Two families of exocomets in the Beta Pictoris system

F. Kiefer[1,2], A. Lecavelier des Etang[1,2], J. Boissier[3], A. Vidal-Madjar[1,2], H. Beust[5], A.-M. Lagrange[5], G. Hébrard[1,2], R. Ferlet[1]

[1] CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98bis bld Arago, F-75014 Paris, France

[2] UPMC, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

[3] IRAM, 300 rue de la piscine, 38406 Saint Martin d'Heres, France

[4] UJF-Grenoble 1/CNRS-INSU, IPAG, UMR 5274, Grenoble, France

The young planetary system surrounding the star β Pictoris harbors active minor bodies [1–6]. These asteroids and comets produce a large amount of dust and gas through collisions and evaporation, as in the early ages of the Solar System [7]. Spectroscopic observations of β Pictoris reveal a high rate of transits of small evaporating bodies [8–11], *i.e.* exocomets. Here we report the analysis of over a thousand spectra gathered between 2003 and 2011, which provides a sample of about 6000 variable absorption signatures due to transiting exocomets. Statistical analysis of the observed properties of these exocomets allow us to identify two populations with highly different physical properties. A first family consists of exocomets producing shallow absorption lines, which can be attributed to old exhausted comets trapped in a mean motion resonance with a massive planet; the second family consists of exocomets producing deep absorption lines, which are possibly related to the recent fragmentation of one or a few parent bodies. Our results show that the evaporating bodies observed for decades in the β Pic system are analogous to the comets in our own Solar System.

From 2003 to 2011, a total of 1106 spectra of β Pictoris have been obtained using the *HARPS* spectrograph. Observations of the Ca II doublet (K-3933.66Å; H-3968.47Å) show a large number of variable absorption features (Fig. 1) varying on timescales of one to six hours. These features simultaneously detected in both Ca II K and Ca II H lines are interpreted as exocomets transiting in front of the stellar disk [7–11]. Since the β Pic Ca II spectrum is typically observed to be stable on 30 minutes timescales, we averaged together spectra in distinct 10 minutes time intervals to limit any possible spectral variability. This results in a total of 357 spectra with signal-to-noise ratio greater than 80. In order to characterize the profile of these transient absorption lines, we divided each of the 357 averaged spectra by a reference spectrum of β Pictoris (Extended Data Fig. 1 and 2) assumed to be free of transiting exocomet's absorption signatures.

Given the *HARPS* resolution and sensitivity, each β Pictoris spectrum shows an average of about 6 variable absorption features due to exocomets. These features have radial velocities ranging from -150 to +200 km s^{-1} with respect to

the β Pictoris heliocentric radial velocity ($\sim 20\,\mathrm{km\,s^{-1}}$). We fitted each feature with a Gaussian profile and obtained the estimates of $p_{K,H}$ their depths in the Ca II K and H lines respectively, v_r the radial velocity of the absorbing cloud, and Δv the line full width at half maximum (FWHM) expressed in units of radial velocity.

A large fraction of the cometary gaseous clouds which pass in front of the star and produce the absorption features, are smaller than the stellar disk. Therefore, the depth $p_{K,H}$ of each feature depends on A, the $\mathrm{Ca^+}$ cloud's opacity (absorption depth, hereafter), and $\alpha = \Sigma_c/\Sigma_\star$ the ratio of the area of the cloud Σ_c over the area of the stellar disk $\Sigma_\star$. The simultaneous fit of the K and H lines yields a non-degenerate determination of both α, A, v_r and Δv for each feature (Extended Data Fig. 3).

Because the transit of an orbiting exocomet can last several hours, we considered the measurements derived from the fit of only one spectrum per observation day. This ensured that each set of measurements was linked to a different independent object. We thus collected a total of 570 individual set of measurements from independent transiting cometary clouds. Among these 570 detections, we discarded variable absorption features compatible with $p_{K,H} < 0.01$ and $\Delta v < 3\,\mathrm{km\,s^{-1}}$, in order to avoid contamination introduced by fitting spurious features. We thus end up with a total of 493 detected cometary clouds. For the statistical analysis, we also excluded detections with $\alpha = 1$, corresponding to clouds covering the full stellar disk and for which some physical characteristics, like the transit distance, cannot be derived.

The plot of the absorption depth as a function of the surface ratio (Fig. 2) shows a depletion of cometary clouds with $0.2 < \alpha < 0.5$ and $A \geq 3$ (or $\log A \geq 0.5$). This depletion divides the data into two well separated clusters, revealing the existence of two distinct populations of exocomets. Statistical cluster analysis[12] in the (p_H, p_K) diagram (Extended Data Fig. 4) shows that these two populations can be distinguished by the value of p_K. A first population, so-called 'population S', corresponding mainly to clouds with small surface ratio ($\alpha \sim 0.1$), produces shallow absorption lines ($p_K < 0.4$); while a second population, corresponding to clouds with large surface ratio ($\alpha \sim 0.8$), so-called 'population D', produces deep absorption lines ($p_K > 0.4$). These two populations present highly different physical properties (Fig. 3).

First, they have different radial velocity and FWHM distributions[13]. Exocomets of population S have a broad distribution of radial velocities with $v_r|_S \sim 36 \pm 55\,\mathrm{km\,s^{-1}}$, while the population D exocomets have a narrow distribution with $v_r|_D \sim 15 \pm 6\,\mathrm{km\,s^{-1}}$. Moreover, population S has a broad distribution of FWHM with $\Delta v_S \sim 55 \pm 55\,\mathrm{km\,s^{-1}}$, while population D has a peaked distribution with $\Delta v \sim 7 \pm 3\,\mathrm{km\,s^{-1}}$. Since the width of the absorption line is expected to decrease as the distance between the exocomet and host star increases[14], the bi-modal FWHM distribution indicates that population S exocomets transit at shorter distances than the exocomets of population D. Furthermore, the narrow distribution of radial velocities for population D suggests that these exocomets are gathered on neighbouring orbits with similar longitude of periastron relative to the line of sight[9], in contrast with exocomets of population S which seem to

be scattered on a wide variety of orbits.

Second, the observed bimodal distribution of cloud's properties can not only be explained by different orbital characteristics between the two populations. As the spatial extent of the cometary gaseous cloud and the calcium production rate of a comet both depend on the distance to the star[14,15], the resulting Ca II absorption depth also depends on the distance. Thus, if exocomets of population D had the same intrinsic properties as exocomets of population S and were only transiting at farther distances, then one would expect to measure significantly smaller Ca II absorption depths, as *not* observed in Fig. 3c. Moreover, if the observed exocomets were originating from a single family spread over a wide range of orbital distances, we would expect a continuum of measurements in any of the quantities presented in Fig. 3. On the contrary, all histograms show a dichotomy in these measurements, which confirms the existence of two families of exocomets orbiting β Pictoris.

The efficiency of converting stellar irradiance incident on a comet into the evaporation of gas (mainly water) and dust from its core depends on the surface properties and size of its nucleus. In support of the above interpretations, estimates of this *evaporation efficiency* show that population D exocomets would have about ten times higher dust production rate if they were located at the same distance to the star as population S exocomets (Extended Data Fig. 5).

Coupling the evaporation efficiency with a dynamical model of evaporation[14–16] allows the distance to the star at the time of transit (d) and the dust production rate ($\dot{M}$) of each detected exocomet to be derived. We find that population D comets orbit twice as far away from β Pic as the population S comets, with $d_D \sim 19 \pm 4 R_\star$ and $d_S \sim 10 \pm 3 R_\star$, as expected from their lower FWHM (Extended Data Fig. 6); while the dust production rates follow $\dot{M}_D / \dot{M}_S \sim 2$. These results show that exocomets of population D present more active surfaces than exocomets of population S; this could be explained either by the nuclei being larger in size or by the nuclei being disrupted, thus exposing fresh layers of ice buried in their core.

Given that the radial velocities are directly measured from the spectra, and assuming that exocomets have near parabolic orbits, the combination of the distance and the radial velocity yields Q, the periastron distance, and ϖ, the longitude of the periastron. The plot of Q as a function of ϖ shows that orbital properties are also different between the two families (Fig. 4). Exocomets of population D have larger periastron distances than exocomets of population S, with $Q_D \sim 18 \pm 4 R_\star$ and $Q_S \sim 9 \pm 3 R_\star$. They also present a narrow distribution of longitude of periastron, indicating that all population D exocomets share similar orbits, with $\varpi_D \sim 7 \pm 8°$. This concentration of a large number of bodies on similar orbits with a nearly constant longitude of periastron can be explained by the disruption of one or a few individual exocomets. These observations resemble the Kreutz Family comets in our own Solar System[17], which are detected with periastron distances ranging from 0.005 AU to 0.01 AU and periastron longitude ranging from $10°$ to $90°$.

Conversely, the population S follows as expected a much broader distribution of longitude of periastron, with

$\varpi_S \sim 22 \pm 25°$. The arc-like structure in Fig. 4 suggests that a fraction of the exocomets of population S present a strong correlation between the periastron distances and the longitudes: starting at $\log Q \sim 1.0$ and $\varpi \sim 50^o$, the periastron distances decrease to $\log Q \sim 0.4$ and $\varpi \sim 100^o$. This is exactly the behavior predicted[18] for comets trapped in a mean motion resonance with a massive planet (see Fig. 11 of ref. 18) such as β Pic b[19,20]. In this scenario, the lower evaporation rate of population S exocomets is explained by the exhaustion of volatiles at the surface of their nucleus caused by a large number of periastron passages as they evolve toward highly eccentric orbits[21,22].

References

1. Roberge, A., et al. High-Resolution Hubble Space Telescope STIS Spectra of C I and CO in theβ Pictoris Circumstellar Disk. *The Astrophysical Journal* **538**, 904-910 (2000).

2. Lecavelier des Etangs, A., et al. Deficiency of molecular hydrogen in the disk of β Pictoris. *Nature* **412**, 706-708 (2001).

3. Okamoto, Y. K., et al. An early extrasolar planetary system revealed by planetesimal belts in β Pictoris. *Nature* **431**, 660-663 (2004).

4. Telesco, C. M., et al. Mid-infrared images of β Pictoris and the possible role of planetesimal collisions in the central disk. *Nature* **433**, 133-136 (2005).

5. Roberge, A., Feldman, P. D., Weinberger, A. J., Deleuil, M., & Bouret, J.-C. Stabilization of the disk around β Pictoris by extremely carbon-rich gas. *Nature* **441**, 724-726 (2006).

6. de Vries, B. L., et al. Comet-like mineralogy of olivine crystals in an extrasolar proto-Kuiper belt. *Nature* **490**, 74-76 (2012).

7. Vidal-Madjar, A., Lecavelier des Etangs, A., & Ferlet, R. β Pictoris, a young planetary system? A review. *Planetary and Space Science* **46**, 629-648 (1998).

8. Ferlet, R., Vidal-Madjar, A., & Hobbs, L. M. The Beta Pictoris circumstellar disk. V - Time variations of the CA II-K line. *Astronomy and Astrophysics* **185**, 267-270 (1987).

9. Beust, H., Vidal-Madjar, A., Ferlet, R., & Lagrange-Henri, A. M. The Beta Pictoris circumstellar disk. X - Numerical simulations of infalling evaporating bodies. *Astronomy and Astrophysics* **236**, 202-216 (1990).

10. Beust, H., Vidal-Madjar, A., Ferlet, R., & Lagrange-Henri, A. M. The Beta Pictoris circumstellar disk. XI - New CA II absorption features reproduced numerically. *Astronomy and Astrophysics* **241**, 488-492 (1991).

11. Vidal-Madjar, A., et al. HST-GHRS observations of β Pictoris: additional evidence for infalling comets. *Astronomy and Astrophysics* **290**, 245-258 (1994).

12. Everitt, B.S., Cluster Analysis, *New York: Halsted Press* (1993)

13. Lagrange, A.-M., et al. The β Pictoris circumstellar disk. XXI. Results from the December 1992 spectroscopic campaign.. *Astronomy and Astrophysics* **310**, 547-563 (1996).

14. Beust, H., Lagrange, A.-M., Plazy, F., & Mouillet, D. The β Pictoris circumstellar disk. XXII. Investigating the model of multiple cometary infalls.. *Astronomy and Astrophysics* **310**, 181-198 (1996).

15. Beust, H., & Tagger, M. A hydrodynamical model for infalling evaporating bodies in the Beta Pictoris circumstellar disk. *Icarus* **106**, 42-(1993).

16. Huebner, W. F. Physics and chemistry of comets. *Physics and Chemistry of Comets* , (1990).

17. Marsden, B. G., The sungrazing comet group, The Astronomical Journal, 72, 1170 (1967)

18. Beust, H., & Morbidelli, A. Mean-Motion Resonances as a Source for Infalling Comets toward β Pictoris. *Icarus* **120**, 358-370 (1996).

19. Lagrange, A.-M., et al. A probable giant planet imaged in the β Pictoris disk. VLT/NaCo deep L'-band imaging. *Astronomy and Astrophysics* **493**, L21-L25 (2009).

20. Lecavelier des Etangs, A., & Vidal-Madjar, A. Is β Pictoris b the transiting planet of November 1981?. *Astronomy and Astrophysics* **497**, 557-562 (2009).

21. Thébault, P., & Beust, H. Falling evaporating bodies in the β Pictoris system. Resonance refilling and long term duration of the phenomenon. *Astronomy and Astrophysics* **376**, 621-640 (2001).

22. Beust, H., & Morbidelli, A. Falling Evaporating Bodies as a Clue to Outline the Structure of the β Pictoris Young Planetary System. *Icarus* **143**, 170-188 (2000).

End Notes

Author Contributions   F.K. led the data analysis, with contributions from A.L.d.E., J.B., A.V.-M., R.F. and G.H; F.K. and A.L.d.E. wrote the paper; H.B computed mean-motion resonance curves and developed theoretical modeling; A.L.d.E, A.V.-M., A.-M.L. and H.B. impulsed and contributed to the conception of the project. All authors discussed the results and commented on the manuscript.

Author information   The data used for this study come from the ESO database and are referenced with the following IDs: 060.A-9036(A), 072.C-0636(A), 072.C-0636(B), 075.C-0234(A), 075.C-0234(B), 076.C-0073(A), 076.C-0279(A), 078.C-0209(A), 078.C-0209(B), 080.C-0032(A), 080.C-0032(B), 080.C-0664(A), 080.C-0712(A), 081.C-0034(B), 082.C-0308(A), 082.C-0412(A), 084.C-1039(A), 184.C-0815(C), 184.C-0815(F). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.K. (kiefer@iap.fr) or A.L.d.E (lecaveli@iap.fr).
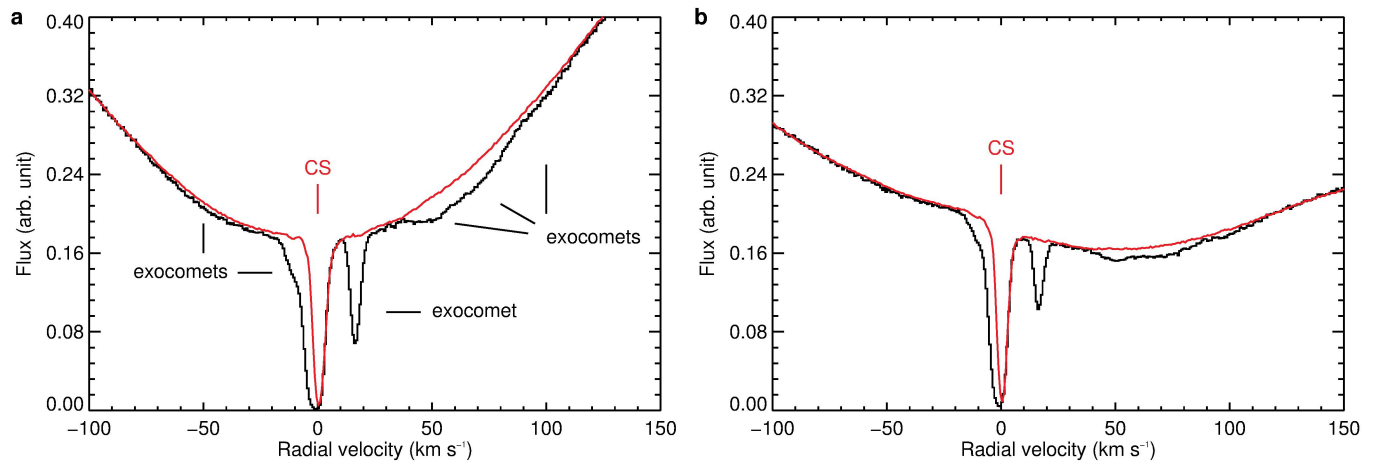
Figures captions

Figure 1: **A typical Ca II spectrum of β Pictoris**. **a,** Ca II K-line (3934Å). **b,** Ca II H-line (3968Å). A typical Ca II spectrum of β Pic (black line) collected on 27 October 2009 is shown together with the derived β Pic stellar spectrum (red line) used as the reference spectrum free of variable absorption features. Radial velocities are given with respect to the star's rest frame. CS indicates the circumstellar disk contribution, whilst solid black lines indicate the changes in flux caused by the transiting exocomets. Each transiting exocomet produces an absorption signature detected at the same radial velocity in both Ca II lines.
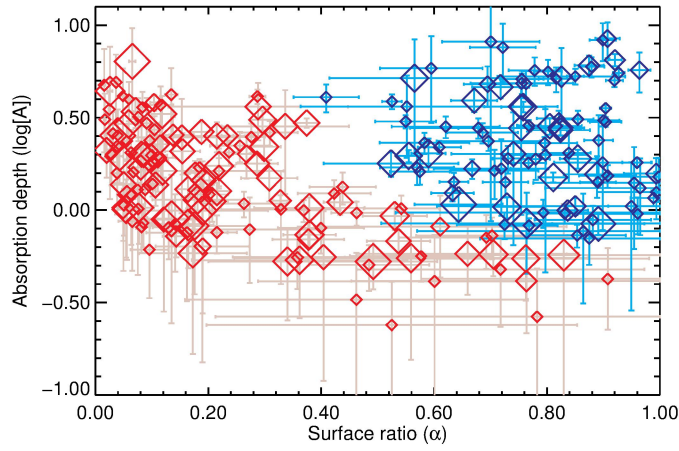
Figure 2: **Coma absorption depths as a function of surface ratio for transiting exocomets.** The absorption depths A are shown on a logarithmic scale as a function of α, representing the cloud sizes in units of the stellar disk's area. Small size symbols correspond to data taken in 2003 and large size symbols to 2011 data. Error bars represent the standard deviation. The exocomets producing shallow absorption lines ($p_K < 0.4$), so-called population S, are plotted in red, whilst exocomets producing deep absorption lines ($p_K > 0.4$), so-called population D, are plotted in blue. The cloud sizes show a bimodal distribution with a deficiency of exocomets with high absorption depths at intermediate sizes.
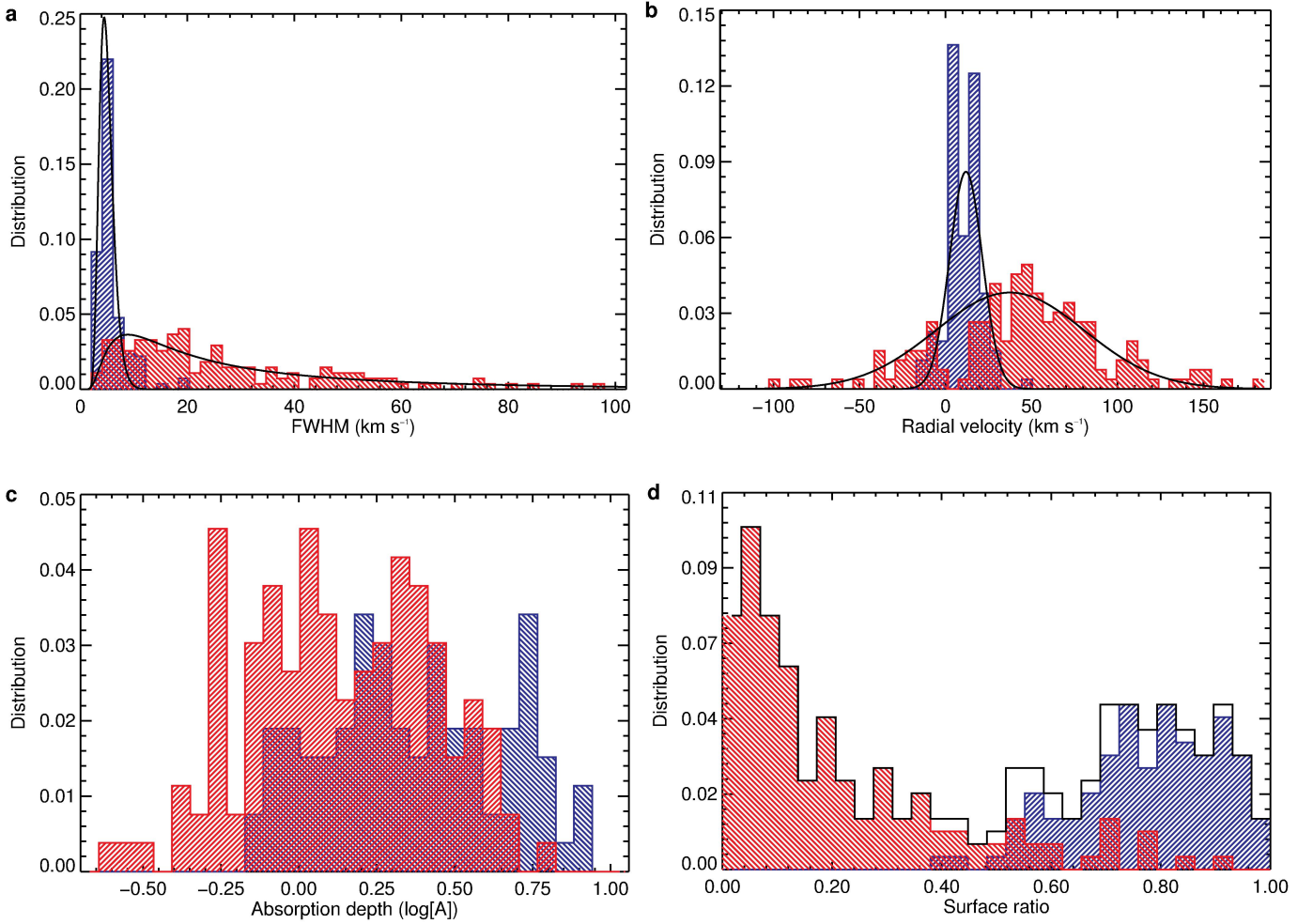
Figure 3: **Measured distribution of the physical properties of transiting exocomets.** Histograms showing the distribution of, **a,** FWHM; **b,** radial velocity; **c,** absorption depth; and **d,** surface ratio α. The S population exocomets (in red) spread on a wide range of FWHM and radial velocities, while population D exocomets (in blue) produce narrow lines redshifted at low radial velocities. The S population is characterized by small cloud sizes, while the D population is composed of large clouds. The cloud sizes for the whole sample of observed exocomets (black line) shows a clear bimodal distribution with a depletion of exocomets at intermediate sizes (panel d).
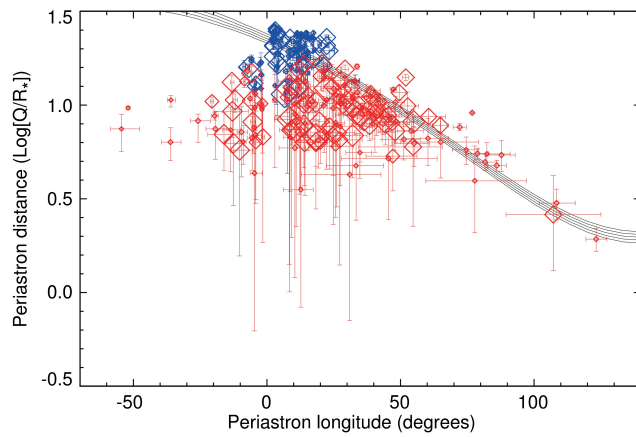
Figure 4: **Periastron distance versus periastron longitude.** Exocomets of population D (in blue) have large periastron distances, $Q_D \sim 18 \pm 4 R_\star$, and a narrow range of longitudes, $\varpi_D \sim 7 \pm 8°$. This population of exocomets could originate from the break-up of some parent bodies, which would liberate large amount of fresh volatiles buried in the cometary core, thus doping the gas and dust evaporation rates. Exocomets of population S (in red) have smaller periastron distances, $Q_S \sim 9 \pm 3 R_\star$, and a wide range of longitudes, $\varpi_S \sim 22 \pm 25°$. The Q-versus-ϖ relationship predicted for comets trapped in 4:1 mean motion resonance with a massive planet [18] (m'$\sim 10 M_J$, a'~ 4.5 AU, $\varpi' \sim -40°$, e'~ 0.04) is shown as black solid lines. Error bars represent the standard deviation.

Methods

This section describes the data analysis, the derivation of the exocomets' physical properties, and details of the interpretation.

Spectra HARPS (High Accuracy Radial velocity planet Searcher) is a high resolution (R~115,000) high precision ($\lesssim 1\,\mathrm{m\,s^{-1}}$) spectrograph installed at the 3.6 meter ESO telescope located at La Silla, Chile. The spectra were extracted from the detector images with the DRS pipeline of HARPS, which includes localization of the spectral orders on the 2D-images, optimal order extraction, cosmic-ray rejection, wavelength calibration, flat-field corrections, and 1D-reconnection of the spectral orders after correction for the blaze. A typical HARPS spectrum of β Pictoris includes the Ca II doublet lines (3933.66Å and 3968.47Å), which show contributions from (i) the rotationally-broadened stellar lines, (ii) the circumstellar gaseous disk, and (iii) variable absorption features (Fig. 1).

In order to compare β Pictoris spectra collected at different epochs, we first normalized all the spectra to the same mean flux level using the mean of the flux in the Ca II line's wings, where no variable absorption features are present. We then checked for possible shifts in wavelength calibration with time by using the Na I lines as a reference. The circumstellar Na I line is steep (Extended Data Fig. 1) and confirms the tremendous stability of the instrument during the observation campaign – as expected for this spectrograph aimed at detecting minute radial velocity variations. Table 2 in ref. 23 shows that the accuracy of HARPS is better than $1\,\mathrm{m\,s^{-1}}$ over several years. We thus have high confidence in the detected spectral variations.

We also checked for stellar variations in the Ca II spectrum over long time scales. To do so, we computed a reference spectrum of the Ca II stellar lines (see next paragraph) for each observational campaign (2003-2004, 2004-2008, 2008-2009, 2009-2011 and 2011). A thorough comparison of these reference spectra allowed us to exclude any significant variation of the stellar lines between 2003 and 2011. We hence decided to use the whole set of spectra to compute one common reference spectrum.

However, variations of the cicumstellar line's width by about $3\,\mathrm{km\,s^{-1}}$ are seen between 2003 and 2011. Since the HARPS spectrograph is not capable of resolving features below $3\,\mathrm{km\,s^{-1}}$ at the Ca II doublet's wavelengths, we discarded the part of the spectra corresponding to the circumstellar line region extending from $18\,\mathrm{km\,s^{-1}}$ to $24\,\mathrm{km\,s^{-1}}$ around the circumstellar line center at $21\,\mathrm{km\,s^{-1}}$, the systemic radial velocity of β Pictoris.

Derivation of the reference spectrum In order to characterize the absorption features, we divided each observed β Pictoris spectrum $F_{\mathrm{obs}}(\lambda)$ by a reference spectrum $F_{ref}(\lambda)$. The reference spectrum was obtained as described below and includes both stellar and circumstellar absorption components in the Ca II doublet. In the absence of exocomets'

transit, the normalized spectrum F_{obs}/F_{ref} is constant and equal to 1.

Each Ca II spectrum of β Pic shows a lot of variable absorption features of different depth and width. As a consequence, none of the 1106 spectra could be considered as an estimate of the reference spectrum. However, at any given wavelength λ_i around the stellar Ca II K and H lines, variable features appear and disappear randomly. As a result, amongst the whole set of flux measurements $F_k(\lambda_i)$ (for k=1,..., N=1106), a small fraction has no or little contamination by variable features, and can be used to compute the reference spectrum. We further assumed that, in the absence of variable absorption features, the noise δF in the flux measurements is Gaussian, and we checked that the RMS of the measurements is proportional to $\sqrt{F}$ with a constant factor independent of the wavelength.

First, we obtained an estimation of the reference spectrum by considering at each wavelength λ_i only the 2.5% highest flux values. Each measurement was subsequently sorted in a decreasing order such that $F_k(\lambda_i) \geq F_{k+1}(\lambda_i)$ (for k=1,..., N=1105). The 2.5% highest flux values ($k \leq 28$) are likely not contaminated by the variable absorption features and can be considered as an upper limit randomly drawn from the Gaussian distribution of the noise centered around the true reference flux. In this case, we can estimate the difference between any $F_k(\lambda_i)$ and the reference spectrum $F_{ref}(\lambda_i)$. Above $F_k(\lambda_i)$ there are k flux measurements that are a fraction k/N of the total number of flux measurements at the given wavelength. Assuming that the noise δF is Gaussian with first momentum σ=RMS, we compute the cut-off value α_k at which the probability that $\delta F/\text{RMS} > \alpha_k$ is k/N. We then obtain a first estimate of the reference flux:

$$F_{\text{ref},k}^{(1)}(\lambda_i) = F_k(\lambda_i) - \alpha_k \text{RMS}. \tag{1}$$

However, variable absorption features randomly appear at each wavelength, disturbing the derivation of the reference spectrum. In order to improve it, we introduced a second step in the computation using a larger number of flux values. For each k and wavelength λ_i, we computed the mean value $\overline{F}_k(\lambda_i)$ of the flux measurements $F_p(\lambda_i)$ for p such that $F_{\text{ref},k}^{(1)}(\lambda_i) < F_p(\lambda_i) < F_k^{\lambda_i}$. These flux measurements are above the first estimate of the reference spectrum and hence unlikely affected by variable absorption lines. This new step takes into account up to 500 flux measurements in the computation of the reference flux, to be compared to the 28 most extreme flux values used in the first step. Using the value $\overline{F}_k(\lambda_i)$, we obtain a new estimate of the reference spectrum given by

$$F_{\text{ref},k}^{(2)} = \overline{F}_k(\lambda_i) - \beta_k \text{RMS}, \tag{2}$$

where β_k is the average value of a normalized Gaussian variable in the range $[0, \alpha_k]$, given by

$$\beta_k = \frac{\int_0^{\alpha_k} x e^{-x^2/2} dx}{\int_0^{\alpha_k} e^{-x^2/2} dx} \tag{3}$$

13

Finally, we compute the final reference spectrum by taking the average of all the $F_{ref,k}^{(2)}$ with k varying from 3 to 28:

$$F_{\text{ref}} = \left\langle F_{\text{ref},k}^{(2)} \right\rangle_{3 \leq k \leq 28} \tag{4}$$

The application of this three-step method at all wavelength λ_i of the Ca II spectrum allowed us to obtain the reference spectrum plotted in Fig. 2. Its accuracy is such that a small interstellar line is detected on the left of the circumstellar line[24]. We performed at each wavelength a χ^2 test to compare the tail distribution of flux measurements going from F_{ref} to $F_{\text{ref}}+3\sigma$ with a Gaussian distribution of photon noise. The agreement is good, with 87% of the reference flux values passing the test at a significance level of 5%.

Fitting method We obtained normalized spectra by dividing each spectrum by the reference spectrum. The normalized spectra show exclusively variable absorption features, as can be seen by comparing Extended Data Fig. 1 and 3. Spectrosocopic variations were typically not seen on timescales less than 30 minutes (corresponding to the minimum duration of a transit), however to limit the effects of any possible spectral variability, the spectra were averaged into separate 10 minutes time intervals. We thus obtained 357 re-sampled spectra with $S/N>80$. Each of them contains an average of 6 variable absorption features with radial velocities between -150 and $200\,\text{km}\,\text{s}^{-1}$ with respect to β Pic. These features can be fitted by a Gaussian profile:

$$p_{K,H} e^{-4\ln 2 \frac{(v-v_0)^2}{\Delta v^2}}, \tag{5}$$

where $p_{K,H}$ are the line-depth in the Ca II K and H line respectively, v_0 is the radial velocity of the coma and Δv the line's FWHM. Because the absorption features are produced by gaseous clouds passing in front of the star, the depths $p_{K,H}$ are related to the cloud-to-star surface ratio $\alpha = \Sigma_c/\Sigma_\star \leq 1$ and the optical depth at the center of the absorption feature in the Ca II K line, or *absorption depth A*, through

$$p_K = \alpha \left(1 - e^{-A}\right)$$
$$p_H = \alpha \left(1 - e^{-A/2}\right) \tag{6}$$

The absorption depth A depends on the density and depth of the medium. This quantity is directly related to the p_K/p_H ratio characterizing the saturation in absorption within the cloud:

$$\frac{p_K}{p_H} = 1 + e^{-\frac{A}{2}} \tag{7}$$

With $\alpha \leq 1$ the relationship between Eqs. 6 and 7 becomes

$$1 \leq \frac{p_K}{p_H} \leq 2 - p_H \tag{8}$$

We fitted each variable absorption feature simultaneously in the Ca II-K and Ca II-H lines of all 357 normalized spectra, providing non-degenerate estimates of α, A, v_0 and Δv for each variable absorption component. Considering the huge number of spectra, and the large number of blended components in each spectrum makes the fitting of these features challenging. We developed a systematic procedure which we used to fit each spectrum automatically by searching for as many lines as required to best fit the data. Since the prior on the number of components is a uniform probability function, we used the Bayesian Information Criterion (BIC) to get the optimal number of components necessary to build a fair model of the normalized spectrum. The BIC is defined as: $\mathrm{BIC} = \chi^2 + k \ln N$, where N is the total number of data points and k the number of parameters. For each spectrum, we took a number n of components such that the fit with an additional component yield a BIC that does not decrease by more than 6. When $\mathrm{BIC}_1 - \mathrm{BIC}_2 < 6$, the model #2 has less than 95% probability to be closer to reality than the model #1.

A typical example of a resulting fit using this procedure is plotted in Extended Data Fig. 3.

Separation of the populations Figure 2 reveals the presence of two clusters of data separated at $\log A \sim 0.5$; one at small surface ratio ($\alpha < 0.2$) and the other one at a larger surface ratio ($\alpha > 0.5$). Extended Data Fig. 4 also suggests the presence of two clusters of features in the distribution of the depth of the K and H lines. We performed a statistical cluster analysis using the k-mean cluster algorithm [12] in the (p_H, p_K) diagram. This algorithm identified two clusters of data with $p_K < 0.4$ on one side and $p_K > 0.4$ on the other side. These two clusters are directly related to the two clusters in Fig. 2. We performed a Kolmogorov-Smirnov test to compare the distributions in parameters α, A, v_r and Δv of these two clusters. For each of these parameters, the two distributions are different with a probability $P > 0.9999$. A similar statistical cluster analysis performed in the (A, α)-space led to an analogous separation of the two populations.

Evaporation efficiency The quantities measured in the ionic Ca^+ cloud transiting in front of β Pictoris can be used to derive the physical properties of the exocomets such as their distance to the star and the dust production rate in their coma. To do so, we estimate the *evaporation efficiency* of each individual detected exocomet, a quantity describing the efficiency with which a comet captures and reprocesses input stellar energy flux into dust and gas evaporation from its nucleus.

1. Definition   After being damped by the opacity of the dust coma, the stellar radiation incident on the comet reaches the icy nucleus surface covered by a thin layer of agglomerated dust[16], and heats it up. A significant part of this heat energy is used to sublimate water from the ice of the nucleus, at a rate Z_{H_2O}; the remaining energy is absorbed by the crust or re-emitted by the surface.

The heat re-emitted or absorbed by the dusty surface is in part used to thermalize the sublimated gas. Assuming that the distance d of a comet to β Pic is around $10R_\star$ and given that $T_{\mathrm{eff}} \sim 8000\,K$, the temperature at the surface of the comet is $T_c = T_{\mathrm{eff}} \sqrt{R_\star/d} \sim 2500\,K$. As the gas heats up, it escapes from the nucleus surface and flows out radially. A water molecule heated at 2500 K is typically accelerated to a radial velocity of $1\,\mathrm{km\,s^{-1}}$ provided that $v_r \sim \sqrt{3\,k_B T / m_{H_2O}}$. While escaping, the gas picks up dust grains from the dust mantle and drags them outward, with a mass rate $\dot{M}$. Kinetic energy is thus transferred from the gas to the dust grains, which are then accelerated to the expansion velocity of water molecules[14,25], $v_n \sim 1\,\mathrm{km\,s^{-1}}$.

The efficiency of this evaporation process, which depends on the surface properties of the nucleus (*e.g.* size, albedo, fragmentation), can be measured by comparing the total energy used per unit time to evaporate dust at a mass rate $\dot{M}$ with the input energy per unit of time reaching the comet, *i.e.* the stellar flux. This leads to the following definition of the evaporation efficiency

$$\eta = \log \frac{Z_{H_2O} L_{H_2O} + \frac{1}{2} \dot{M} v_n^2}{F_\star(d)}. \tag{9}$$

We introduced $L_{H_2O} \sim 3 \times 10^3\,\mathrm{kJ\,kg^{-1}}$ is the latent heat of vaporization for water, and $F_\star(d)$ is the stellar flux at a distance d, which is related to the stellar luminosity $L_\star$ by $F_\star = L_\star/4\pi d^2$, with $L_\star = 8.6 L_\odot$. We neglected the gas kinetic energy that is an order of magnitude lower than the latent heat of water sublimation.

Reducing numerical factors, and assuming the dust-gas mass ratio is constant and close to 1 such that $Z_{H_2O} \sim \dot{M}$, the evaporation efficiency can be expressed as

$$\eta \simeq \log(\dot{M} d^2) - 1.9 \tag{10}$$

The distance d is expressed in stellar radius units and $\dot{M}$ in $\mathrm{kg\,s^{-1}}$. Typical values[9] are expected to be $10^7 < \dot{M} < 5 \times 10^8\,\mathrm{kg\,s^{-1}}$ and $10 < d < 50 R_\star$, yielding $7 < \eta < 10$.

2. Measurements of the evaporation efficiency   The evaporation efficiency can be estimated from the measured values of α and A using the conservation of momentum in the cometary cloud: the total momentum carried by the stellar photons which are absorbed by the Ca^+ cloud equates the total momentum gained by these Ca^+ atoms.

On one side, we thus consider the amount of photons absorbed by the Ca^+ cloud, accounting for contribution of both K and H lines, integrated over its projected surface $\Sigma_c = \pi R_\star^2 \alpha$ and $F_{\star,d}$ the stellar flux at distance d from the star.

This leads to an absorbed momentum flux:

$$\frac{dP}{dt}\bigg|_{\text{abs}} = \Sigma_c \int \left(1 - e^{-A}\right) e^{-4\ln 2 \frac{(v-v_0)^2}{\Delta v^2}} F_{\star,d}(\lambda) \frac{d\lambda}{c} \tag{11}$$

$$+ \Sigma_c \int \left(1 - e^{-A/2}\right) e^{-4\ln 2 \frac{(v-v_0)^2}{\Delta v^2}} F_{\star,d}(\lambda) \frac{d\lambda}{c}, \tag{12}$$

with $v - v_0 = c(\lambda - \lambda_0)/\lambda_0$ and $\lambda_0 = 3950 \text{Å}$.

On the other side, we consider the total momentum P gained per unit of time by the Ca^+ ions when they are accelerated from their initial velocity v_i to a final velocity v_f:

$$\frac{dP}{dt}\bigg|_{\text{acc}} = \dot{M}_{Ca^+} \left\langle v_f - v_i \right\rangle_{v_f} \tag{13}$$

We assume v_f to follow a distribution of typical width Δv and mean v_0. The variation of momentum is averaged with respect to this distribution, with $\left\langle v_f \right\rangle = v_0$.

These two independent expressions for the momentum flux can be computed for each line as a function of A, α and Δv following:

$$\frac{dP}{dt}\bigg|_{\text{abs}} = \alpha \pi R_\star^2 \frac{(19.3\,\text{pc})^2 F_\oplus(\lambda_0)}{d^2 R_\star^2} \frac{\lambda_0 \Delta v}{c^2} \sqrt{\frac{\pi}{4\ln 2}} \left(2 - e^{-A} - e^{-A/2}\right)$$

$$\frac{dP}{dt}\bigg|_{\text{acc}} = \dot{M}_{Ca^+}(v_0 - v_i) \sim \dot{M}_{Ca^+} \times \Delta v, \tag{14}$$

where $F_\oplus(\lambda_0) = 0.25 \times 10^{-10}\,\text{erg cm}^{-2}\text{s}^{-1}\text{Å}^{-1}$ is the β Pic stellar flux measured from Earth around wavelength $\lambda_0 \sim 3950\text{Å}$, accounting for a 0.25-reduction factor at the bottom of K and H lines ; the distance d is expressed in stellar radius units.

Momentum conservation implies that $dP/dt|_{\text{acc}} = dP/dt|_{\text{abs}}$, leading to

$$\dot{M}_{Ca^+} d^2 \sim 1.3 \times 10^9 \, \alpha \left(2 - e^{-A} - e^{-A/2}\right). \tag{15}$$

Taking into account the expected abundance of calcium in silicate[9] ($\dot{M}_{Ca^+} \sim 0.01 \dot{M}_{\text{dust}}$) we conclude:

$$\eta \sim 9.2 + \log \left[\alpha \left(2 - e^{-A} - e^{-A/2}\right)\right] \tag{16}$$

3. Discussion   The distribution of the measured evaporation efficiencies for the observed exocomets is in good agreement with typically expected values, $7 < \eta < 10$ (Extended Data Fig. 5). The two populations of exocomets have

distinct evaporation efficiency distributions with $\eta_S \sim 8.6 \pm 0.4$ and $\eta_D \sim 9.4 \pm 0.1$. Exocomets of population D are thus almost ten times more efficient at capturing and converting input stellar energy into gas and dust evaporation than exocomets of population S. In other words, they would have ten times higher dust production rate, if they were located at the same distance to the star as population S.

The uncertainties quoted above on the measure of the evaporation efficiency in each population do not include the effects of the model's approximations, in particular in the estimation of the cometary cloud area and the velocity of the Ca^+ ions. We estimated that model uncertainties by a factor of about 2 or less on these two quantities lead to an error bar of about ± 0.3 for the evaporation efficiency, which is about 3 times smaller than the difference seen between the two populations ($\Delta\eta \sim 0.8$). This is therefore additional evidence of the existence of two families of exocomets with distinct intrinsic properties.

Distance and dust production rate   Using the model of ref. 15, we can derive the distance between the star and the exocomets at the time of their transit. In this model, ions are supersonically dragged by the evaporating gas flowing out from the nucleus. At a large distance from the nucleus, the radially expanding ions are slowed down by the radiation pressure to a subsonic velocity and are further accelerated in the anti-stellar direction.

As a result, a shock surface is formed at the distance r_0 from the nucleus

$$r_0 = \left(\frac{\gamma F \dot{M} d^2}{8\pi \beta G M_\star a_1} \right)^{1/3} . \tag{17}$$

We introduced $\gamma = 0.01$ the typical calcium abundance in silicate grains; $\beta = 77$ the ratio of radiation pressure to gravity for Ca^+ ions in the β Pic environment; $M_\star = 1.7 M_\odot$ the β Pic mass; and $a_1 = \mu/\sigma$ a constant factor accounting for the shock surface[15] depending on μ the Ca^+ mass and σ the effective cross-section of the stellar flux absorption by the Ca^+ ions. The F factor[15] depends on the gas production rate Z (typically equal to dust production rate $\dot{M}$ in solar system comets)

$$F = \frac{\dot{M}}{2 v_e \mu \, m_{H_2O}} \sqrt{\frac{4 m_H \alpha_H e^2}{4\pi \varepsilon_0}}, \tag{18}$$

where m_{H_2O} is the mass of the water molecule, m_H is the hydrogen mass; $\alpha_H = 6.67 \times 10^{-31} \mathrm{m}^3$ is the polarizability of the hydrogen atom; and $v_e \sim 10\,\mathrm{km\,s}^{-1}$ is the dust expansion velocity just below the shock surface. The effective cross-section of stellar flux absorption by the Ca^+ ions is

$$\sigma = \sum_{K,H} \frac{\lambda_{K,H}}{\Delta v} \frac{1}{4\pi \varepsilon_0} \frac{\pi e^2}{m_e c} f_{K,H}, \tag{19}$$

introducing $f_K=0.69$ and $f_H=0.34$ the oscillator strengths of Ca II K and H lines; and $\Delta v \sim 1\,\text{km}\,\text{s}^{-1}$ the estimated transition line width taking into account the natural, thermal and collisional broadenings. Provided $r_0 \leq R_\star$, the surface ratio $\alpha = \Sigma_c/\Sigma_\star$ is roughly given by $\alpha \simeq r_0^2/R_\star^2$, yielding

$$\alpha = 4.5 \times 10^{-14} \dot{M}^{4/3} d^{4/3}. \tag{20}$$

Using Eq.15, we obtain the distance of each exocomet to β Pic at the time of the transit given in stellar radius

$$d = 6.2 \times 10^{-9}\, 10^{\eta}\, \alpha^{-3/4}. \tag{21}$$

Our measurements lead to estimates of the distance between 1 and $30\,R_\star$ (Extended Data Fig.6) as expected. The dust production rate $\dot{M}$ is then deduced from eq. (21) and (10).

Periastron distance and longitude   Assuming that each exocomet exhibit a near parabolic orbit, an estimate of the distance to the star together with the measurement of the radial velocity at the time of the transit allows an estimation of the periastron orientation and distance. We define ϖ the longitude of the periastron, which is the true anomaly of the line of sight with respect to the exocomet's periastron, and Q the periastron distance in stellar radius unit. These two quantities can be expressed with respect to the distance and the radial velocity by solving:
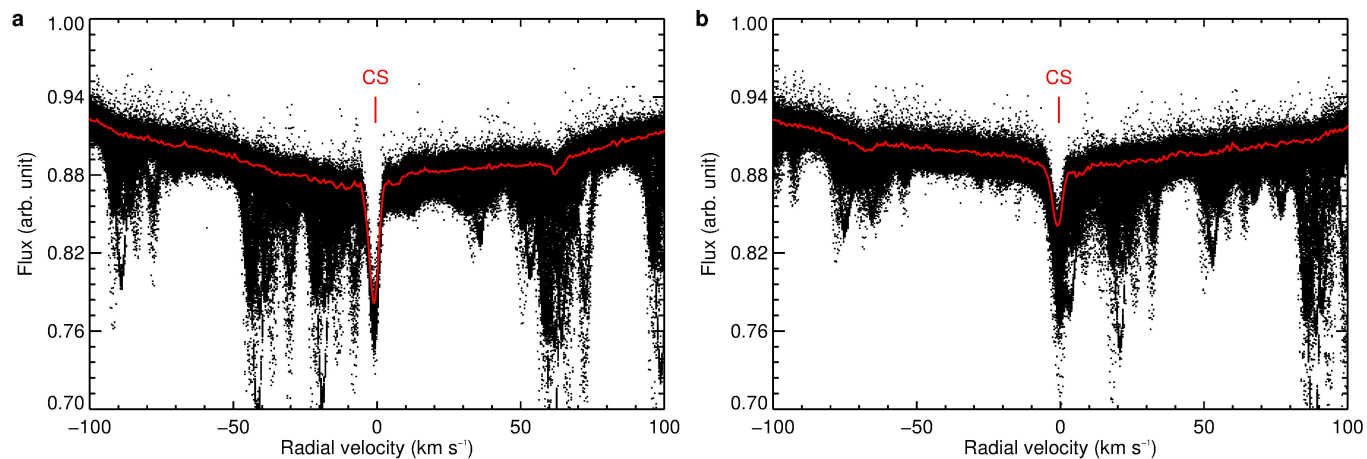
$$v_\perp = \sqrt{\frac{GM_\star}{R_\star d}} \frac{\sin \varpi}{\sqrt{1+\cos \varpi}}$$
$$d = \frac{2Q}{1+\cos \varpi} \tag{22}$$

The first equation is solved using a numerical inversion method, and the second is solved once ϖ is known. We plot the $(\varpi, \log Q)$ diagram in Fig. 4.
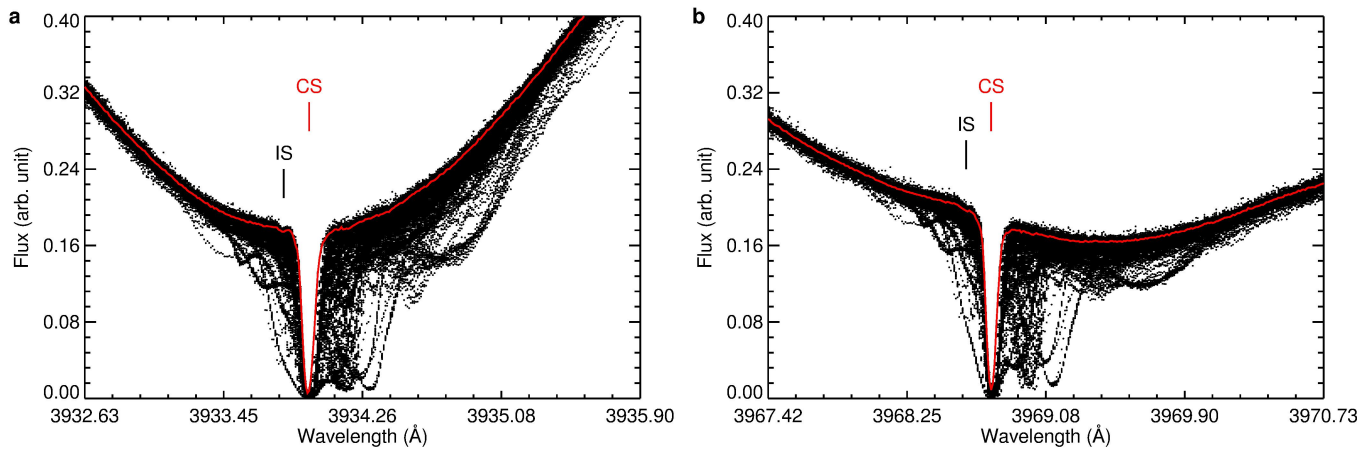
References

23. Pepe, F., Lovis, C., et al., The HARPS search for Earth-like planets in the habitable zone. I. Very low-mass planets around HD 20794, HD 85512, and HD 192310, Astronomy and Astrophysics, 534, A58 (2011)

24. Lallement, R., Ferlet, R., Lagrange, A. M., Lemoine, M., & Vidal-Madjar, A. Local Cloud structure from HST-GHRS. *Astronomy and Astrophysics* **304**, 461-(1995).

25. Beust, H., Lagrange-Henri, A. M., Vidal-Madjar, A., & Ferlet, R. The Beta Pictoris circumstellar disk. IX - Theoretical results on the infall velocities of CA II, AL III, and MG II. *Astronomy and Astrophysics* **223**, 304-312 (1989).
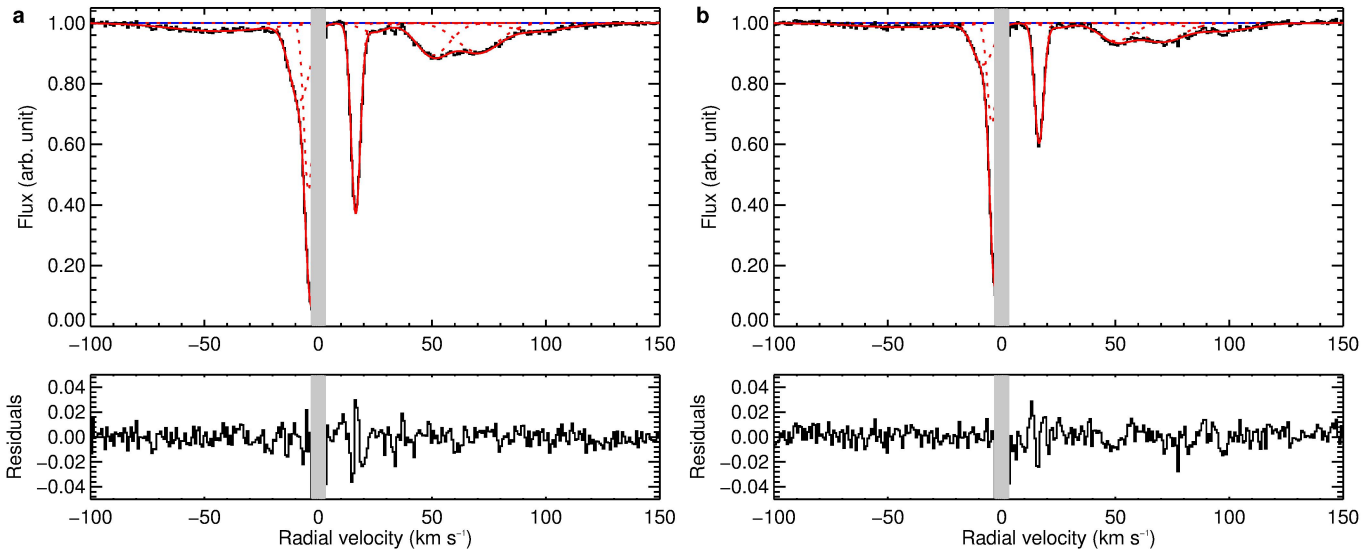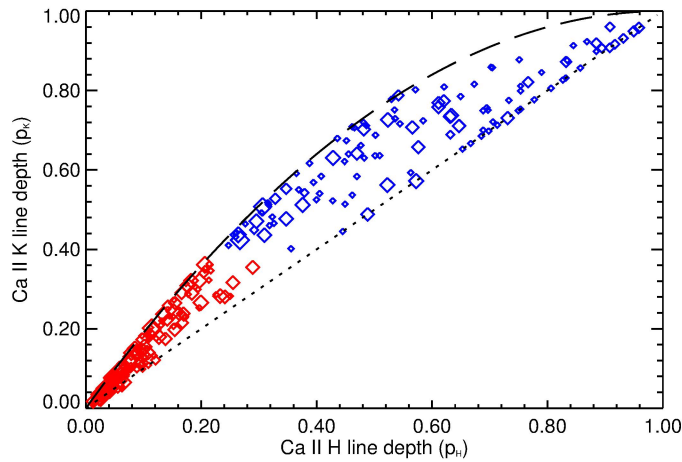
Extended data figures captions

Extended Data Figure 1: **The Na I spectrum of β Pictoris. a,** Na I D2-line spectrum (5890Å). **b,** Na I D1-line spectrum (5896Å). It shows the superposition of all Na I spectra of β Pic (black dots) compared with the stellar reference spectrum (red line). Radial velocities are given in the star's rest frame. The stable Na I line centered at the star's radial velocity is identified as due to the circumstellar (CS) disk. The sharpness of the Na I D1 and D2 lines and steadiness of this circumstellar feature in all spectra confirm the stability of HARPS over the timescales of years. The narrow absorption lines seen in most of the spectra and not in the calculated reference spectrum are due to atmospheric water.
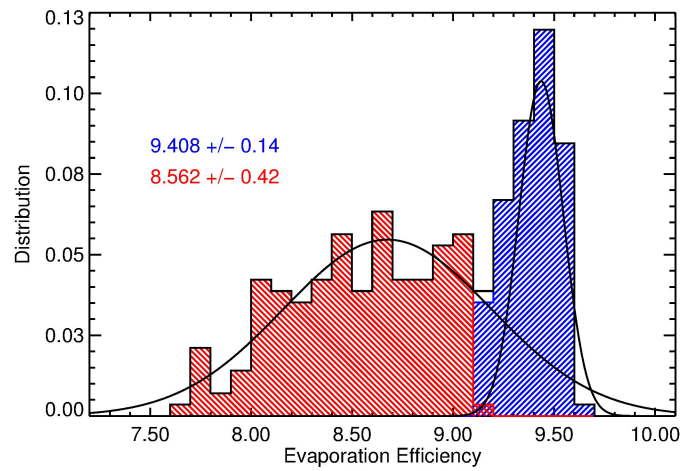
Extended Data Figure 2: **The Ca II reference spectrum of β Pictoris. a,** Ca II K-line spectrum (3933.66Å). **b,** Ca II H-line spectrum (3968.47Å). It shows the superposition of all the Ca II spectra of β Pic (black dots) compared with the stellar reference spectrum (red line). The stable circumstellar (CS) line is centered at the star's radial velocity. Variable absorption features are revealed by their diffuse shapes with respect to the dark upper envelop of the cloud of points. The predominance of redshifted absorption features is clearly visible. A small interstellar line is noticeable on the left of the circumstellar line[24].
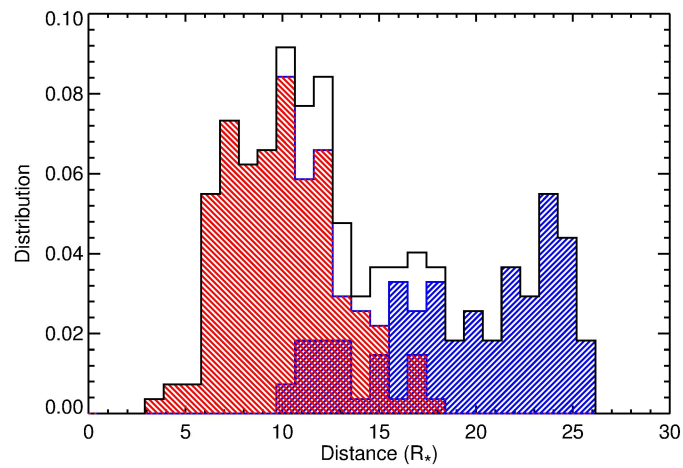
Extended Data Figure 3: **A typical fitted Ca II normalized spectrum. a,** Ca II K-line normalized spectrum. **b,** Ca II H-line normalized spectrum. The Ca II normalized spectrum (black line) is obtained through the division of the corresponding regular spectrum collected on the 27 October 2009 (Fig. 1) by the reference spectrum plotted in Extended Data Fig. 2. Radial velocities are given with respect to the stellar rest frame. The fit of each detected feature is detailed with red dashed lines, and their superposition with a solid red line. The bottom panels show the residuals of the fit. The gray-zone indicates the $\pm 3\,\mathrm{km\,s^{-1}}$ excluded CS-region, where variable absorption features due to exocomets are not resolved from the circumstellar line. This spectrum presents all types of variable absorption features: a broad and shallow absorption at large radial velocity ($\sim\pm 50\,\mathrm{km\,s^{-1}}$) and a sharp and deep absorption at small radial velocity ($\sim 20\,\mathrm{km\,s^{-1}}$).

23

Extended Data Figure 4: **Diagram of the Ca II line depths.** Plot of the Ca II K line depth, p_K, as a function of the Ca II H line depth, p_H, for independent absorptions features due to individual transiting comets observed between 2003 and 2011. Using cluster analysis of these line depths' measurements, two populations of exocomets show up: the population S generates the shallow absorption lines with $p_K < 0.4$ (in red) and the population D generates the deep absorption lines with $p_K > 0.4$ (in blue). The dotted line represents the full saturation limit $p_K = p_H$ and the dashed line represents the $\alpha = 1$ limit corresponding to cometary cloud with a projected area greater than the stellar disk area.

Extended Data Figure 5: **Histogram of the evaporation efficiency of transiting exocomets.** The plot of the histogram of η, the evaporation efficiencies (in black) shows a clear bimodal distribution: population S (in red) is centered on $\eta_S=8.6\pm0.4$, while population D (in blue) is centered on $\eta_D=9.4\pm0.1$. The solid black line represents the distribution of evaporation efficiency for the whole sample of observed exocomets. The two dashed curves are obtained by fitting the black curve with the superposition of two Gaussians.

Extended Data Figure 6: **Histogram of the distances between β Pic and the exocomets at the time of transit.**
The comets of population D (in blue) are located further away from the star than the comets of population S (in red),
with $d_D{\sim}19{\pm}4\,R_\star$ and $d_S{\sim}10{\pm}3\,R_\star$. Distances are expressed in stellar radius units. The solid black line represents the
distribution of distances for the whole sample of observed exocomets.