LETTER

doi:10.1038/nature13785

Water vapour absorption in the clear atmosphere of a Neptune-sized exoplanet

Jonathan Fraine[1,2,3], Drake Deming[1,4], Bjorn Benneke[3], Heather Knutson[3], Andrés Jordán[2], Néstor Espinoza[2], Nikku Madhusudhan[5], Ashlee Wilkins[1] & Kamen Todorov[6]

Transmission spectroscopy has so far detected atomic and molecular absorption in Jupiter-sized exoplanets, but intense efforts to measure molecular absorption in the atmospheres of smaller (Neptune-sized) planets during transits have revealed only featureless spectra[1–4]. From this it was concluded that the majority of small, warm planets evolve to sustain atmospheres with high mean molecular weights (little hydrogen), opaque clouds or scattering hazes, reducing our ability to observe the composition of these atmospheres[1–5]. Here we report observations of the transmission spectrum of the exoplanet HAT-P-11b (which has a radius about four times that of Earth) from the optical wavelength range to the infrared. We detected water vapour absorption at a wavelength of 1.4 micrometres. The amplitude of the water absorption (approximately 250 parts per million) indicates that the planetary atmosphere is predominantly clear down to an altitude corresponding to about 1 millibar, and sufficiently rich in hydrogen to have a large scale height (over which the atmospheric pressure varies by a factor of e). The spectrum is indicative of a planetary atmosphere in which the abundance of heavy elements is no greater than about 700 times the solar value. This is in good agreement with the core-accretion theory of planet formation, in which a gas giant planet acquires its atmosphere by accreting hydrogen-rich gas directly from the protoplanetary nebula onto a large rocky or icy core[6].

We observed transits of HAT-P-11b[7] (mass, $M_p = (25.8 \pm 2.9)M_\oplus$ ($M_\oplus$, Earth mass); radius, $R_p = (4.37 \pm 0.08)R_\oplus$ ($R_\oplus$, Earth radius); equilibrium temperature, $T_{eq} = (878 \pm 50)$ K) in a joint programme involving NASA's Hubble and Spitzer space telescopes. Our Hubble observations comprised 1.1–1.7 μm grism spectroscopy using the Wide Field Camera 3 (WFC3) in spatial scanning mode. We also integrated these data over wavelength to produce WFC3 photometry[1–4,8]. Our Spitzer observations comprised photometry during two transits in each of the 3.6 and 4.5 μm bands of the Infrared Array Camera[9] (IRAC). Because the planet lies in the field of view of NASA's Kepler spacecraft[10], precision optical photometry (~642 nm) was obtained simultaneously with our Spitzer observations, although not simultaneously with our Hubble observations. Table 1 summarizes specific details of our observations, and Fig. 1a shows our transit photometry and model fits. Because HAT-P-11 is an active planet-hosting star[11–13], we show that starspots on the stellar surface are not sufficiently cool, nor sufficiently prevalent, to mimic the effect of water vapour absorption in the planet[14]. Our simultaneous Spitzer and Kepler photometry was critical to defining the temperature of the starspots that could otherwise, potentially mimic the effect of water vapour absorption in the planetary atmosphere.

HAT-P-11b crosses starspots on virtually every transit[12,13], as seen prominently Fig. 1a. Our WFC3 photometry has the sensitivity to detect starspot crossings[2], but none were observed when Hubble observed the system. Our WFC3 observations contain large temporal gaps because Hubble passes behind the Earth[1–4,8,15], but not during the transit. Therefore, unocculted starspots, rather than occulted ones, potentially affect our transmission spectrum[16,17]. When the planet blocks unspotted portions of the stellar photosphere, the absorption lines in cool unocculted spots become relatively more prominent[12,13].

Figure 1a shows the binned and normalized light curves of our four simultaneous Kepler–Spitzer transits and our WFC3 band-integrated light curve. We fitted analytic transit light curves to all of time series with PyMC[18] to generate Markov chain Monte Carlo (MCMC) distributions to estimate the planetary parameters[19,20]. We reanalysed the phased and binned Kepler data using improved limb-darkening coefficients derived from stellar model atmospheres[21]. To fit the Spitzer and WFC3 transits, we held the orbital distance and inclination constant at our Kepler-derived values. Although the uncertainties in the Kepler-derived parameters were smaller than in previous studies[12,13], our purpose was to implement the updated limb-darkening law and derive orbital parameters for all of our observations.

Each of the Kepler light curves obtained concurrently with our Spitzer observations shows starspot crossings as deviations in the light curves between ~0.3 and 0.7 h after mid-transit. The amplitude of these deviations is a function of both the area and the temperature of the occulted spots[12,13]. Because the Kepler and Spitzer photometries were concurrent, the relative intensity is independent of the starspots' areas. However, because the contrast between starspot temperature and the photosphere is a chromatic effect, the amplitude of these deviations varied with wavelength[16,17]. The spot crossings are not obvious in the Spitzer data because thermal radiation produces a much smaller contrast between the stellar photosphere and spot fluxes in the infrared than in the optical. The ratio between the Spitzer and Kepler spot crossing amplitudes constrained lower limits on the starspot temperatures for the crossed starspots.

We included the relative shape of the spot crossings, sliced from each residual Kepler light curve, and scaled their amplitudes as free parameters in our MCMC analysis with our Spitzer transits. The distributions of the Spitzer/Kepler spot crossing amplitude ratios are shown in Fig. 1b. The dashed black lines represent the predicted spot crossing amplitude ratios for given spot temperature contrasts. We calculated these temperatures by representing the spots using model stellar atmospheres at various temperatures[22]. Using χ^2 difference (δ^{χ^2}) tests, we indirectly detected spot crossings only at 3.6 μm because only these Spitzer observations resulted in positive, bounded photosphere-to-spot temperature contrasts. The 4.5 μm Spitzer observations are consistent with zero, or a non-detection at infrared wavelengths. These measurements, especially the non-detections, imply that the starspots crossed during each transit are too hot to mimic water vapour absorption features in the planetary spectrum[12,14]. Our starspot analysis is described in Methods along with the distribution of Kepler spot crossing amplitudes, for comparison with those observed concurrently with Spitzer.

The activity of HAT-P-11[7,11–13] produces variations in the total brightness of the star owing to the rotation of spots in and out of view, which will change the band-integrated transit depth measured at different epochs. If the relative stellar brightness at the epoch of each observation is known, then the transit depths can be corrected to a common

[1]Department of Astronomy, University of Maryland, College Park, Maryland 20742-2421, USA. [2]Instituto de Astrofísica, Pontificia Universidad Católica de Chile, 7820436 Macul, Santiago, Chile. [3]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. [4]NASA Astrobiology Institute's Virtual Planetary Laboratory, Seattle, Washington 98195, USA. [5]Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK. [6]Department of Physics, ETH Zürich, 8049 Zürich, Switzerland.

Table 1 | Summary of observations

Date (UT)	Start time	End time	Observatory (instrument)	Band pass (μm)	Spectral resolution	Cadence (s)	Number of observations
7 Jul. 2011	23:11:42	06:37:52	Warm Spitzer (IRAC Channel 1)	3.6	~4	0.4	62,592
5 Aug. 2011	07:02:48	14:28:58	Warm Spitzer (IRAC Channel 2)	4.5	~4	0.4	58,112
15 Aug. 2011	01:49:20	09:15:30	Warm Spitzer (IRAC Channel 1)	3.6	~4	0.4	52,633
29 Aug. 2011	17:37:18	01:03:28	Warm Spitzer (IRAC Channel 2)	4.5	~4	0.4	62,592
18 Oct. 2012	17:37:18	01:03:28	Hubble (WFC3 G141)	1.13–1.64	~60–89	123	113
24 Dec. 2012	23:56:56	03:05:58	Hubble (WFC3 G141)	1.13–1.64	~60–89	123	99

We observed HAT-P-11b during four warm Spitzer observations, two transits at both 3.6 and 4.5 μm with the IRAC instrument[9], and two observations using the Hubble WFC3 G141 grism spectrometer, spanning 1.1–1.7 μm. Concurrent Kepler observations were retrieved for comparison with our warm Spitzer observations, but were unavailable for our Hubble spectroscopic observations.

value. Kepler measured HAT-P-11's relative brightness during all four Spitzer observations, but not during our WFC3 observation. The unknown stellar brightness during this observation introduced an additional uncertainty in our estimate of the WFC3 transit depth, relative to the Spitzer and Kepler observations, of ±51 p.p.m. In Fig. 2, the offset between the WFC3 spectrum and the best-fit model is ~93 p.p.m. on average.

Figure 2 shows our HAT-P-11b transmission spectrum with Kepler, WFC3 and Spitzer transits combined. We constrain the atmospheric composition using the SCARLET tool, which is a new version of the Bayesian retrieval framework described in previous studies[23,24]. Our primary results are a robust 5.1σ detection of water absorption in the WFC3 data and a 3σ upper limit on HAT-P-11b's atmospheric metallicity (the proportion by mass of elements heavier than hydrogen and helium) of ~700 times the solar metallicity[5], corresponding to a mean molecular weight of ~10.2 g mol^{-1} at the 10 mbar level (Fig. 3). Transmission spectra of selected atmospheric models[23–26] are plotted for a comparison with the observations in Fig. 2, with colour-matched symbols in Fig. 3. Although the significance of the water vapour detection is unaffected by uncertainties in the stellar activity because all wavelengths in the water band are measured simultaneously, this uncertainty made it prohibitively difficult to place robust constraints on the methane and carbon dioxide abundances and, therefore, the C/O ratio of HAT-P-11b's atmosphere[26].

Figure 3 shows that constraints on the atmospheric metallicity and cloud-top pressure are correlated. Atmospheric compositional scenarios along a curved distribution agree with the data at 3σ, spanning a range of atmospheric metallicities from 1 to 700 times the solar metallicity. Figure 2 shows that a representative 10,000-times-solar (water-dominated) spectrum is robustly excluded by the data. The high mean molecular weight of this atmosphere would not allow the significant water absorption feature observed in the WFC3 band pass.

We found that models with atmospheric metallicities corresponding to solar metallicity require the presence of small-particle hazes to match the HST and Kepler data points simultaneously. The fit to the data improves towards higher metallicities, reaching the best-fit value at 190 times solar metallicity. The presence of the water absorption feature in the WFC3 spectrum requires that any cloud deck must be at an altitude higher than the 10 mbar pressure level (Fig. 3), and the Kepler and Spitzer transit depths impose a similar lower limit on the cloud top pressure.

The atmospheric and bulk compositions of exoplanets provide important clues to their formation and evolution. Mass and radius alone do not provide unique constraints on the bulk compositions of these planets,

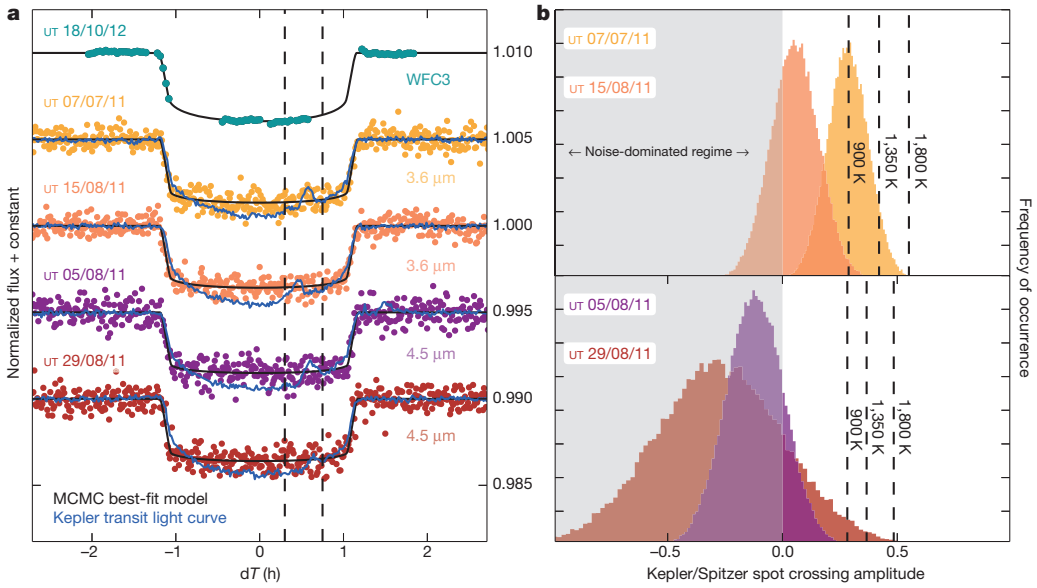


Figure 1 | White-light transit curves and starspot crossing temperature estimates. a, Transit curves from the Hubble WFC3 and warm Spitzer, aligned in phase and shifted in flux for clarity. The four warm Spitzer transits at both 3.6 and 4.5 μm (ref. 9) are binned for illustration. Starspot crossings are seen as deviations near +0.5 h in the Kepler photometry (blue). **b**, We estimated the starspot temperatures by dividing the Spitzer transit residuals by the Kepler transit residuals (colours as in **a**). The dashed lines represent the photosphere-to-starspot temperatures for three stellar model atmospheres[22]. Water vapour has been detected in sunspots as cool as 3,000 K, corresponding to a contrast of ~1,800 K here[14]. There is essentially no starspot temperature that can produce sufficiently strong water absorption to mimic our result.

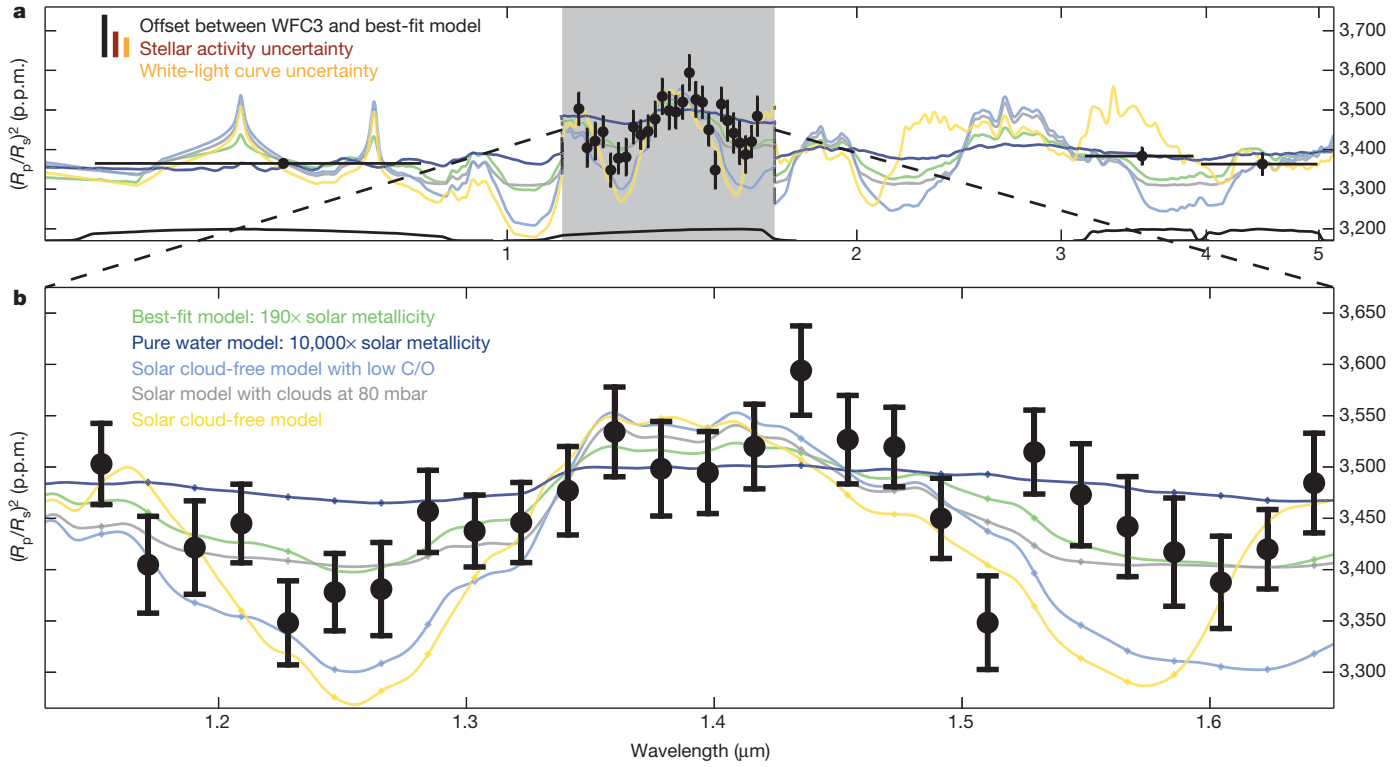


Figure 2 | The transmission spectrum of HAT-P-11b. **a**, Our WFC3 observations show transit depth variations in agreement with a hydrogen-dominated atmosphere. The coloured, solid lines[23,24] correspond to matching markers displayed in Fig. 3. The error bars represent the standard deviations over the uncertainty distributions. An atmosphere with a high mean molecular mass (dark blue line) is ruled out by our observations by $>3\sigma$. The WFC3 spectrum was allowed to shift, as a unit, over these uncertainties. R_s, stellar radius. **b**, Detailed view of our WFC3 spectrum. For the purposes of visually comparing the spectral significance, we shifted all of the models by 93 p.p.m. in the grey region in **a** and in **b**.

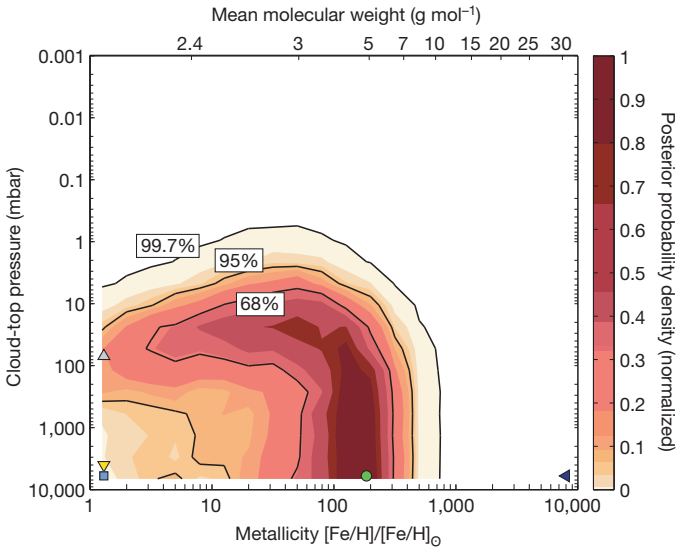


Figure 3 | Spectral retrieval results of our transmission spectrum. The coloured regions indicate the probability density as a function of metallicity (relative to solar) and cloud-top pressure derived using our Bayesian atmospheric retrieval framework[23,24]. Mean molecular weight was derived for a solar C/O ratio at 10 mbar. Black contours mark the 68%, 95% and 99.7% Bayesian credible regions. The depth of the observed water feature in the WFC3 spectrum required the presence of a large atmospheric scale height that can only be self-consistently obtained for an atmospheric metallicity below 700 times solar at 3σ (99.7%) confidence. The atmosphere is probably predominately cloud-free at least down to the 1 mbar level. We indicate the matching models plotted in Fig. 2 with coloured markers.

which are degenerate for various combinations of rock, ice and hydrogen gas[27]. By measuring the mean molecular weight of the atmosphere using transmission spectroscopy, we can resolve these degeneracies and provide stronger constraints on the interior compositions of these planets[5,24,27,28]. Observations of water vapour dominate the shape of the infrared spectral features for warm (planetary temperature, $T_p \approx 1{,}000$ K) exoplanets. In contrast, the featureless transmission spectra observed for several similarly small planets[1–4,16] ($R_p \approx 3R_\oplus$–$4R_\oplus$) imply that scattering hazes, clouds or high mean molecular weights exist in those atmospheres, obscuring absorption features[5,23,24] and limiting our ability to understand their interiors directly[5,24,27]. HAT-P-11b is the smallest and coldest planet with an absorption signature measured by transmission; this allows the estimation of its atmosphere's mean molecular weight, providing new insights into the formation history of this Neptune-mass planet[5,24,27–30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 April; accepted 7 August 2014.

1. Knutson, H. *et al.* A featureless transmission spectrum for the Neptune-mass exoplanet GJ 436b. *Nature* **505,** 66–68 (2014).
2. Kreidberg, L. *et al.* Clouds in the atmosphere of the super-Earth exoplanet GJ 1214b. *Nature* **505,** 69–72 (2014).
3. Knutson, H. *et al.* Hubble Space Telescope near-IR transmission spectroscopy of the super-Earth HD 97658b. Preprint at http://arxiv.org/abs/1403.4602 (2014).
4. Ehrenreich, D. *et al.* Near-infrared transmission spectrum of the warm-Uranus GJ 3470b with the Wide Field Camera-3 on the Hubble Space Telescope. Preprint at http://arxiv.org/abs/1405.1056v3 (2014).
5. Moses, J. *et al.* Compositional diversity in the atmospheres of hot Neptunes, with application to GJ 436b. *Astrophys. J.* **777,** 34 (2013).
6. D'Angelo, G., Durisen, R. H. & Lissauer, J. J. in *Exoplanets* (ed. Seager, S.) 319–346 (Univ. Arizona Press, 2010).

7. Bakos, G. *et al.* HAT-P-11b: a super-Neptune planet transiting a bright K star in the Kepler field. *Astrophys. J.* **710,** 1724–1745 (2010).
8. Deming, D. *et al.* Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophys. J.* **774,** 95 (2013).
9. Fazio, G. *et al.* The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J.* **154,** 10 (2004).
10. Borucki, W. *et al.* Kepler Planet-Detection Mission: introduction and first results. *Science* **327,** 977–980 (2010).
11. Knutson, H., Howard, A. & Isaacson, H. A correlation between stellar activity and hot Jupiter emission spectra. *Astrophys. J.* **720,** 1569–1576 (2010).
12. Deming, D. *et al. Kepler* and ground-based transits of the exo-Neptune HAT-P-11b. *Astrophys. J.* **740,** 33 (2011).
13. Sanchis-Ojeda, R. & Winn, J. Starspots, spin-orbit misalignment, and active latitudes in the HAT-P-11 exoplanetary system. *Astrophys. J.* **743,** 61 (2011).
14. Bernath, P. Water in sunspots and stars. *Int. Astron. Union Symp.* **12,** 70 (2002).
15. Berta, Z. *et al.* The GJ1214 super-Earth System: stellar variability, new transits, and a search for additional planets. *Astrophys. J.* **736,** 12 (2011).
16. Fraine, J. D. *et al.* Spitzer transits of the super-Earth GJ1214b and implications for its atmosphere. *Astrophys. J.* **765,** 127 (2013).
17. Sing, D. *et al.* Hubble Space Telescope transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *Mon. Not. R. Astron. Soc.* **416,** 1443–1455 (2011).
18. Patil, A., Huard, D. & Fonnesbeck, C. PyMC: Bayesian stochastic modelling in Python. *J. Stat. Softw.* **35,** 4–85 (2010).
19. Ford, E. Quantifying the uncertainty in the orbits of extrasolar planets. *Astron. J.* **129,** 1706–1717 (2005).
20. Ford, E. Improving the efficiency of Markov chain Monte Carlo for analyzing the orbits of extrasolar planets. *Astrophys. J.* **642,** 505–522 (2006).
21. Castelli, F. & Kurucz, R. New grids of ATLAS9 model atmospheres. Preprint at http://arxiv.org/abs/astro-ph/0405087 (2004).
22. Husser, T.-O. *et al.* A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. *Astron. Astrophys.* **553,** A6 (2013).
23. Benneke, B. & Seager, S. Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy. *Astrophys. J.* **753,** 100 (2012).
24. Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile-dominated super-Earths. *Astrophys. J.* **778,** 153 (2013).
25. Madhusudhan, N. *et al.* A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b. *Nature* **469,** 64–67 (2011).
26. Madhusudhan, N. C/O ratio as a dimension for characterizing exoplanetary atmospheres. *Astrophys. J.* **758,** 36 (2012).
27. Rogers, L. & Seager, S. A framework for quantifying the degeneracies of exoplanet interior compositions. *Astrophys. J.* **712,** 974–991 (2010).
28. Fortney, J. *et al.* A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775,** 80 (2013).
29. Chiang, E. & Laughlin, G. The minimum-mass extrasolar nebula: in situ formation of close-in super-Earths. *Mon. Not. R. Astron. Soc.* **431,** 3444–3455 (2013).
30. Hu, R. & Seager, S. Photochemistry in terrestrial exoplanet atmospheres. III. Photochemistry and thermochemistry in thick atmospheres on super Earths and mini Neptunes. *Astrophys. J.* **784,** 63 (2014).

Acknowledgements J.F., A.J. and N.E. acknowledge support from project IC120009 'Millennium Institute of Astrophysics (MAS)' of the Millennium Science Initiative, Chilean Ministry of Economy; FONDECYT project 1130857; and BASAL CATA PFB-06. N.E. is supported by CONICYT-PCHA/Doctorado Nacional. We thank P. McCullough for his assistance in the planning and execution of our observations. We are grateful to I. Crossfield, L. Kreidberg and E. Agol for providing their open-source, Python code banks on their individual websites. We are also grateful for discussions with M. Line, J. Fortney and J. Moses about the nature of photochemistry and interior structures. We thank the ATLAS and PHOENIX teams for providing stellar models. We also thank the SciPy and NumPy associations for providing extensive and rigorous numerical routines for an assortment of mathematical and computational techniques.

Author Contributions J.F. led the data analysis for this project with contributions from D.D., H.K., N.E., A.J. and A.W. A.W. supplied Hubble spectral fitting routines and interpretations. N.E. and A.J. supplied Python routines for MCMC, wavelet and transit curve analyses specific to transiting exoplanets. D.D., H.K., N.E. and A.J. provided computational equipment and administration. D.D., N.M., H.K. and K.T. successfully applied for and provided data from Hubble. B.B. and N.M. provided atmospheric models and accompanying fits. B.B. performed atmospheric retrieval analysis and provided figures and interpretations. N.E. supplied stellar limb-darkening coefficients calculated from both ATLAS and PHOENIX models.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F. (jfraine@astro.umd.edu).

METHODS

Observations. We observed two transits of HAT-P-11b using Hubble Space Telescope (HST) spectroscopy near 1.4 μm, and we obtained photometry for two transits at 3.6 μm and two at 4.5 μm using warm Spitzer. We also analysed 208 archival Kepler transits to assess the effect of starspots and update the optical transit depth[12] (Extended Table 1); Kepler observed HAT-P-11 during our Spitzer, but not HST, observations.

Hubble WFC3 spectroscopy and white-light photometry. We observed HAT-P-11 using the HST Wide Field Camera 3 (WFC3) in spatial scan mode (Table 1). We used the G141 grism, with a binned, four-column spectral resolution from 60–89 over the wavelength range 1.1–1.7 μm. Gaps in the HST observations (Extended Data Fig. 1) occur every ~45 min during occultations of the Earth.

We scanned the spectrum in the cross-dispersion direction to maximize efficiency[8,31]. Each scan covered 135 pixels in 44 s (~0.3981″ s^{-1}), yielding ~45,000 electrons per pixel (~70% of saturation). The average photon-limited signal-to-noise ratio is ~220 per pixel, integrating to a signal-to-noise ratio of ~2,500 per column, for 113 spectral images in transit 1. We were unable to use the second transit because HST's fine guidance sensors lost positional stability, which is not uncommon in this observing mode[2,8]. This also occurred eight times during transit 1. We further removed the entire first orbit and the first image of each orbit—a common practice to ameliorate instrumental effects[8,15,32]—yielding 72 images for photometric and spectroscopic measurement.

Warm Spitzer IRAC. Spitzer transits were critical to establish that starspots on HAT-P-11 are not sufficiently cool to exhibit stellar water absorption masquerading as planetary atmospheric absorption. Spitzer also provided a long-wavelength baseline for the planet's radius, minimizing the effect of Rayleigh scattering while remaining sensitive to absorption from carbon-containing molecules such as methane and carbon monoxide.

Kepler archival transits. We used all 208 archival transits of HAT-P-11 that Kepler observed at ~0.6 μm. The out-of-transit photometry yielded constraints on the disk-integrated activity of the host star (see below). We re-fitted the phased and binned Kepler light curve that was analysed previously[12,13], using fourth-order limb-darkening coefficients to improve the optical radius and geometric parameters of the system. The occurrence of 298 starspot crossings allowed us to characterize the amplitude distribution of spots crossed by HAT-P-11b during transit.

Limb-darkening coefficients. We used a single set of planetary orbital parameters for all analyses, and wavelength-dependent, four-parameter, nonlinear limb-darkening coefficients[33] (LDCs). For the Kepler band pass, the WFC3 band-integrated band pass, and the Spitzer band pass, we computed the LDCs by integrating stellar model intensities over each instrument's response function. For the 128 individual wavelengths from the WFC3 grism, we used the intersection of the WFC3 response function and a one-column (4.71 nm) square window centred at each wavelength. We held the computed LDCs constant during subsequent analyses.

We represented HAT-P-11 using an ATLAS model[21] (http://kurucz.harvard.edu/grids.html) with a stellar effective temperature of $T_{\rm eff}$ = 4,750 K (ref. 7), [M/H] = +0.3 and log(g) = 4.5. To ensure that our exoplanet spectrum is not sensitive to the stellar parameters, we also derived it using LDCs with $T_{\rm eff}$ = 4,500 K and $T_{\rm eff}$ = 5,000 K. Both the exoplanet spectrum and the white-light (band-integrated) transit depth varied negligibly (~1 and ~5 p.p.m., respectively) between the three stellar models. Repeating our analysis with both quadratic and three-parameter limb-darkening laws, we found similarly negligible effects (~1 and ~10 p.p.m.).

Hubble Wide Field Camera 3. Each WFC3 spatial scan comprised seven non-destructive reads, with 7.35 s of exposure per read. We combined them by subtracting each read from the previous read, applying a spatial mask to the difference and adding all of the masked differences to an initially blank image to create the spectral frame[1–3,8]. We used edge detection algorithms[34] to determine the edges of each combined image in the scanning direction and masked all pixels within 20 pixels of these edges in the scanning direction, to keep sky background and other noise from accumulating in the final spectral image. We identified bad pixels in the spectral images using a median filter with a 4σ threshold, over a seven-pixel window, and assigned bad pixel values to the median of the window. Extended Data Fig. 2 shows a spectral frame from the first visit (the first Hubble observation of an individual transit epoch) displaying the spectral (dispersion) and spatial (cross-dispersion) dimensions. The curves in Extended Data Fig. 2b show the averaged, column-integrated spectral template (red), before (top) and after (bottom) being fitted to the example spectrum (blue) for both the wavelength solution and the white-light photometry, as a function of time.

We calculated the wavelength flat-field calibration using standard procedures[35]. We fitted a two-dimensional Gaussian to the spectral images, and found that the spectra shifted by at most 1.12 pixels in the wavelength direction; we corrected this shift during the template fitting described below. The final, column-integrated spectra were derived by dividing each raw spectral image by its corresponding flat field and subtracting the per-column background values, and then summing the detector in the cross-dispersion direction (down the columns).

The sky background was calculated per-column as the median of the portion of the spectral image not scanned by the instrument. In Extended Data Fig. 2a, this corresponds to the blue regions above and below the red/orange spectral information. The background values varied by <3% from 1.1 to 1.7 μm, but were uncorrelated[34] with the resultant planetary spectrum.

To derive the WFC3 spectrum for HAT-P-11b, we used the established technique[1,3,8,32] of spectral template fitting. We formed the spectral template by averaging the out-of-transit spectra, and fitted it to individual grism spectra in both wavelength and amplitude[34], using both Levenberg–Marquardt and spline interpolation algorithms[34]. The fitted amplitudes as functions of time yield the band-integrated white-light curve (WLC) (Extended Data Fig. 1). The WLC defines the average transit depth over the total WFC3 band pass.

WFC3 exponential and linear baselines. The raw WLC contained both the transit and instrument effects in the form of exponential ramps[1,2,8,15] over each orbit, and a linear trend over each visit. We simultaneously fitted for exponential parameters as a function of HST's orbital phase, $E_{\rm R}(\theta; A, S)$, a linear trend as a function of time, $L(t; m, b)$, and an additive offset for the second half of the in-transit data, $O(t_i; \theta, O_{\rm o})$ ($i \in \{0, 1, \ldots, 113\}$) (see below), with an analytic transit light curve[36], MA(t; p, $T_{\rm c}$, P, b, a, u_1, u_2):

$$E_{\rm R}(\theta; A, S) = 1 + A\exp\{S(\theta_i - \theta_{\rm min})\}$$

$$O(t_i; \theta, O_{\rm o}) = O_{\rm o}\delta(\theta_i - \theta_2)\chi(t_i)$$

$$L(t; m, b) = m(t_i - t_{\rm min}) + b$$

$$\text{Model} = \text{MA}(t; p, T_{\rm c}, P, b, a, u_1, u_2)E_{\rm R}(\theta; A_{\rm o}, S_{\rm o})L(t; m, b) + O(t_i; \theta, O_{\rm o})$$

Here θ represents HST's orbital phase, $\{A, S\}$ are the exponential amplitude and scale factor, and $\{m, b\}$ are the slope and intercept of the linear function. $O(t_i; \theta, O_{\rm o})$ is a small (~100 ± 50 p.p.m.) step function, $\chi(t_i)$, correcting an unexplained offset in the band-integrated photometry (see Extended Data Fig. 1 between phases ~0.2 and ~0.7). The offset is probably related to a small shift in the position of the spectrum that does not occur in other HST orbits. Including the offset in our model improved the WLC fits significantly, without degrading the Bayesian information criterion[37,38] (BIC) or altering the significance of the water detection.

We tested four different models for WFC3's WLC exponential baseline[8,15,37], and selected among them on the basis of the BIC. The BIC for our adopted model differed only slightly from optimum ($\Delta\text{BIC} \ll 2$)[37,38], but the transit depth was more physically realistic when compared to atmospheric models. WFC3 WLCs are known to have noticeable red noise[1,8,32], so we also implemented a wavelet analysis to include both the white noise ($\sigma_{\rm w}$ = 12.81 p.p.m.) and the red noise ($\sigma_{\rm r}$ = 61.89 p.p.m.) components of the residuals into our final uncertainties[39].

We use a MCMC procedure[18] to fit the transit and instrument parameters simultaneously, thereby incorporating correlations between parameters into our reported uncertainties. Extended Data Fig. 3 compares the posteriors graphically. The Pearson correlation coefficient over each parameter showed the correlations to be insignificant ($-0.10 < \max_{i,j}(P) < 0.10$ for all $i, j \in$ {fitted parameters}).

HST WFC3 exoplanet spectrum derivation. We calculated the planetary spectrum differentially relative to the WLC, by dividing the spectral template, $S_{\rm T}(\lambda)$, into each individual spectrum $S(\lambda, t_i)$ (calculating $S(\lambda, t_i)/S_{\rm T}(\lambda)$), allowing for small wavelength shifts (<1.12 pixels). The planetary spectrum, $P(\lambda)$, is derived from the normalized residuals, or differential light curves:

$$\text{DLC}(\lambda, t_i) = (S(\lambda, t_i) - S_{\rm T}(\lambda, t_i))/S_{\rm T}(\lambda, t_i)$$

We fitted the DLC(λ, t_i), with differential analytic light curves, DALC(λ, t_i), by renormalizing the analytic WLCs[36], MA(t_i):

$$\text{DALC}(\lambda, t_i) = \frac{\text{MA}(t_i) - \max_t\{\text{MA}(t_i)\}}{\max_t\{\text{MA}(t_i)\} - \min_t\{\text{MA}(t_i)\}} + \max_t\{\text{MA}(t_i)\}$$

where the minima and maxima are taken over the time domain.

We fitted the normalization amplitude, $P(\lambda)$, of the DALCs, simultaneously with wavelength-dependent linear trends, using linear matrix inversion

$$\text{DLC}(\lambda, t_i) = P(\lambda)\text{DALC}(\lambda, t_i) + m_\lambda(t_i - t_{\rm min}) + b_\lambda$$

Extended Data Fig. 4 shows all 32 wavelength dependent light curves (blue to red) and corresponding analytic light curves (black) ranging from 1.167 to 1.675 μm

with 18 nm spacing in wavelength. We used the linear matrix fits as initial conditions for MCMCs to probe posterior distributions for each wavelength, forming the final planetary spectrum $P(\lambda)$ shown in Fig. 2b and detailed in Extended Table 2. The DALCs were modified to include wavelength-dependent limb darkening described above. Our analysis also included smoothing in wavelength with a triangle function to reduce the effect of known spectral undersampling from the WFC3 detector[1,3,8]. The full-width at half-maximum of the smoothing triangle was four columns, resulting in 32 DLCs at a spectral resolution of $R \approx 75$.

Warm Spitzer IRAC. We performed aperture photometry on the Spitzer images (Table 1), after subtracting a background value determined as a median of the pixel values well away from the stellar image. We applied a newly developed base vector analysis to the subsequent photometry to decorrelate the well-known Spitzer intra-pixel effect[16,40–44]. We tested this decorrelation method on both published and unpublished Spitzer exoplanet eclipses, and found that it consistently reduces both the BIC and red-noise by comparison with previous methods[40–44]. Our base vector algorithm fits linear coefficients to the nine pixel values centred at and surrounding the stellar point spread function[16] over time:

$$\mathrm{Flux}(t) = a^{-1}(t)\sum_n a_n \mathrm{pixel}_n(t), \quad n = \{1, \ldots, 9\}$$

where $a(t)$ is the normalization for each 3×3-pixel box as a function of time. Our Spitzer fitting simultaneously solved for the transit depth and amplitudes of starspot crossings, by scaling the more precise (and strictly simultaneous) Kepler spot crossings as described below.

Kepler archival transits. We used the Kepler data for two completely independent purposes. The first purpose was to derive an improved optical transit depth and geometric parameters. The second purpose was to characterize the nature of starspots on HAT-P-11.

Improving the optical transit depth. To improve the optical transit depth, we improved on the previous analysis[12], using all four years of Kepler data Q0-Q16, incorporating fourth-order limb darkening in the analysis. We detrended the stellar variations by fitting linear trends to the normalized fluxes within one transit duration of first and fourth contacts. After dividing by linear baselines centred at each transit, we phase-folded all 208 short-cadence Kepler archival transits into a single, very high-precision transit having 60 s time resolution. We fitted an analytic transit model[36] to this phase-folded light curve, including the LDCs described above, and to geometric parameters from radial-velocity measurements[45]. This determined both the transit depth and the geometric parameters (for example impact parameter) to high precision. All of the values were within 1σ of the original analysis[12], but now followed the same limb-darkening model as all of our data sets.

The nature of starspots on HAT-P-11. Our second analysis of the Kepler data characterized the nature of starspots that occur on the disk of HAT-P-11. We used the properties of spots crossed by the planet during transit to verify that uncrossed spots cannot mimic water absorption in the planetary spectrum. This involves determining the relative temperature (spot versus photosphere) of the spots crossed during the Spitzer transits, and also demonstrating that those spots are representative of the uncrossed spots.

To determine the relative amplitude (Spitzer versus Kepler) of the spot crossings, we first measured their profiles in the Kepler transits as the residual of the individual Kepler light curves relative to the phased and binned model. We include the Kepler spot-crossing profile as a term in the MCMC fitting[18–20] to the Spitzer transits, fitting the spot crossing amplitude simultaneously with the transit parameters. Negative ratios are statistically possible here because of the low signal-to-noise ratio of the spot crossing profiles. We allowed our MCMC chains to probe these unphysically negative values of spot crossing amplitudes. In Fig. 1b, the grey region represents the noise-dominated regime and the white region represents the physically relevant regime. Because the spot area crossed by the planet is the same for the simultaneous Spitzer and Kepler transits, the relative amplitudes can be converted to relative temperatures in Fig. 1b.

To establish that the crossed spots are typical of the uncrossed ones, we fitted a Gaussian profile to each of the 298 Kepler starspot crossing profiles—during the 208 Kepler transits observed over four years—to determine the full distribution of the spot crossing amplitudes (Extended Data Fig. 5), as discussed below.

Constraining the significance of water vapour in starspots. The amplitude of the starspot effect that mimics exoplanetary absorption is given as $f\varepsilon\delta$ (ref. 8), where f is the fractional coverage of spots on the stellar disk, ε is the depth of the transit and δ is the water vapour absorption line depth (relative to the spectral continuum) in the spatially resolved starspot spectrum, at the observed spectral resolution. Because the planet's orbit is not synchronous with the star's rotation, spot crossings in the Kepler data were used to estimate a flux deficit of 0.0179 (ref. 12). Accounting for the (small) optical intensity of dark starspots lead to $f = 0.02$. $\varepsilon = 0.0036$ is measured directly by our WFC3 photometry (Fig. 1a). Because we cannot obtain spatially resolved spectra of the starspots on HAT-P-11, we here estimate δ from Phoenix model stellar atmospheres[22], using a relative temperature constraint for the starspots based on Fig. 1b.

To establish that the spots occulted in our simultaneous Kepler and Spitzer photometry are typical of HAT-P-11, we plotted the distribution of spot crossing amplitudes (Extended Data Fig. 5) over the entire set of Kepler photometry for HAT-P-11 (Extended Data Fig. 6).

Spots that cover the disk of the star cause a brightness variation as the star rotates. This variation has a peak-to-peak modulation of approximately 2%, consistent with the spot coverage inferred from our previous Kepler study[12]. The times of our Spitzer and WFC3 observations are indicated in Extended Data Fig. 6 by blue and red lines, respectively. Similarly, the spot crossing amplitudes during our Spitzer observations are identified in Extended Data Fig. 5 by the coloured, dashed lines. From this, we concluded that the total effect of spots on the disk of HAT-P-11 is approximately the same during our observations as during other times, and that spots crossed during our Spitzer observations are typical of the unocculted spots during our WFC3 observations.

Next we must determine δ, the depth of the 1.4 μm water absorption feature in the spectrum of starspots. We approximate the starspot spectrum as equivalent to a star of the same abundance and surface gravity as HAT-P-11, but having a lower temperature. We examined Phoenix model atmospheres that are enriched in oxygen by +0.3 in the log of abundance[22], and we convolved their spectra to the resolution of WFC3.

Even in the extreme case with a temperature contrast of 1,800 K (Fig. 1b), the Phoenix spectrum shows that $\delta \approx 0.024$ at 1.4 μm, yielding $f\varepsilon\delta < 2$ p.p.m., which is two orders of magnitudes less than the absorption we derived for the exoplanetary atmosphere. Because f and ε are small, there is essentially no starspot temperature that can produce sufficiently strong water absorption to mimic our result, given our inferred values for f and ε.

Self-consistent atmospheric retrieval for exoplanets. We interpret the observed transmission spectrum using a new variant of our atmospheric retrieval framework described in previous studies[23,24]. The new SCARLET framework combines a self-consistent, line-by-line atmospheric forward model with the nested-sampling technique to efficiently compute the joint posterior probability distribution of the atmospheric parameters. We probe the multidimensional parameter space spanned by the metallicity (that is, the overall abundance of heavy elements), the C/O ratio, the cloud-top pressure, the planetary radius at the 1 bar level and the planetary Bond albedo.

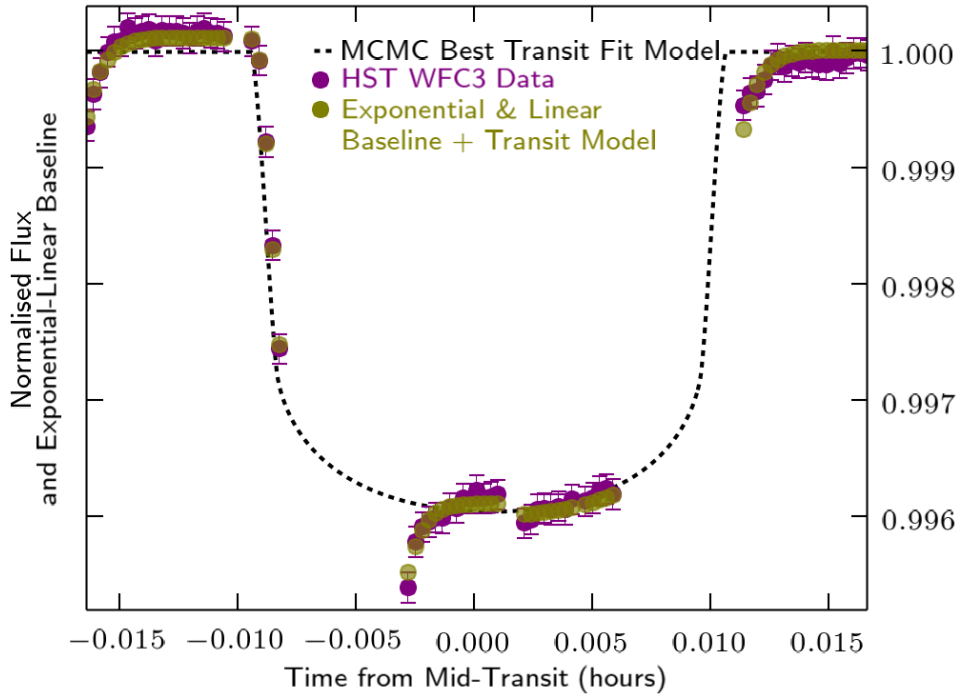
For a given set of parameters, the atmospheric forward model self-consistently computes the molecular abundances in chemical equilibrium and the temperature pressure profile in radiative–convective equilibrium. Line-by-line radiative transfer based on pre-calculated opacity look-up tables enables us to model molecular absorption accurately for the entire range of compositions. Rayleigh scattering is included using the two-stream approximation. In this study, we included clouds as a grey opacity source that cuts off the transmission of starlight below the parameterized cloud-top pressure.

We included the planetary Bond albedo as a free parameter to capture the uncertainty in the atmospheric composition introduced by the unknown albedo. For a given atmospheric composition, the Bond albedo introduced the dominant uncertainty in the planetary temperature profile, which (via the scale height) affects the relation between observed depths of the absorption features and the mean molecular mass.

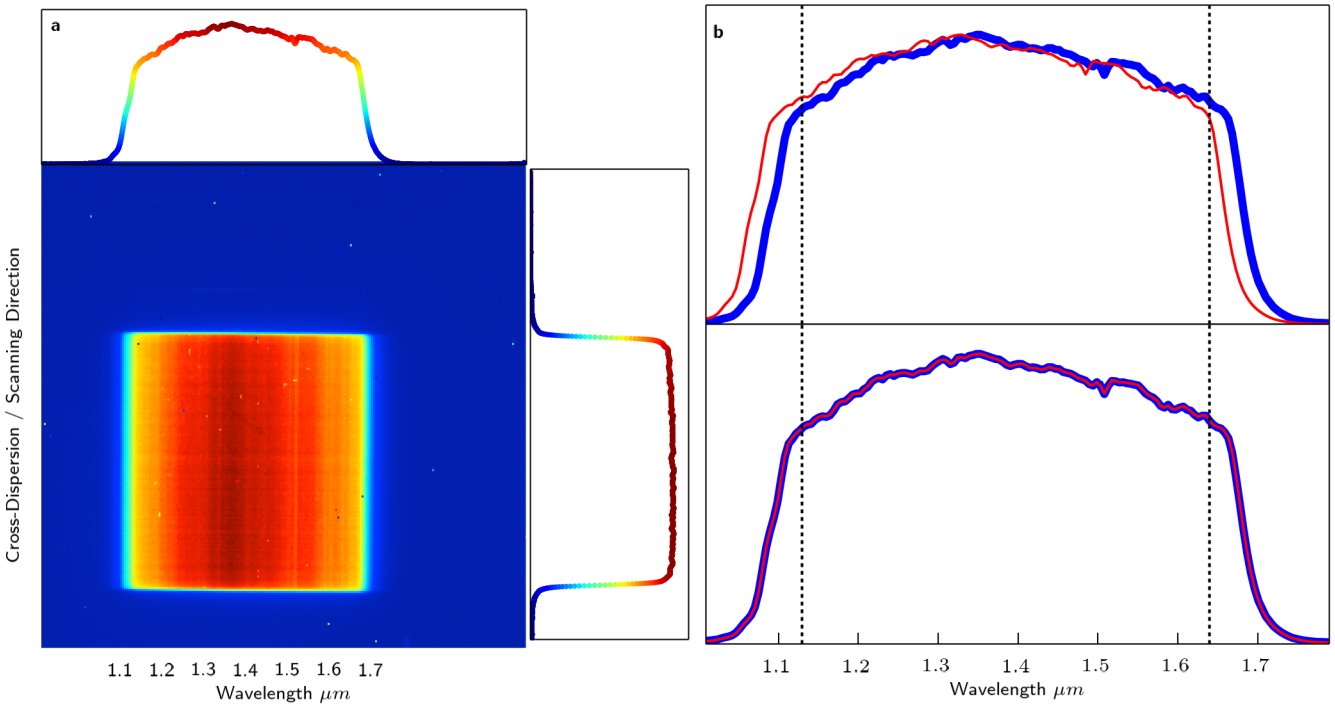
The nested-sampling algorithm repeatedly invoked the atmospheric forward model to probe the agreement between model spectra and the observational data throughout the multidimensional parameter space. In total, several 10^4 self-consistent, line-by-line atmospheric models were computed. The algorithm was initiated by randomly sampling 1,000 active samples within the full multidimensional parameter space. The active samples then iteratively migrated towards the regions of high likelihood by replacing the lowest-likelihood active sample, that is, the worst fit to the observations, with a new, better-fitting random sample[24]. Convergence was obtained once the logarithm of the Bayesian evidence, Z, computed from the active sample no longer changed by more than $\Delta(\log(Z)) = 0.0001$. The algorithm is robust to multimodal posterior distributions and highly elongated curving degeneracies frequently encountered in exoplanet atmospheric retrieval studies[24].

31. McCullough, P. M. & MacKenty, J. *Considerations for Using Spatial Scans with WFC3*. Instrum. Sci. Rep. WFC3 2012–8 (Space Telescope Science Institute, 2012).
32. Wilkins, A. *et al.* The emergent 1.1-1.7 μm spectrum of the exoplanet Corot-2b as measured using the Hubble Space Telescope. *Astrophys. J.* **783,** 113 (2014).
33. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 \le \log[M/H] \le +1$, $2000\,K \le T_{eff} \le 50000\,K$ at several surface gravities. *Astron. Astrophys.* **363,** 1081–1190 (2000).
34. Oliphant, T. Python for scientific computing. *Comput. Sci. Eng.* **9,** 10–20 (2007).
35. Rajan, A. *et al. WFC3 Data Handbook 2011* (Space Telescope Science Institute, 2011).

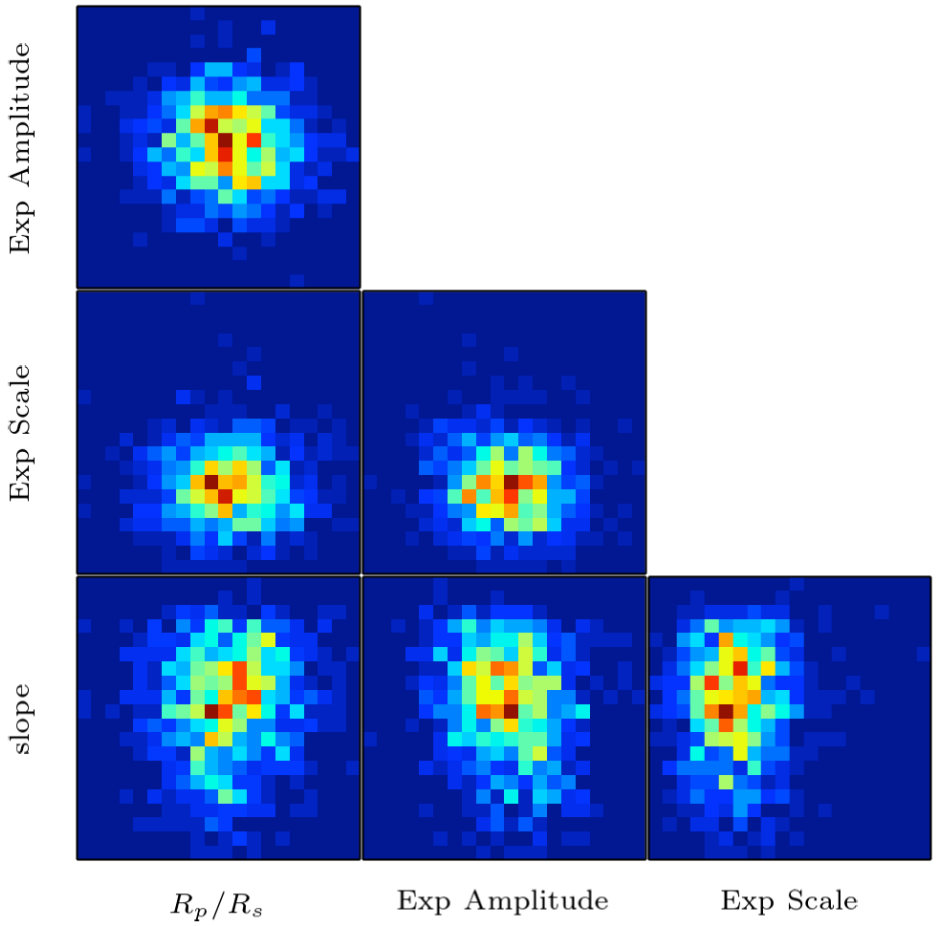
36. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580,** L171 (2002).
37. Mandell, A. *et al.* Exoplanet transit spectroscopy using WFC3: Wasp-12 b, Wasp-17 b, and Wasp-19 b. *Astrophys. J.* **779,** 128 (2013).
38. Kass, R. & Raftery, A. Bayes factors. *J. Am. Stat. Assoc.* **90,** 773–795 (1995).
39. Carter, J. A. & Winn, J. N. Parameter estimation from time-series data with correlated errors: a wavelet-based method and its application to transit light curves. *Astrophys. J.* **704,** 51–67 (2009).
40. Lewis, N. *et al.* Orbital phase variations of the eccentric giant planet HAT-P-2b. *Astrophys. J.* **766,** 95 (2013).
41. Knutson, A. *et al.* The 3.6-8.0 μm broadband emission spectrum of HD 209458b: evidence for an atmospheric temperature inversion. *Astrophys. J.* **673,** 526–531 (2008).
42. Ballard, S. *et al.* A search for a sub-Earth-sized companion to GJ 436 and a novel method to calibrate warm Spitzer IRAC observations. *Publ. Astron. Soc. Pacif.* **122,** 1341–1352 (2010).
43. Knutson, H. *et al.* 3.6 and 4.5 μm phase curves and evidence for non-equilibrium chemistry in the atmosphere of extrasolar planet HD 189733b. *Astrophys. J.* **754,** 22 (2012).
44. Todorov, K. *et al.* Warm Spitzer observations of three hot exoplanets: XO-4b, HAT-P-6b, and HAT-P-8b. *Astrophys. J.* **746,** 111 (2012).
45. Eastman, J., Siverd, R. & Gaudi, S. Achieving better than 1 minute accuracy in the heliocentric and barycentric Julian dates. *Publ. Astron. Soc. Pacif.* **122,** 935–946 (2010).
46. Knutson, H. *et al.* Friends of hot Jupiters. I. A radial velocity search for massive, long-period companions to close-in gas giant planets. *Astrophys. J.* **785,** 126 (2014).



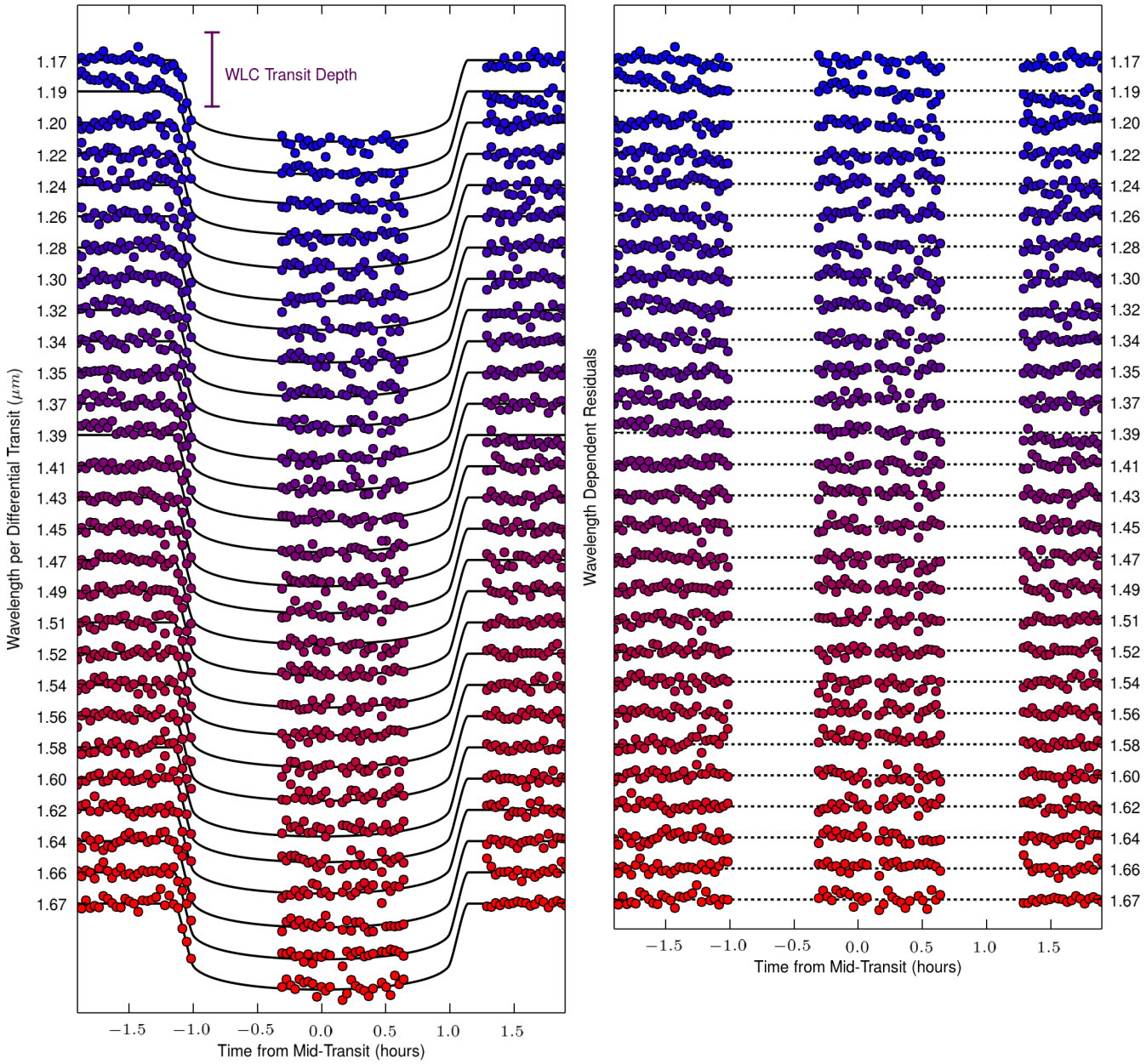
Extended Data Figure 1 | HST white-light curve with exponential ramp effects. The gaps resulted when HAT-P-11 was occulted by the Earth during Hubble's ~96 min orbit. We decorrelated the ramp effect by fitting an average, two-parameter (scale and amplitude) exponential function over time.



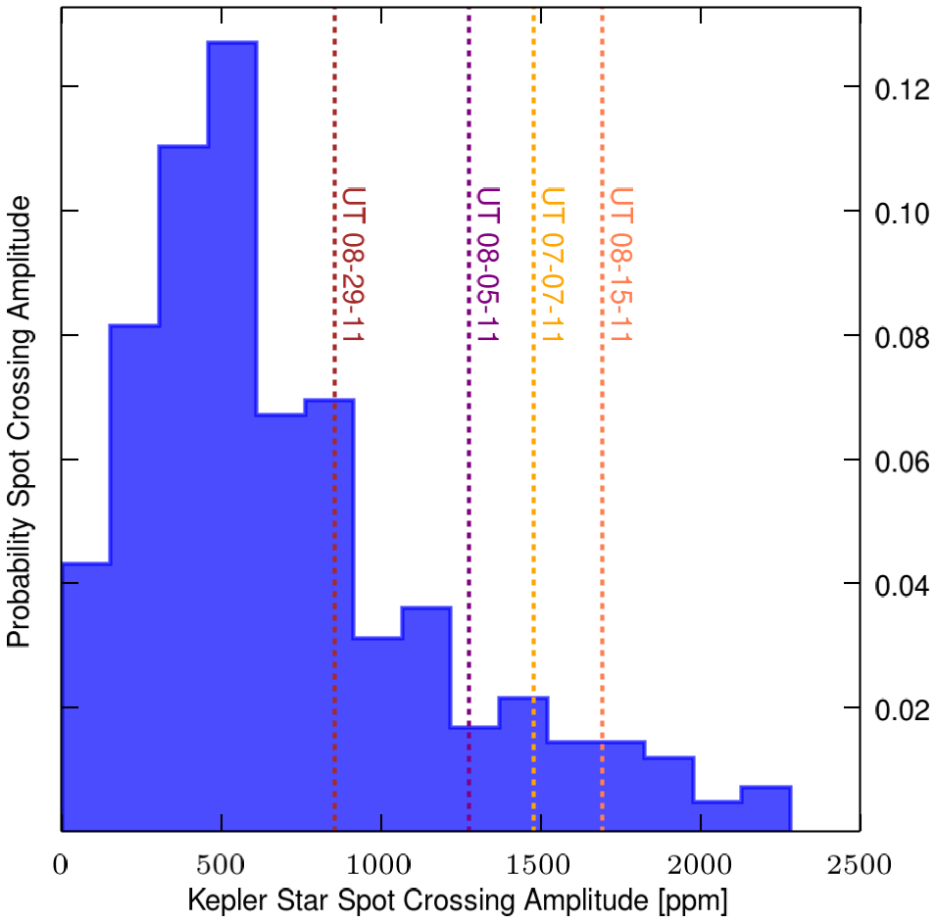
Extended Data Figure 2 | An example of WFC3 scanning-mode observation spectral images. **a**, Example spatial scan spectral image with the normalized summations in the dispersion (upper) and cross-dispersion or scanning (right) directions. **b**, Integrated spectrum (blue) and spectral template (red) before (top) and after (bottom) fitting; the amplitudes and colours are normalized to 1.0.



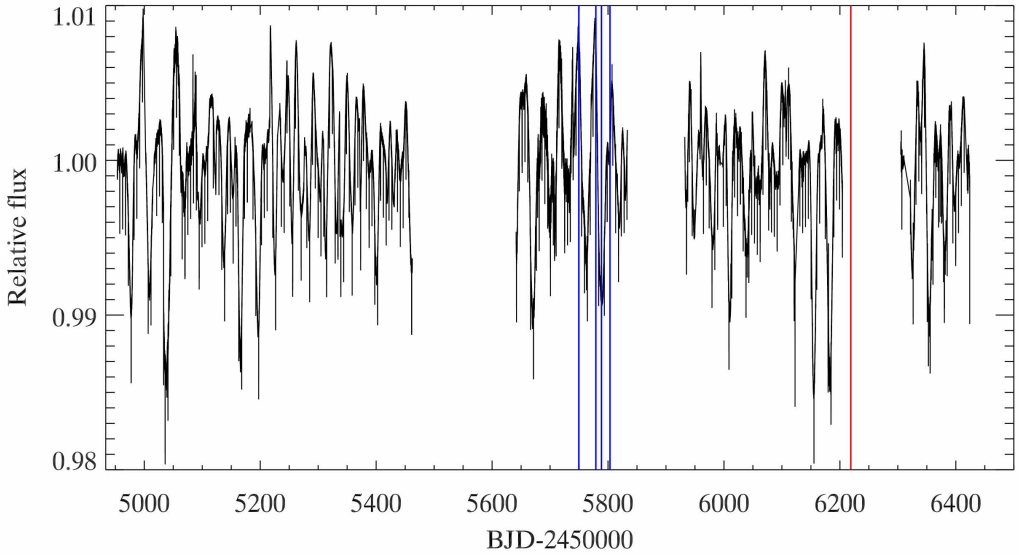
Extended Data Figure 3 | Correlations between all fitted parameters for our HST WFC3 white-light curve. We calculated the Pearson correlation coefficient over the posteriors of each parameter, and found the correlations to be small (<0.10 in magnitude), or in most cases negligible (<0.01 in magnitude). Blue represents regions of lesser posterior density and red represents regions of greater posterior density, with green and yellow in the middle.



Extended Data Figure 4 | Wavelength-dependent transit light curves. The coloured points are the wavelength light curves, ranging from blue (1.17 μm) to red (1.67 μm) with 18 nm spacing. The black lines represent the best-fit transit light curves over the wavelength range from 1.1 to 1.7 μm. The curves are shifted for display purposes only. The differential light curves were fitted with differential analytic transit curves to derive the planetary spectrum seen in Fig. 1. We added the white-light curve into the differential light curves to derive the data above.



Extended Data Figure 5 | The distribution of Kepler starspot crossing anomalies. We fitted a Gaussian profile to each of the 298 spot crossings seen during the 208 transits observed by Kepler. Here we show the distribution of starspot amplitudes, calculated as the height minus baseline of the fitted Gaussian profile in parts per million. The dashed lines represent the starspot crossing amplitudes observed during our four concurrent Spitzer observations. In particular, note that all four spot crossings with concurrent Spitzer observations are at the larger end of the distribution. In addition, the spot crossing on UT 7 July 2011, the largest spot crossing feature observed during our concurrent Spitzer observations, crossed a spot with $\Delta T \approx 900$ K.



Extended Data Figure 6 | HAT-P-11 Kepler light curve for ~4 yr of short cadence. The variation in flux has a peak-to-peak modulation of ~2%, consistent with the spot coverage inferred from our previous Kepler study[12]. The times of our Spitzer observations are marked with vertical blue lines. The time of our HST WFC3 observation, included in this analysis, are marked with a vertical red line.

Extended Data Table 1 | Transit depths as a function of wavelength for Kepler, HST WFC3, Spitzer IRAC1 and Spitzer IRAC2

Wavelength [microns]	Transit Depths [ppm]	Transit Depth Uncertainties [ppm]	Bandpass- / Bin- width [microns]
0.642*	3365	8	0.2415
1.153	3502	40	0.0094
1.172	3407	47	0.0094
1.190	3421	46	0.0094
1.209	3445	38	0.0094
1.228	3350	41	0.0094
1.247	3377	38	0.0094
1.266	3380	45	0.0094
1.284	3457	40	0.0094
1.303	3436	35	0.0094
1.322	3448	39	0.0094
1.341	3476	43	0.0094
1.360	3536	44	0.0094
1.379	3499	46	0.0094
1.397	3498	40	0.0094
1.416	3524	41	0.0094
1.435	3591	44	0.0094
1.454	3524	43	0.0094
1.473	3520	39	0.0094
1.492	3447	39	0.0094
1.510	3344	45	0.0094
1.529	3513	41	0.0094
1.548	3471	50	0.0094
1.567	3438	49	0.0094
1.586	3414	53	0.0094
1.604	3383	45	0.0094
1.623	3415	38	0.0094
1.642	3480	48	0.0094
1.661	3498	60	0.0094
1.680	3376	74	0.0094
3.521†	3384	20	0.3685
4.471†	3363	27	0.5021
3.521	3421	29	0.3685
3.521	3347	29	0.3685
4.471	3321	38	0.5021
4.471	3407	37	0.5021

* Kepler transit depth determined from all 208 phased and binned Kepler transits.
† Weighted mean of the two other independent Spitzer transits for each channel.

Extended Data Table 2 | The system and planetary parameters of HAT-P-11b

	Kepler	HST WFC3	Spitzer IRAC1 UT 07-07-11	Spitzer IRAC1 UT 08-15-11	Spitzer IRAC2 UT 08-05-11	Spitzer IRAC2 UT 08-29-11
Wavelength (μm)	0.641	1.419	3.521	3.521	4.471	4.471
Period (days)*	4.8878018 ± 7.1 x 10^{-6}	--	--	--	--	--
T_{center} (HJD_{UTC})†	54811.1786727 ± 0.0000082	54811.1785254 ± 0.0000663	54811.1791713 ± 0.0001272	54811.1788248 ± 0.0001239	54811.1790119 ± 0.0001541	54811.1789256 ± 0.0001591
Inclination (°)‡	89.549 ± 0.114	--	--	--	--	--
b^c (impact parameter)	0.135 ± 0.034	--	--	--	--	--
a / R_s‡	17.125 ± 0.060	--	--	--	--	--
R_p / R_s§	0.05852 ± 0.00007	0.05887 ± 0.00025	0.05849 ± 0.00025	0.05785 ± 0.00033	0.05762 ± 0.00033	0.05837 ± 0.00032
Transit Depth‡ (ppm)	3424 ± 8	3466 ± 29	3421 ± 29	3347 ± 29	3321 ± 38	3407 ± 37
c1‖	0.7547 ± --	0.6718 ± --	0.5750 ± --	0.5750 ± --	0.6094 ± --	0.6094 ± --
c2‖	-0.9164 ± --	-0.1618 ± --	-0.3825 ± --	-0.3825 ± --	-0.7325 ± --	-0.7325 ± --
c3‖	1.6411 ± --	0.2855 ± --	0.3112 ± --	0.3112 ± --	0.7237 ± --	0.7237 ± --
c4‖	-0.6328 ± --	-0.1551 ± --	-0.1099 ± --	-0.1099 ± --	-0.2666 ± --	-0.2666 ± --
Eccentricity¶	0.232 ± 0.054	--	--	--	--	--
Longitude of Periastron¶	7° ± 25°	--	--	--	--	--

A dash indicates that the value was held constant for all MCMC chains.
* Ref. 12.
† Ref. 45.
‡ Derived from MCMC posteriors over our phased and binned Kepler transit.
§ Transit depths are uncorrected for stellar activity.
‖ Derived from ATLAS models.
¶ Ref. 46.