function of adjusted planet radius near the radius gap for single-transiting systems (*Top*) and multi-transiting systems (*Bottom*). The "adjusted radius" is calculated following refs. 29 and 57 to account for the period dependence of the radius gap. Planets in the gap show tentative evidence of elevated ⟨*e*⟩ compared to planets on either side of the gap.

use the former because it incorporated radius uncertainties that are comparable to the narrowest bin widths.

Both singles and multis exhibit a similar relationship between ⟨*e*⟩ and $R_p$ (Fig. 3). Eccentricity is low between 0.5 and 2.0 $R_\oplus$ and then rises to an elevated plateau above 6 $R_\oplus$. The transition between low and high eccentricity is somewhat more gradual for singles. Overall, ⟨*e*⟩ is higher for singles than for multis, consistent with previous analyses (11, 13, 14, 16).

To quantify the transition feature in this relationship, we fit the ⟨*e*⟩-$R_p$curve with a logistic sigmoid

$$f(x) = B + \frac{L}{1 + e^{-k(x-x_t)}}. \quad [10]$$

Here, $x \equiv \log R_p$, $B$ encodes the baseline eccentricity, $L$ encodes the overall normalization, and $\{x_t, k\}$ encodes the location and rate of change from low to high $e$. In physical terms, $e_{\rm low} = B$ and $e_{\rm high} = L + B$.

The shape of the ⟨*e*⟩-$R_p$curve is similar for singles and multis. The transition from low-to-high $e$ occurs at $R_p = 3.3 \pm 0.4\ R_\oplus$for singles and at $R_p = 4.2 \pm 0.9\ R_\oplus$for multis. For singles, $e_{\rm high}/e_{\rm low} = 3.9^{+1.9}_{-1.0}$ while for multis $e_{\rm high}/e_{\rm low} = 2.5^{+1.0}_{-0.8}$. Thus, both the location and amplitude of the transition agree within $1\sigma$ between singles and multis.

Overall, the mean eccentricity of singles is higher than that of multis by a factor of $2.4^{+1.5}_{-0.7}$. In the flat region of the curve on the small planet end ($R_p < 1.5\ R_\oplus$), we compute $e_{\rm singles}/e_{\rm multis} = 1.7^{+0.5}_{-0.4}$, while on the large planet end ($R_p > 6.0\ R_\oplus$) we compute $e_{\rm singles}/e_{\rm multis} = 2.5^{+1.4}_{-0.7}$.

For intermediate sizes where the curves are most visually discrepant, $e_{\rm singles}/e_{\rm multis}$ ranges from 1.3 to 6.2 (95% confidence). Thus, the single-to-multi-enhancement factor for eccentricity is consistent with a constant (or nearly constant) value across the full range of planet sizes.

This result is broadly consistent with measured single-to-multi-enhancement factors of $8 \pm 7$ (13), $4.8 \pm 1.8$ (14), $3.9 \pm 1.1$ (11), and $6.8 \pm 3.7$ (16). However, previous works investigated different samples of planets, so one should not necessarily expect the offsets to be identical.

Small planets are bifurcated into two populations: rocky super-Earths ($R_p \lesssim 1.5\ R_\oplus$) and gas-rich sub-Neptunes ($R_p \gtrsim 2.0\ R_\oplus$), with few planets in between, a demographic feature known as the "radius gap" or the "radius valley" (5). More precisely, the radius gap is a 2 d feature in $P$-$R_p$ space (Fig. 1). The center of the gap is a gently declining function of orbital period (Fig. 1)

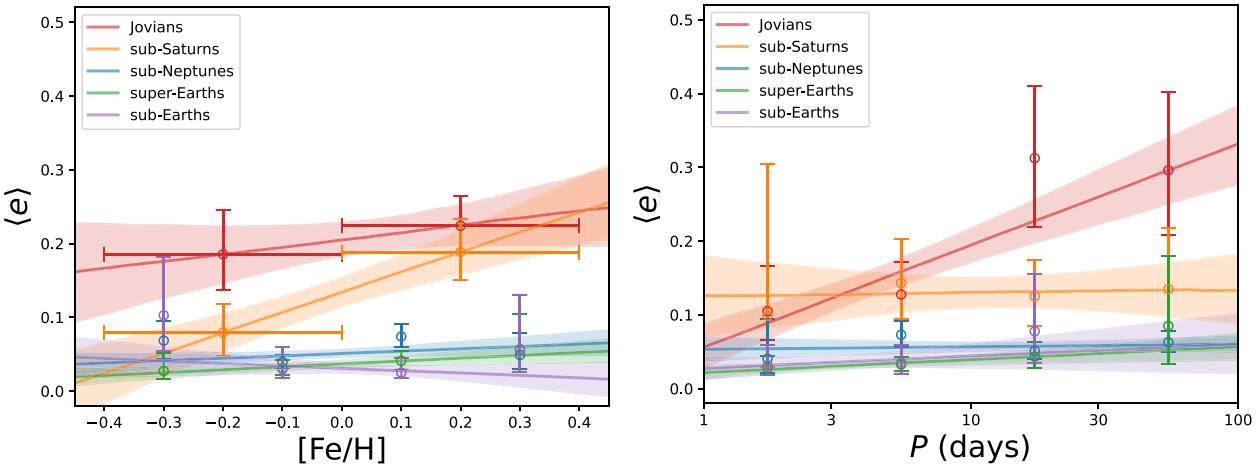


**Fig. 5.** Mean eccentricity ⟨*e*⟩ as a function of (*A*) host star metallicity [Fe/H] and (*B*) orbital period *P* for each planet size class. Planets larger than four Earth-radii exhibit elevated eccentricities compared to their smaller counterparts across all values of [Fe/H] and *P*. *Left*: small planets exhibit a weak association between [Fe/H] and ⟨*e*⟩, whereas large planets exhibit a modest positive correlation. *Right*: associations between eccentricity and period depend on planet size. Jovians exhibit a positive correlation between ⟨*e*⟩ and *P*, sub-Saturns are consistent with no correlation at a consistently elevated ⟨*e*⟩, and small planets are consistent with no correlation at low ⟨*e*⟩.

**Table 1. Size bins used to determine eccentricity–radius relationship**

| $R_0$ | $\sigma(R)$ | $f_{\mathrm{FWHM}}$ |
|---|---|---|
| 0.548 | 0.093 | 0.4 |
| 0.789 | 0.101 | 0.3 |
| 1.037 | 0.088 | 0.2 |
| 1.245 | 0.106 | 0.2 |
| 1.494 | 0.127 | 0.2 |
| 1.793 | 0.152 | 0.2 |
| 2.151 | 0.183 | 0.2 |
| 2.582 | 0.219 | 0.2 |
| 3.099 | 0.263 | 0.2 |
| 3.719 | 0.316 | 0.2 |
| 4.889 | 0.622 | 0.3 |
| 7.041 | 1.196 | 0.4 |
| 10.141 | 1.722 | 0.4 |
| 14.605 | 2.481 | 0.4 |

To determine the $\langle e \rangle$-$R_p$ relationship, we divided planets into fourteen size bins, where $R_0$ are the bin centers, $\sigma(R)$ are the bin widths, and $f_{\mathrm{FWHM}}$ is the fractional full-width-half-max corresponding to $\sigma(R)/R$. See Eq. **8** and corresponding text for mathematical formulation.

and roughly follows a power-law $R_{\mathrm{p,valley}} \approx 1.8\ R_\oplus\ (P/10\ \mathrm{d})^\alpha$, where $\alpha = -0.11$ (29). Thus, to investigate the eccentricities of planets in and around the radius gap, we separated planets into bins sliced parallel to the radius gap, following ref. 57. We find tentative evidence that the $\langle e \rangle$ for planets in the radius gap are elevated compared to planets on either side (Fig. 4). This feature is detected at $\sim 2\sigma$ confidence, and the physical implications are discussed below.

One complication is that planets that appear to reside in the gap may have scattered there due to $R_p/R_\star$ measurement errors. However, as discussed, $R_p/R_\star$ is covariant with $b$ and ultimately with $e$. To explore possible confounding effects, we ran a suite of injection/recovery experiments described in *SI Appendix*. We found these effects, by themselves, are insufficient to produce the observed peak at its observed significance.

**The Relationship between Eccentricity and Stellar Metallicity.** A star's metallicity is its inventory of elements heavier than hydrogen and helium. It is parameterized as $[\mathrm{Fe/H}] = \log_{10}(n_X/n_{X,\odot})$, where $n_X$ is the number density of all elements heavier than hydrogen and helium and $n_{X,\odot}$ is the solar value. Metals constitute the bulk of sub-Jovian planets by mass and form the cores of giant planets, so it is natural to explore whether eccentricity is associated with metallicity.

Fig. 5 (*Left*) shows mean eccentricity $\langle e \rangle$ as a function of stellar metallicity [Fe/H] for different planet size classes. For the three smallest size classes, we divided planets into four equal-size bins spanning $-0.4$ dex to $+0.4$ dex. Due to the low numbers of Jovians and sub-Saturns, we divided them into two bins corresponding to sub- and supersolar metallicities.

After determining $\langle e \rangle$ for each of the [Fe/H]-$R_p$ bins, we fit a linear model

$$\langle e \rangle = m \times [\mathrm{Fe/H}] + b \qquad [11]$$

to assess the strength and significance of any relationships between $\langle e \rangle$ and [Fe/H].

Eccentricity is consistently higher for large planets ($R_p > 4\ R_\oplus$) compared to small planets ($R_p < 4\ R_\oplus$) across all values of [Fe/H]. For the three small planet size classes, we do not observe any strong metallicity trends. For sub-Saturns and Jovians, however, we find a modest positive correlation between [Fe/H] and $\langle e \rangle$. The trend for sub-Saturns is significant at the $3\sigma$ level, while the trend for Jovians is consistent with a flat line within $1\sigma$.

**The Relationship between Eccentricity and Orbital Period.** Fig. 5 (*Right*) shows mean eccentricity $\langle e \rangle$ as a function of orbital period $P$ for different planet size classes. The association between $\langle e \rangle$ and $P$ varies depending on planet size. We quantified the strength of these $\langle e \rangle - P$ relationships by fitting the following model

$$\langle e \rangle = m \times \log P + b. \qquad [12]$$

For the largest planets (Jovians), we find a positive correlation between $\langle e \rangle$ and $P$, with planets inside $P < 3$ d having $\langle e \rangle \sim 0.1$, rising to $\langle e \rangle \sim 0.3$ for planets beyond $P > 10$ d. For sub-Saturns, the trend is consistent with a constant value $\langle e \rangle \sim 0.1$ across periods. For small planets ($R_p < 4\ R_\oplus$) we find no apparent correlation between period and eccentricity.

## Discussion

We detected numerous statistically significant features in the relationships between $\langle e \rangle$, $R_p$, $P$, and [Fe/H]. We discuss the astrophysical implications of these relationships below.

**Distinct Formation Pathways for Planets Larger or Smaller Than Four Earth-Radii.** There is a sharp break in the eccentricity distribution of small vs. large planets. Intriguingly, the transition radius is consistent with several other features in the demographics of close-in planets. There is a precipitous drop in exoplanet occurrence rates above $\sim 3.5\ R_\oplus$ (58). Furthermore, planets above $\sim 4\ R_\oplus$ are strongly associated with high stellar metallicity, while planets below this threshold are not (59, 60). Stated plainly, large planets are rare, but those that do exist tend to be found with elevated eccentricities $\langle e \rangle \approx 0.2$ around metal-rich stars ([Fe/H] $> 0$ dex). In contrast, small planets are common, possess low eccentricities $\langle e \rangle \approx 0.05$, and show no apparent preference for high or low metallicity hosts (Fig. 6).

The conspicuous alignment of a sharp transition in planet occurrence, eccentricity, and host star metallicity at the same planetary size threshold suggests that large planets have experienced a distinct formation pathway compared to small planets. From inspection of occurrence rate alone (Fig. 6, *Top*), it is clear that small planets form more readily than large planets. While these small planets form across a wide range of metallicities between [Fe/H] = $-0.5$ to $+0.5$, increasing metallicity does not seem to boost the number of small planets (60). In contrast, large planets are associated with high metallicity, but having high metallicity does not guarantee that a system will form a large planet. Indeed, Fischer and Valenti 61 found that only 10% of Sun-like stars with [Fe/H] = 0.25 dex have a giant planet companion inside $P \leq 4$ y. This incongruity may be understood by considering the stochastic nature of planet formation.

Planets with cores that are between 1 and 10 $M_\oplus$ and orbital periods less than a year occur at a rate of one per star (60). The fraction of stars with such cores is more uncertain but estimates range from 30 to 50% (3, 62, 63). However, only a small fraction of these cores are able accrete sufficient H/He envelopes to become super-Neptune-sized planets (with envelope mass fractions of $f_{\mathrm{env}} \gtrsim 10\%$) before the circumstellar gas disk disperses ($t_{\mathrm{disk}} \approx 1$ to 10 Myr; see, e.g., ref. 64). Increasing metallicity clearly increases this small fraction. However, making a large planet may require a confluence of other factors besides high metallicity such as a high accretion rate, long disk lifetime,

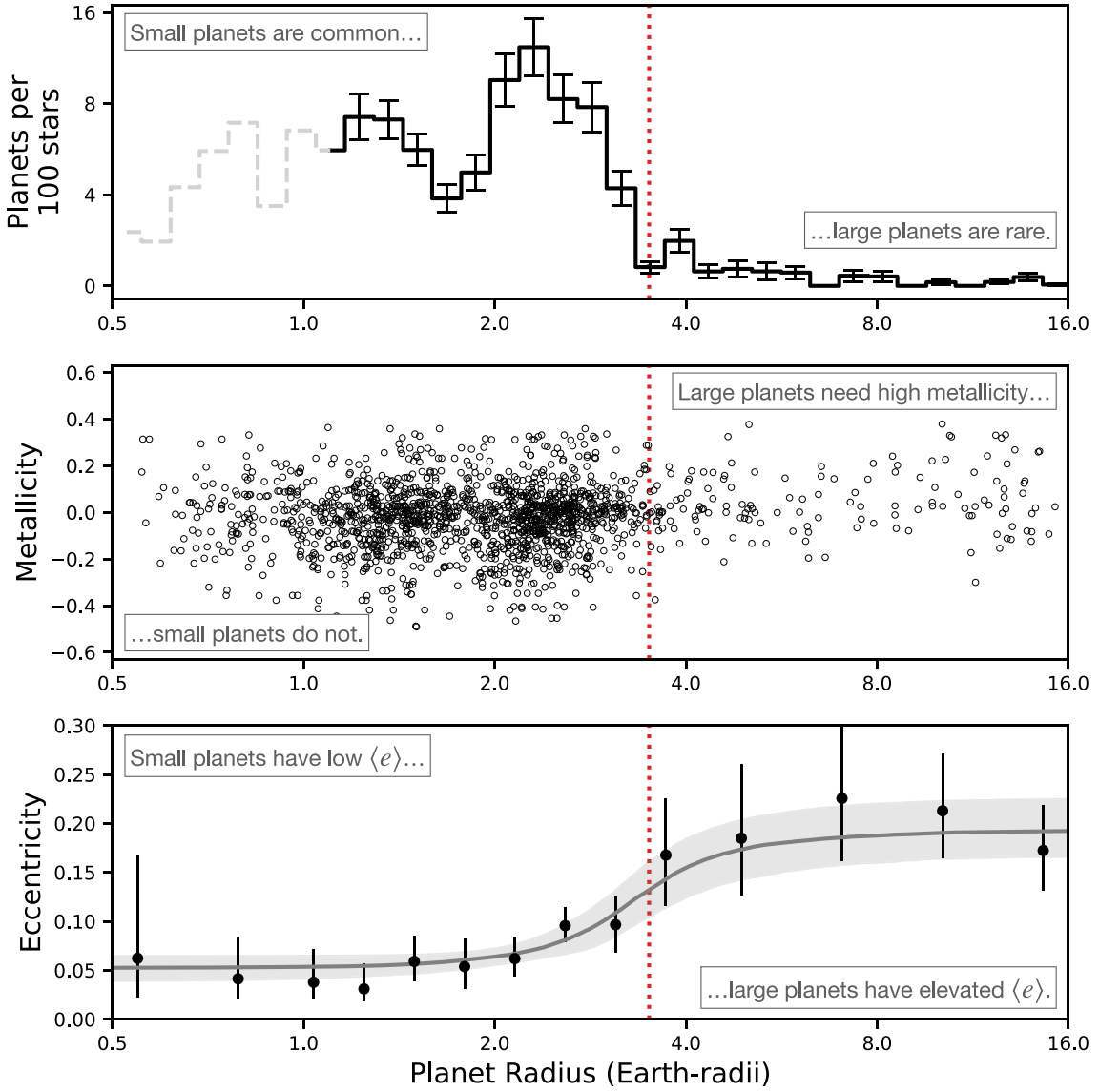


**Fig. 6.** Planets above and below a threshold of ~3.5 Earth-radii possess qualitatively distinct occurrence rates, host star metallicities, and eccentricities. *Top*: Small planets are common, but larger planets are rare. Data reproduced with permission from ref. 58. *Middle*: Small planets can be found orbiting stars of any metallicity, but large planets prefer high metallicity hosts (see ref. 59). *Bottom*: Small planets tend to have low mean eccentricities $\langle e \rangle \approx 0.05$, whereas large planets have elevated mean eccentricities $\langle e \rangle \approx 0.2$. Taken together, these trends suggest distinct formation pathways for planets smaller/larger than ~3.5 Earth-radii.

and high local gas disk mass. We do not expect a critical [Fe/H] that guarantees the formation of a large planet.

How then does eccentricity fit into this picture? Small planets almost universally have low eccentricities $e \lesssim 0.1$. These planets form within the circumstellar disk before the gas disperses, so damping will rapidly circularize any initial nonzero eccentricities. In contrast, some giant planets have elevated eccentricities $e \gtrsim 0.3$; these may result from self-excitation through secular interactions and/or from planet–planet scattering events. Other giant planets have low eccentricities $e \sim 0$; perhaps they form with low $e$ and planet–planet interactions are insufficient to excite them, or perhaps they are excited while the gas disk is sill present and there is sufficient damping to circularize. The moderate mean eccentricity observed among large planets would then be the average of low-$e$ and high-$e$ formation histories.

Because small planets form at both high and low metallicities, we should expect there to be some systems of small planets in which eccentric giant planet formation also occurred. In this case, there is a possibility that the eccentric giants would excite their smaller siblings. However, occurrence rates suggest that these outcomes are rare. Among the high metallicity stars, only 10% produce giant planets; of those, only a fraction are high-$e$; and, of those, only a fraction have orbits that can effectively transfer eccentricity to the inner system. We estimate that the fraction of small planets with distant-giant-excited eccentricities to be no more than about 5%. Even if every small planet in such systems were highly eccentric ($e \gtrsim 0.3$), the preponderance of low-$e$ small planets means that any observed $\langle e \rangle$ integrated over the population will also be low. This "eccentricity dilution" is strong enough that $\langle e \rangle$ for small planets may be kept low even around high [Fe/H] stars. Conversely, if high-$e$ small planets are identified, this might indicate the presence of a (perhaps unseen) giant outer companion.

**An Eccentricity Peak in the Radius Valley.** Planets in the radius valley exhibit somewhat elevated mean eccentricity compared to planets on either side of the valley. Although the effect only

marginally significant (2.1$\sigma$; see *SI Appendix* for calculation), several lines of evidence suggest that the peak may indeed be physical in origin, rather than a spurious fluctuation arising from the small number of planets in the radius valley.

Numerous previous analyses have endeavored to "clear out" the radius valley by improving measurements of stellar and planetary radii (e.g., refs. 29, 40, and 57), but none has produced an entirely pristine radius valley, implying that an astrophysical population of "gap planets" does indeed exist. Proposed mechanisms for producing such planets include variable mass-loss rates due to the intrinsic scatter in the XUV output of host stars, a diversity of core composition (i.e., different admixtures of iron, rock, and water), or planet–planet collisions (57). Furthermore, recent efforts have detected anomalous patterns in the size ratios (65) and resonance behaviors (65, 66) of purported gap planets, implying they may have experienced unusual dynamical histories. Any processes which would have disrupted the usual size and spacing regularity found in super-Earth and sub-Neptune systems would also be likely to have excited eccentricities. While no single line of evidence conclusively demonstrates the existence of an exotic population of gap planets, the existing measurements can be readily synthesized into a coherent picture.

If the eccentricity peak in the radius valley is indeed real, we hypothesize that the population could be generated by planet–planet collisions in postaccretion systems. Such objects could be built either from the bottom–up via mergers or from the top–down via atmospheric stripping. In the first scenario, consider two 5 $M_\oplus$ super-Earths with an Earth-like bulk composition. Such objects would have a radius of $R_p = 1.0\ R_\oplus (M_p/M_\oplus)^{0.25} = 1.5\ R_\oplus$ (67), and the merger of these two objects would result in a 1.8 $R_\oplus$planet. In the second scenario, consider a sub-Neptune with a 10$M_\oplus$ Earth composition core and a 3% by mass H/He envelope. This object would have $R_p \approx 3.6\ R_\oplus$; the large change in $R_p$ occurs because a small fractional mass of H/He can greatly increase a planet's radius (e.g., ref. 68). An impacting body with mass $M = 0.1 M_\oplus$ would have sufficient kinetic energy to unbind the atmosphere, leaving a bare rock with $R_p = 1.8\ R_\oplus$. In either scenario, at $P = 20$ d around a Sun-like star, the final planet would fall squarely in the middle of the radius valley and would possess a nonexistent or tenuous atmospheres and elevated eccentricity. These predictions can be tested with a focused Doppler campaign using existing extreme precision radial velocity instruments or with secondary eclipse measurements using JWST.

**A Common Parent Population for Single- and Multitransiting Systems.** Early analyses of the Kepler census found an overabundance of single-transiting systems compared to models developed from the statistics of multitransiting systems (69). The origin of this so-called "Kepler dichotomy" has remained a mystery, with both one- and two- population models being invoked to explain the data (e.g., refs. 62 and 69–72. While it is possible that an intrinsic population of singles exists, the dichotomy may instead result from observational biases. One plausible explanation is that the single-transiting systems belong to a dynamically hot population with elevated eccentricities and mutual inclinations, which produces fewer observations of multitransiting systems even if intrinsic multiplicity is high (e.g., ref. 69). Alternatively, Zink et al. (73) explained the Kepler dichotomy as a byproduct of suppressed transit detection efficiency in multiplanet systems.

The similarity of the $\langle e \rangle$-$R_p$ curves for singles and multis (Fig. 3) points to a common parent population and formation channel. If the single transiting systems represent some distinct population of planets with a unique formation history, we would expect a different shape to the $\langle e \rangle$-$R_p$ relationship. We therefore interpret Kepler's single-transiting systems as the high-$e$ high-$\Delta i$ tail of the intrinsic multiplanet population, rather than a qualitatively distinct group.

$N$-body models of planetesimal growth have found a tendency toward equipartition of random orbital velocities which produces a tight correlation between eccentricity and inclination $\langle e \rangle \simeq 2\Delta i$ (74). This tendency has been observed over a wide range of size and angular momentum scales that includes the Kepler population (13). Thus, eccentricity (which has a weak effect on observed multiplicity) can be used as a proxy for mutual inclination (which has a strong effect on observed multiplicity). Our observed eccentricity enhancement factor $e_{\rm singles}/e_{\rm multis} \sim 2.5$ therefore implies a similar disparity in mutual inclinations which biases the number of detected planets in dynamically hot vs. dynamically cool systems.

## Conclusions

In this work, we measured the distribution of eccentricities for close-in planets based on 1,646 individual planets discovered by Kepler. We summarize our main findings below:

1. The eccentricity distribution $f(e)$ peaks at $e = 0$ and falls toward zero at $e = 1$. The shape—but not the spread—is consistent across a range of planet sizes. Of the commonly used parametric models in the literature, the beta distribution provides a good approximation, but the Rayleigh and half-normal distributions do not.
2. On average, large planets are ~3 to 4× more eccentric than small planets. The transition between low-to-high $\langle e \rangle$ at $R_p \sim 3.5\ R_\oplus$ matches other features in the exoplanet population, specifically transitions in planet occurrence rate and host star metallicity. Taken together, these trends suggest that large and small planets form via two distinct formation channels.
3. Planets in the radius valley show tentative evidence of having elevated eccentricities compared to slightly larger or smaller planets. Such planets may have experienced major planet–planet collisions in their past which substantially altered their core mass and envelope fraction, moving them into the sparsely populated radius valley.
4. Planets in single- vs. multitransiting systems have the same $\langle e \rangle$-$R_p$ relation, except that, on average, singles are ~2.5× more eccentric than multis. This correspondence suggests that single-transiting and multitransiting systems belong to the same parent population and experienced similar formation histories.
5. Eccentricity and metallicity lack a strong correlation (positive or negative). However, subtle correlations among subpopulations may exist and warrant further investigation.

Planet eccentricities are related to other planetary properties in intricate, subtle, and surprising ways. Surely, fundamental aspects of planet formation physics are encoded in these relationships. We eagerly await new models that can account for these patterns.

**Data, Materials, and Software Availability.** Kepler Object of Interest Q1-Q17 light curves (22) were accessed via the NASA Exoplanet Archive (31) on 1 October 2024. Light curves are hosted at the Mikulski Archive for Space Telescopes: https://archive.stsci.edu. Stellar properties from the Gaia-Kepler Stellar Properties catalog are described in Berger et al. 2020 (26).

**ACKNOWLEDGMENTS.** We acknowledge the heroic efforts by the Kepler and Gaia teams, which provided the starting datasets for this project. We benefited from insightful discussions with Sarah Ballard, Konstantin Batygin, Brendan

Bowler, Dan Foreman-Mackey, Mason MacDougall, James Rogers, Jason Rowe, Sheila Sagear, Hilke Schlichting, Vincent Van Eylen, Josh Winn, and Jon Zink. Funding for this work was provided by a University of California, Los Angeles set-up award to E.A.P. and by the Heising-Simons Foundation Award #2022-3833. We thank the anonymous referees for suggestions which greatly improved the quality of the analysis.

1. P. S. Laplace, *Exposition Du Système Du Monde* (Paris, Courcier, 1796).
2. K. Tsiganis, R. Gomes, A. Morbidelli, H. F. Levison, Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435**, 459–461 (2005).
3. E. A. Petigura, A. W. Howard, G. W. Marcy, Prevalence of Earth-size planets orbiting Sun-like stars. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19273–19278 (2013).
4. A. W. Howard *et al.*, Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. *ApJS* **201**, 15 (2012).
5. B. J. Fulton *et al.*, The California–Kepler survey. III. A gap in the radius distribution of small planets. *AJ* **154**, 109 (2017).
6. S. J. O'Toole *et al.*, Selection functions in doppler planet searches. *MNRAS* **392**, 641–654 (2009).
7. F. A. Rasio, E. B. Ford, Dynamical instabilities and the formation of extrasolar planetary systems. *Science* **274**, 954–956 (1996).
8. B. Jackson, R. Greenberg, R. Barnes, Tidal evolution of close-in extrasolar planets. *ApJ* **678**, 1396–1406 (2008).
9. R. I. Dawson, J. A. Johnson, Origins of hot Jupiters. *ARAA* **56**, 175–221 (2018).
10. D. M. Kipping, Parametrizing the exoplanet eccentricity distribution with the beta distribution. *MNRAS* **434**, L51–L55 (2013).
11. V. Van Eylen *et al.*, The orbital eccentricity of small planet systems. *AJ* **157**, 61 (2019).
12. D. C. Fabrycky *et al.*, Architecture of Kepler's multi-transiting systems. II. New investigations with twice as many candidates. *ApJ* **790**, 146 (2014).
13. J. W. Xie *et al.*, Exoplanet orbital eccentricities derived from LAMOST-Kepler analysis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11431–11435 (2016).
14. S. M. Mills *et al.*, The California-Kepler survey. VIII. Eccentricities of Kepler planets and tentative evidence of a high-metallicity preference for small eccentric planets. *AJ* **157**, 198 (2019).
15. V. Van Eylen, S. Albrecht, Eccentricity from transit photometry: Small planets in Kepler multi-planet systems have low eccentricities. *ApJ* **808**, 126 (2015).
16. S. Sagear, S. Ballard, The orbital eccentricity distribution of planets orbiting M dwarfs. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2217398120 (2023).
17. J. B. Pollack *et al.*, Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
18. E. B. Ford, F. A. Rasio, Origins of eccentric extrasolar planets: Testing the planet-planet scattering model. *ApJ* **686**, 621–636 (2008).
19. D. Fabrycky, S. Tremaine, Shrinking binary and planetary orbits by Kozai cycles with tidal friction. *ApJ* **669**, 1298–1315 (2007).
20. J. Li, D. Lai, K. R. Anderson, B. Pu, Giant planet scatterings and collisions: Hydrodynamics, merger-ejection branching ratio, and properties of the remnants. *MNRAS* **501**, 1621–1632 (2021).
21. P. Goldreich, R. Sari, Eccentricity evolution for planets in gaseous disks. *ApJ* **585**, 1024–1037 (2003).
22. S. E. Thompson *et al.*, Planetary candidates observed by Kepler. VIII. A fully automated catalog with measured completeness and reliability based on data release 25. *ApJS* **235**, 38 (2018).
23. J. L. Christiansen *et al.*, Measuring transit signal recovery in the Kepler pipeline. IV. Completeness of the DR25 planet candidate catalog. *AJ* **160**, 159 (2020).
24. S. Bryson *et al.*, The occurrence of rocky habitable-zone planets around solar-like stars from Kepler data. *AJ* **161**, 36 (2021).
25. T. A. Berger, D. Huber, E. Gaidos, J. L. van Saders, Revised radii of Kepler stars and planets using Gaia data release 2. *ApJ* **866**, 99 (2018).
26. T. A. Berger *et al.*, The Gaia-Kepler stellar properties catalog. I. Homogeneous fundamental properties for 186,301 Kepler stars. *AJ* **159**, 280 (2020).
27. E. Furlan *et al.*, The Kepler follow-up observation program. I. A catalog of companions to Kepler stars from high-resolution imaging. *AJ* **153**, 71 (2017).
28. M. L. Wood, A. W. Mann, A. L. Kraus, Characterizing undetected stellar companions with combined data sets. *AJ* **162**, 128 (2021).
29. E. A. Petigura *et al.*, The California-Kepler survey. X. The radius gap as a function of stellar mass, metallicity, and age. *AJ* **163**, 179 (2022).
30. M. C. Stumpe *et al.*, Kepler presearch data conditioning. I–Architecture and algorithms for error correction in Kepler light curves. *PASP* **124**, 985 (2012).
31. R. L. Akeson *et al.*, The NASA Exoplanet Archive: Data and tools for exoplanet research. Kepler Objects of Interest Properties. https://exoplanetarchive.ipac.caltech.edu/. Accessed 1 October 2024.
32. T. Holczer *et al.*, Transit timing observations from Kepler. IX. Catalog of the full long-cadence data set. *ApJS* **225**, 9 (2016).
33. D. M. Kipping, Characterizing distant worlds with asterodensity profiling. *MNRAS* **440**, 2164–2184 (2014).
34. Gaia Collaboration *et al.*, Gaia Data Release 2. Summary of the contents and survey properties. *A&A* **616**, A1 (2018).
35. T. O. Husser *et al.*, A new extensive library of PHOENIX stellar atmospheres. *A&A* **553**, A6 (2013).
36. H. Parviainen, S. Aigrain, LDTk: Limb darkening toolkit. *MNRAS* **453**, 3821–3826 (2015).
37. H. Akaike, A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
38. D. M. Kipping, Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *MNRAS* **435**, 2152–2160 (2013).
39. G. J. Gilbert, Accurate modeling of grazing transits using umbrella sampling. *Astron. J.* **163**, 111 (2022).
40. E. A. Petigura, Two views of the radius gap and the role of light curve fitting. *AJ* **160**, 89 (2020).
41. G. J. Gilbert, M. G. MacDougall, E. A. Petigura, Implicit biases in transit models using stellar pseudo density. *AJ* **164**, 92 (2022).
42. J. Skilling, "Nested sampling" in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, American Institute of Physics Conference Series*, R. Fischer, R. Preuss, U. V. Toussaint, Eds. (AIP, 2004), vol. 735, pp. 395–405.
43. E. Higson, W. Handley, M. Hobson, A. Lasenby, Dynamic nested sampling: An improved algorithm for parameter estimation and evidence calculation. *Stat. Comput.* **29**, 891–913 (2019).
44. J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS* **493**, 3132–3158 (2020).
45. M. G. MacDougall, G. J. Gilbert, E. A. Petigura, Accurate and efficient photoeccentric transit modeling. *AJ* **166**, 61 (2023).
46. E. B. Ford, S. N. Quinn, D. Veras, Characterizing the orbital eccentricities of transiting extrasolar planets with photometric observations. *ApJ* **678**, 1407–1418 (2008).
47. R. I. Dawson, J. A. Johnson, The photoeccentric effect and proto-hot Jupiters. I. Measuring photometric eccentricities of individual transiting planets. *ApJ* **756**, 122 (2012).
48. J. N. Winn, "Exoplanet transits and occultations" in *Exoplanets*, S. Seager, Ed. (University of Arizona Press, Tucson, AZ, 2010), pp. 55–77.
49. D. W. Hogg, A. D. Myers, J. Bovy, Inferring the eccentricity distribution. *ApJ* **725**, 2166–2175 (2010).
50. B. P. Bowler, S. C. Blunt, E. L. Nielsen, Population-level eccentricity distributions of imaged exoplanets and brown dwarf companions: Dynamical evidence for distinct formation channels. *AJ* **159**, 63 (2020).
51. J. W. Barnes, Effects of orbital eccentricity on extrasolar planet transit detectability and light curves. *PASP* **119**, 986–993 (2007).
52. C. J. Burke, Impact of orbital eccentricity on the detection of transiting extrasolar planets. *ApJ* **679**, 1566–1573 (2008).
53. D. M. Kipping, Bayesian priors for the eccentricity of transiting planets. *MNRAS* **444**, 2263–2269 (2014).
54. D. Foreman-Mackey, D. W. Hogg, T. D. Morton, Exoplanet population inference and the abundance of earth analogs from noisy, incomplete catalogs. *ApJ* **795**, 64 (2014).
55. K. Masuda, E. A. Petigura, O. J. Hall, Inferring the rotation period distribution of stars from their projected rotation velocities and radii: Application to late-F/early-G Kepler stars. *MNRAS* **510**, 5623–5638 (2022).
56. M. A. Newton, A. E. Raftery, Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat.* **56**, 3–26 (2018).
57. C. S. K. Ho, V. Van Eylen, A deep radius valley revealed by Kepler short cadence observations. *MNRAS* **519**, 4056–4073 (2023).
58. B. J. Fulton, E. A. Petigura, The California-Kepler survey. VII. Precise planet radii leveraging Gaia DR2 reveal the stellar mass dependence of the planet radius gap. *AJ* **156**, 264 (2018).
59. L. A. Buchhave *et al.*, An abundance of small exoplanets around stars with a wide range of metallicities. *Nature* **486**, 375–377 (2012).
60. E. A. Petigura *et al.*, The California-Kepler survey. IV. Metal-rich stars host a greater diversity of planets. *AJ* **155**, 89 (2018).
61. D. A. Fischer, J. Valenti, The planet-metallicity correlation. *ApJ* **622**, 1102–1117 (2005).
62. W. Zhu, C. Petrovich, Y. Wu, S. Dong, J. Xie, About 30% of sun-like stars have Kepler-like planetary systems: A study of their intrinsic architecture. *ApJ* **860**, 101 (2018).
63. D. C. Hsu, E. B. Ford, D. Ragozzine, K. Ashby, Occurrence rates of planets orbiting FGK stars: Combining Kepler DR25, Gaia DR2, and Bayesian inference. *AJ* **158**, 109 (2019).
64. J. P. Williams, L. A. Cieza, Protoplanetary disks and their evolution. *ARAA* **49**, 67–117 (2011).
65. Q. Chance, S. Ballard, Evidence that planets in the radius gap do not resemble their neighbors. arXiv [Preprint] (2024). https://arxiv.org/abs/2410.02150 (Accessed 2 October 2024).
66. F. Dai *et al.*, The prevalence of resonance among young, close-in planets. *Astron. J.* **168**, 239 (2024).
67. J. E. Owen, Y. Wu, The evaporation valley in the Kepler planets. *ApJ* **847**, 29 (2017).
68. E. D. Lopez, J. J. Fortney, Understanding the mass-radius relation for sub-Neptunes: Radius as a proxy for composition. *ApJ* **792**, 1 (2014).
69. J. J. Lissauer *et al.*, Architecture and dynamics of Kepler's candidate multiple transiting planet systems. *ApJS* **197**, 8 (2011).
70. S. Ballard, J. A. Johnson, The Kepler dichotomy among the M dwarfs: Half of systems contain five or more coplanar planets. *ApJ* **816**, 66 (2016).
71. M. Y. He, E. B. Ford, D. Ragozzine, D. Carrera, Architectures of exoplanetary systems. III. Eccentricity and mutual inclination distributions of AMD-stable planetary systems. *AJ* **160**, 276 (2020).
72. S. C. Millholland *et al.*, Evidence for a nondichotomous solution to the Kepler dichotomy: Mutual inclinations of Kepler planetary systems from transit duration variations. *AJ* **162**, 166 (2021).
73. J. K. Zink, J. L. Christiansen, B. M. S. Hansen, Accounting for incompleteness due to transit multiplicity in Kepler planet occurrence rates. *MNRAS* **483**, 4479–4494 (2019).
74. E Kokubo, "Dynamics of planetesimals: The role of two-body relaxation" in *IAU Colloq. 197: Dynamics of Populations of Planetary Systems*, Z. Knežević, A. Milani, Eds. (2005), pp. 41–46.
75. Astropy Collaboration *et al.*, The Astropy Project: Building an open-science project and status of the v2.0 core package. *AJ* **156**, 123 (2018).
76. L. Kreidberg, batman: BAsic Transit Model cAlculatioN in Python. *PASP* **127**, 1161 (2015).
77. D. Foreman-Mackey, E. Agol, S. Ambikasaran, R. Angus, Fast and scalable Gaussian process modeling with applications to astronomical time series. *AJ* **154**, 220 (2017).
78. D. Foreman-Mackey *et al.*, exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series. *J. Open Source Software* **6**, 3285 (2021).
79. C. R. Harris *et al.*, Array programming with NumPy. *Nature* **585**, 357–362 (2020).
80. W. McKinney, "Data structures for statistical computing in Python" in *Proceedings of the 9th Python in Science Conference*, S. van der Walt, J. Millman, Eds. (2010), pp. 51–56.
81. J. Salvatier, T. V. Wiecki, C. Fonnesbeck, Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).
82. P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
83. R. Luger *et al.*, starry: Analytic occultation light curves. *AJ* **157**, 64 (2019).