Peer Review File

Manuscript Title: A Highly Magnified Star at Redshift 6.2

Editorial Notes:

Redactions – unpublished data

Parts of this Peer Review File have been redacted as indicated to maintain the confidentiality of unpublished data.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

A. Summary of the Key Results: "An Extremely Magnified Individual Star Observed in the First Billion Years" announces the discovery of a star (or small system of stars) at z = 6.2 via strong gravitational lensing.

B. Originality and Significance: Although cluster lenses have previously been been used as 'cosmic telescopes' to detect galaxies and SNe at high redshifts, and now individual stars at lower redshifts, the detection of a single star at the epoch of reionization in the first billion years of the Universe, if confirmed, would be a spectacular discovery and merit publication in Nature.

C. Data and Methodology: Validity of Approach, Quality of Data, Quality of Presentation: This result relies on HST RELICS survey data, which is solid. A major limitation is 5 sigma detection of the star in just the F110W filter, so the mass, temperature, and metallicity of the star cannot be determined. The authors need to clarify some issues with their methods as explained below and perhaps correct some of them.

D. Appropriate Use of Statistics and Treatment of Uncertainties: I have some questions about the authors' approach as discussed below.

E. Conclusions: Robustness, Validity, Reliability: Although HST photometry prevents the authors from pinpointing the properties of the star(s), the small angular diameter of the lensed object, suggesting it is just one or a few stars, is too compelling to ignore and outweighs limitations imposed by photometry. They applied a disciplined approach to ruling out other interpretations of the object. Also, models for cluster lenses are subject to well-known discrepancies even for a given cluster, so the fact that four distinct models produced similar constraints on the properties of the star reinforces the authors' conclusions. I think this result will hold up in the long run. Furthermore, the authors have successfully made the case for this object being a z = 6.2 star to the JWST allocation committee and have secured telescope time for spectroscopic followup.

F. Suggested Improvements: Experiments, Data for Possible Revision

See below.

G. References: Appropriate Credit to Previous Work?

Sufficient, no significant omissions.

H. Clarity and Context: Lucidity of Abstract/Summary, Appropriateness of Abstract, Introduction and Conclusions

Overall, the paper is well written although considerable work will be required to properly partition it into Nature format as a Letter, Supplement, and perhaps Extended Data sections that dovetail well without repetition. The language of the paper in some places needs to be made more professional by elimination of terms such as 'incredibly' and 'superlative' (in other words, no editorializing).

I therefore would be happy to reconsider this result for publication after the authors address my points below.

----

(0) Minor general comment: the authors will have to adjust most of their figures so that just the first word of each axis label is capitalized to conform with Nature format.

Title

(1) Please change it to something more efficient along the lines of "An extremely magnified star at redshift z = 6.2". Sometimes less is more.

Abstract

(2) It's initially confusing to classify this as a massive (> 50 Msun) star and then quote temperatures ranging from 8000 - 60000 K. Please clarify that the low-temperature limit corresponds to a post-main sequence star by saying something like "... is consistent with very massive (> 50 Msun) star on the main sequence with temperatures as high as 60000 K or on the post-main sequence with temperatures as low as 8000 K". Note that 8000 K is not usually considered to be a 'hot' star, please revise your language.

(3) Later in your paper you adopt a metallicity of 0.05 Zsun for the star based on analysis that is reasonable given the information present. Please state that you find a likely metallicity of 0.05 Msun based on the properties of its host galaxy.

(4) Please omit the point about this discovery giving the opportunity to study a low-metallicity star. It's problematic for two reasons. First, you don't really know the metallicity of the star, it could be 0.001 - 1 Zsun based on observations of metals in other high-z galaxies and a number of numerical simulations. Second, some would disagree that there are few low-metallicity stars in the local universe. You won't have any real chance of constraining the metallicity of the star without NIRspec measurements. This statement can be removed without loss of impact -- the importance of discovering a single star at z = 6.2 speaks for itself.

Introduction

(5) para 3: please define CLASH.

Results

(6) para 1: Although it's commonly understood by the lensing community, please reword "arc at z = 6.2..." to something like "arc of a lensed background galaxy at z = 6.2 ..." The arc is not at z = 6.2.

(7) 2.1, first sentence: You state "Earendel was identified as having an extreme magnification in a model-independent way by the symmetry of two adjacent bright clumps.." How? It's not clear how you flagged this object to be extremely magnified. Did you designate some minimum magnification as 'extreme'? Please explain.

(8) bottom of p.3: Please describe in more detail how you carried out your MCMC campaign, either here or on p.14. For example, what is varied between runs?

(9) Table 1: you state uncertainties for each quantity on the bottom line, do they apply for all the entries in a given column? I would expect them to vary with lens model. Please clarify and attach specific errors to each entry in the table if they vary.

(10) Is it possible to dial down the range of magnifications in your

models any further? A factor of 40 is quite large.

(11) section 2.3, para 1: Do you take into account absorption of flux from the star by dust in the galaxy? One of the authors, Keren Sharon, is an expert on corrections to SN Ia spectra due to local extinction by their host galaxies so you should be able to elaborate. Also, at z = 6.2 one would expect some Lya scattering due to residual neutral H in the IGM, was this folded into your analyses? If not, both will need to be done if they are required.

(12) section 2.3, para 2: You state that green shaded regions indicate uncertainties in lensing of a factor or 7, but you state earlier this uncertainty is a factor of 40 -- please clarify.

(13) section 2.3, para 3 / Figure 3: You quote 8000 K as a lower limit but that is not the final temperature that any of your stellar models reach in the post-main sequence phase, which appears to be 4000 - 5000 K. The transition from high MS temperatures to low post-MS temperatures happens fairly quickly, so it seems much less likely that the star would happen to be found below 20000 K. It therefore seems that the high temperatures are really all that matter. Do you factor the short time it takes for these stars to evolve to your lower limit in temperature into your analyses and Figure 3? Please clarify, I could not find any discussion of this in your text.

(14) section 2.3, para 6/7: Not to be unkind, but you will have to tone down your claim that Earendel presents an opportunity to study a

low-metallicity star because you have no real idea of its metallicity, just one based on a reasonable model. In reality, observations and numerical simulations have shown that stars in z ~ 6 galaxies can have metallicities ranging from 0.001 - 1 Zsun (and even have supersolar metallicities at z ~ 7). You can omit this discussion without any loss of impact.

(15) section 2.3: Before this result can be published you have to take a good look at how Figure 3 would change with the metallicity of the star. In principle, this would require running additional grids of stellar evolution models at Z = 0.001, 0.1 and 1 Zsun and pairing them with atmosphere models to determine spectra for low-temperature stars. However, I'm willing to be flexible on how you address this point as long as it's convincing -- for example, considering just stars at high temperatures or using precomputed stellar evolution models from open databases like Heger et al. 2010.

(16) section 2.5, para 2: You say this star formed recently but later state that it formed 900 Myr after the BB. I know what you are trying to say but instead just state that the lifetime of the lensed star is much shorter than that of its host galaxy. I think this was what you intended.

(17) section 2.5, para 4: Please briefly discuss the timescales on which redshifted flux from stellar mass BHs would vary and compare them to the cadence in RELICS. Can you rule out a BH based on the fact that its flux should be transient?

Referee #2 (Remarks to the Author):

I first reply briefly to the questions as listed on the review form,

then provide detailed comments by section and paragraph.

[A] Summary of the key results

The authors state to have observed an object that could be an object at redshift z~6. They cannot constrain much of its properties but state that it might be hot and massive. They cannot even constraint whether it is a single or multiple star, or a star at all.

[B] Originality and significance: if not novel, please include reference

Finding a single star at such high redshift is, to my knowledge, would be novel, and an observational 'jackpot', if confirmed.

They do not provide any probability estimates, so significance, in a statistical sense, of the result is zero.

[C] Data & methodology: validity of approach, quality of data, quality of presentation

There is insufficient t data for useful conclusions.

[D] Appropriate use of statistics and treatment of uncertainties

A prime example of how not to do statistical analysis with model comparison.

The text is threaded with qualitative statements such as "incredibly small" w/o quantification. Inadequate for scientific paper.

[E] Conclusions: robustness, validity, reliability

There is no scientifically useful conclusions based on the available data and theoretical models.

[F] Suggested improvements: experiments, data for possible revision

I think it is unfair to your colleagues to send such a paper for review.

[G] References: appropriate credit to previous work?

References on reviews of stellar evolution and outcomes of the stars under consideration are thin.

[H] Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions

Abstract and summary are flawed.

Pointless speculation that are wrong (summary) and/or uninteresting (abstract).

Abstract

Naming of stars is sole authority of IAU.

Disregarding this is not acceptable.

Please use the standard way to show error bars for upper and lower limit for radius. Provide centre values plus one-sigma errors (could be asymmetric).

The magnification factor is uncertain by factor 40!

The star temperature of 8kK to 60kK is ridiculously uncertain, with unclear spectral type.

Being consistent with a hot massive star does not imply it is inconsistent with a cold faint star.

The star being low Z is pure speculation.

On top, this is what would be expected in the first place, so it is not even an interesting speculation.

The statement that low-metallicity stars cannot be found locally is incorrect, e.g., in SMC and LMC have low metallicty massive stars, and Zw 18 has low-metallicty star-forming region.

https://www.aanda.org/articles/aa/full_html/2013/05/aa20948-12/aa20948-12.html

Introduction

P1. What is the maximum angular size of infinitesimal lensed objects? It depends on how the product of magnified region size and magnification factor scales, and is likely a lens property and can be solved using a modest amount of maths. Please clarify.

P2. Do cluster crossing events 'live'? Is that the correct expression?

P3. What is "CLASH" (please define all abbreviations used, this is customary in astronomy literature).

2 RESULTS

Similar to abstract, naming of astronomical objects is sole IAU authority. Hence any reference to fantasy names "Sunrise Arc" and "Earendel" need to be removed.

The statement of being a high-z arc is at least wrong, if not misleading. Arcs are apparent phenomena at the observers' side - there is no are at high redshift; please represent astronomy correctly.

2.1

P1. "extremely high magnification" - please quantify

P1. "a star sitting incredibly close" - sitting stars? Please quantify "incredibly"

P3. MCMC model - From the information provided no one can deduce whether this statement is of any significance because it is absolutely unclear what has been done. The statement is suggestive but the dependable information is zero.

Figure 1. "LTM" is used in figure caption before it appeases in the text. Please define in caption.

Table 1. Please show the specific errors for each case in the table with the values rather than providing broad range multipliers below and in the caption. (common notation)

P4. "microlensing suppression would more likely be a transient phenomenon. Therefore we conclude the two lensed images are most likely unresolved in the current HST WFC3/IR imaging." Please quantify.

P4. "This is consistent with our original lens model-independent interpretation suggesting it is directly on the critical curve." But is it inconsistent with other things?

Last par. "ostensibly up to infinity" (see note above) Please study mathematics and astrophysics of lensing. A star will be an object of finite size and be a at finite distance, the maximum brightness you can achieve is to have the entire 4 pi steradians - the entire sky - shine at that surface temperature. This would be very bright, but definitively less than infinite magnification. As you go to increasingly larger magnifications, the area of the source is increasingly smaller, at some point it is smaller than the object. At this point magnification has no more effect on brightness for the same image size (Liouville's theorem). Finite size of the lens plays a role in the size of the image you can get.

2.2

P2. "Our largest radius constraint is a factor of two smaller than this superlative star cluster, while our tightest constraint is nearly an order of magnitude smaller." - Why do you have different constraints? Would it not be most statistically significant to combine the different constraints, even if this results in a "systematic" uncertainty?

P3. It is rather uncommon to define massive stars using a lower mass limit of 15 Msun. Please consult common textbooks or reviews.

P4. The binary part is pure speculation with no observational consequence.

P4. Also the possibility of unrelated projected binaries or multiples is not discussed.

2.3

P1. This remains rather unclear what has been done here. I can only speculate that the flux in F110W has been used, based on atmosphere models to derive a luminosity for a given surface temperature. If that is the case, please say so.

P1. How about extinction due to dust or absorption beyond Ly break?

P2. Where is the flux uncertainty represented, it seems missing. Certainly it remains unmentioned.

P2. You use [M/H]=-1 atmosphere models, but the stars shown are 0.05 Zsun = [M/H]=-1.3. Is it really necessary to dial together things apparently randomly?

P3. How do you estimate this? Is there any mathematical basis? Or do you speculate rather than estimate (which can be a scientific method)?

P3. There is no 500 Msun star in the plot. How do you get to 500 Msun?

Figure 3. Such HRDs are not new, such track have been available since 30 yr. Additionally, there is the question of the Humphrey-Davidson limit (

https://en.wikipedia.org/wiki/Eddington_luminosity#Humphreys%E2%80%93Davidson_limit)

Figure 3. This is again very suggestive, but as seen in the plots, it appears as if there was a flat prior for the different star masses and evolution stages. But this is not so. Maybe you were misguided by the plot showing circles where they had models - potentially based on varying-size time steps or data files, and sizes of the circles potentially related to star radius rather than likelihood of finding a star there.

Figure 3. A proper analysis would require folding the probability of a star of that mass in that evolution stage based on some IMF assumption. This is standard Bayesian analysis commonly used in current scientific literature. Standard practise.

P3. The constraints are so open, they are not really useful.

P4. This is mere speculation, not results. What is the take-out here? Looking at Figure 3, I can see that a track of ~20 or 25 Msun would intersect with the light green shaded region of shifter up by 0.3 or 0.5 dex? If that is what the authors refer to, it is completely wrong. See section above about Bayesian analysis. for one, the folding needs to be done, and for second, those stars spend a very small fraction of their lifetime during that phase. This is on top of the metallicity of the stellar grid and that of the observational data shown differ.

P6. The star is so distant, it can't be really studied in any meaningful way with these observations: You do not know metallicity. You estimate temperature up to about 1 dex uncertainty. You do not know multiplicity of the star by 0.6 dex (factor 4). SO you are really trying to tell the reader you can learn something about evolution of low-metallicity stars form this? This is an insult to the readers' intellect.

P6. Whereas binary BH may be formed by close binaries at low Z, it will still be only a small fraction of them that make such binary BH systems that lead to GW events. There chance that the system you observed is one of those few select is rather negligibly small.

P7. Low Z. See comment on abstract.

P8. So there is hope for data in the future of which the present paper is devoid.

2.4

(I cannot comment on this)

2.5

P1. Various grids of Pop III and other low-Z stars exits already.

Again, a detailed Bayesian analysis would need to be done for meaningful results.

P1. A statement of the kind "X is consistent with ... " has not merit as it does not constrain what is not consistent. Not having a bank statement to show for it is consistent with being a billionaire.

P2. Formed recently? Later it is stated it formed 900 Myr after the Big Bang.

P2. Whereas lifetimes are short what would be the effect of apparent time dilation of lifetime by a factor 7 due to high red shift?

P2/P3. There is a history dating at last 15 years back that Pop III stars and their IMF are not all of a kind. It also depends on the environment and the ionisation stage of the gas. For example, from early on, after the first SN go off, cosmic rays will propagate, affecting the ionisation stage of the gas and potentially enriching the gas slightly. Then there is the Lyman and Lyman-Werner radiation from the surrounding stars. Finally, an enrichment of [Z]=-4 or less will affect the IMF; a metallicty of [Z]=-8 or less will affect the evolution of star so that they already differ from truly primordial stars - such star may not be distinguishable from metal-fee stars by their lines bit still evolve differently. They are not Pop III stars. The present discussion on Pop III stars add no value.

P4. X-rays would also be red-shifted. What would be the variability of such an accreting BH? Also, 200 Msun may not be a likely mass due to the pair-SN mass gap.

P5. "the incredibly small area" - quantity. It is probably not an area (unit cm^2) but a solid angle.

P6. "it would be highly unlikely" -quantify.

3.

--

P1. "observed 900 Myr after the Big Bang" - I though it was a recent observation.

P1. This also shows there basically very little constraints, hardly any dependable data.

P1. "Likely" properties of the star are not supported by manuscript.

P2. Speculations about the future only.

P3. "Very massive low-metallicity stars are extremely rare, and thus few are known even locally." There is 3 things in this sentence that require quantification.

P3. very massive stars refer to those above 100 Msun which make pulsation pair SN supernovae, or sometimes those that make pair SNe and leave no black hole.

P3. Stars that may indeed enter the giant branch may be too far separated to become the binary BH of LIGO.

P3. On top, this is confluence of opposing speculations made in the manuscript. It is stated the star may be massive with > 40 Msun if single, or 25 Msun if binary or maybe 20 if multiple. By all good faith, beyond the low probability a ransom massive binary system making a binary BH that merges as a LIGO source, a binary of two 25 Msun stars will not result in a binary of two 30 Msun BHs.

4.

--

I will not comment on methods section.

Additional concerns and feedback

================================

Critical

The naming of celestial objects and stars would be the sole authority of the International Astronomical Union (IAU), and in the interest of science and standards, neither the authors nor nature should bypass this.

Minor

Typesetting has various minor flaws and should be improved for consistency and legibility. (Whereas I assume some of this will be rescued by nature editing, it is always better to start with something done carefully.)

The authors may want to look up the difference between indices and subscripts for variables.

Spacing between numbers and units is not applied consistently.

Various commas are missing.

Referee #3 (Remarks to the Author):

I have read the paper entitled "An Extremely Magnified Individual Star Observed in the First Billion Years" by Brian Welch et al with great interest and enjoyment, and would like to congratulate the authors on their discovery and a very nice paper. To summarise, the paper itself presents the discovery of a lensed star (system) from HST RELICS data that likely resides at high redshift, and performs detailed analyses in order to infer the source's nature as a single star or star cluster and places the result in context by aiming to establish the star's mass and metallicity. Although the data themselves are insufficient to rule out a binary star system, the discovery is one of the highest-redshift discoveries of single star system and provides a useful benchmark for future studies and follow up observations. The methodologies adopted by the authors are appropriate and thorough, with clear explanations and presentation of analyses and results. While I am convinced that this object is likely a single star (system), I would have several revisions/requests I would like the authors to perform before recommending the paper for publication:

Section 2.1:

1. Figure 1: What filters are used to create the image? Could the authors please add small circles to each component/clump of the systems used in the lens modelling (perhaps with different according to each system so they can clearly be identified), in similar fashion to the interloper? Apart from 1.1a and 1.1b, the other clumps are faint enough that they are challenging to identify in the image. Additionally, to what significance are each of the clumps (including the star) used in the lens modelling detected and in what filter?

2. For the remaining 20% of the models where Earendel does *not* fall within D_Crit<0.1", at what D_crit does the star fall?

3. Is there a particular reason why the LTM and Glafic (c=1) results are largely different from the others?

4. Could the authors please restructure Table 1 to include the uncertainties on the individual values? At present it is difficult to evaluate the meaning on the individual values without their associated uncertaities (for instance, there's a big difference if the uncertainty on mu=8400 is x0.7 or x5.0, neither of which is explicitly associated). In particular, what are the error bars for the radius and do the tabulated upper limits include those uncertainties?

Section 2.2:

5. Similarly to point 6 above, do the upper limits to the radius values quoted here include the uncertainties? And how are those uncertainties derived?

6. In addition to referencing the smallest star clusters, please add a reference to Bouwens et al. (2017; https://ui.adsabs.harvard.edu/abs/2017arXiv171102090B/abstract) to highlight that the estimated radius of the star is also smaller than very low luminonsity galaxies at high redshift.

7. Based on the text and caption of Figure 3, it is not immediately clear to me: I assume all of the colored tracks are the stellar evolution models from BoOST? Are these all dwarf galaxy predictions with 0.05 Z_Msol or single stars? Additionally, if tracks of various metallicities are present in the

Figure (as suggested in paragraph 5), this is not clear to me. Please clarify (both in the text and the Figure caption).

Section 2.4:

8. Would the results of such a simulation change significantly if a different velocity were assumed? Are there any (reasonable) star velocities that would induce significant magnification variations not observed in the photometry, accounting for the uncertainties?

9. Presumably the the gray bands in the top right panel of Figure 4 are the 1sig uncertainties? Please clarify.

Section 2.5:

10. Assuming ~100 stars per arcmin2 in a 1 arcmin2 field centered on Earendel, what is the probability of a star landing close to Earendel's position if such a density was injected into the image e.g., 10000 times? Are there any that would land close?

11. Second-to-last paragraph: I would like the authors to expand on reasons for Earendel not being a Brown Dwarf via SED-fitting tests described for Section 4.2 below.

12. Last paragraph: similarly, I am not convinced of the reasons stated by the authors to exclude a low-z solution: the discovery of Earendel is already a fortuitous event (quoted at P(100 M_sol)~2-4% in Section 4.6) and thus it is not unreasonable to assume other low-probability events could mimick the discovery. I suggest expanding this paragraph and the one prior to include mention of the tests requested for Section 4.2 below.

Section 4.2:

While I agree with the authors that Earendel's size is the strongest constraint for it being a single star or star cluster, I find details of this section of the paper somewhat lacking in establishing the high-z nature of the galaxy and Earendel, and would like to authors to clarify some aspects and perform some additional tests to ensure that Earendel in particular is not a low-z interloper or brown dwarf.

13. How are the stacks of the arc flux and uncertainties performed? Please clarify and state in the text.

14. What redshift prior and range are the authors adopting for BPZ? For the latter, presumably 0.1<=z<=13.

15. Do the BPZ models include those of brown dwarfs (e.g., the SpecX library) and z~2 dust starburst galaxies? If they do, what are the likelihoods of (i) the best-fit brown dwarf model and (ii) a low-z interloper compared to the best-bit z~6 solution? If not, I would like the authors to please fit the SpecX brown dwarf library as well as add a model of a z~2 dusty starburst galaxy to evaluate those likelihoods.

16. What star formation history and which parameter ranges are adopted in the Bagpipes fitting, and is the redshift left as a free parameter or fixed? If the redshift is fixed, please repeat the fit allowing the same redshift range as the BPZ range and state how the z_phot and P(z) solutions compare with BPZ.

17. While the original HST photometry clearly does not have the power for any meaningful SED constraints on Earendel, the 3-year photometry might. Could the authors please repeat the BPZ and Bagpipes exercises (including the SpecX ad low-z interloper exercises) for only the 3-year photometric points of Earendel and compare likelihoods, z_phot and P(z) solutions.

18. Please could the authors tabulate the extracted photometry (and 1sig uncertainties) of each component in a Table that get used in the fitting process.

19. Have the authors checked whether Spitzer/IRAC data is available for the cluster/star? RELICS has decent Spitzer overlap (e.g., SURFS UP, PI: Bradac, Strait et al. 2021, etc)? If so, such photometry should be included in the above fitting tests/procedures and would probably be of great aid in the SED fitting.

Referee #4 (Remarks to the Author):

This paper presents the discovery of a z~6.2 star that is highly-magnified by a foreground galaxy cluster. This discovery is the first of its kind, and its position close to the lensing critical curve enables it to be visible at high magnification for an extended period of time where it can be followed-up with future observations. The authors' analysis is thorough and convincing, and the paper clearly presents the data, methodology, and reasoning behind their conclusions. I do have some minor comments on the paper that I think would help clarify and strengthen certain points. Once these are addressed, I will recommend that the paper be accepted for publication. My comments are listed below:

Section 1, line 73: It is stated that the magnified stars (the source) are moving relative to compact objects in the foreground lensing cluster, causing microlensing variations. Is this entirely (or mostly) due to the motion of the source, or does the motion of the microlenses in the cluster matter as well? I would expect the proper motion of the foreground stars in the lens galaxy to be greater, so this statement seemed somewhat counterintuitive. This also seems to contradict a statement later in the paper (Section 2.4, line 224).

Section 2.1, line 107: Unless it is computationally unfeasible, generating an MCMC chain of only 100 models seems too short to draw any conclusions. Can the authors run a longer chain, or explain further why this sample is sufficient?

Table 1 (and in the text where relevant): Do the uncertainties in the table apply to each lens model, the best (LTM) model, or is it the total uncertainty accounting for differences between the various lens models? If it's the former, I find it unlikely that the models all have the same level of uncertainty given the difference in methodology. If it accounts for differences among models, the authors should provide some effective weighted average.

Section 2.1: Are there any line-of-sight foreground objects that could potentially affect the lens model?

Section 2.2, line 144: Does the statement about the typical radii of star clusters apply even at z~6? What do theory/simulations predict?

Section 2.4 and 4.6: If Earendel is a binary or multiple-star system, would a lack of microlensing variability in the future be able to indicate this? Or would this be degenerate with the angle of motion of Earendel relative to the caustic?

Section 3, line 319: Alternatively, knowing the temperature of Earendel from JWST spectroscopy could provide a direct magnification constraint at its location in the lens model. Would this potentially supersede any lens model improvements from additional multiply-imaged systems?

Section 4.3, line 382: Rather than taking the weighted average of all four IR bands in the stack, what if you just used the two bands closest to F110W? This would make the effective wavelength closer, even if the uncertainty is larger.

Section 4.4.1 and 4.4.2: When allowing the relative weights of the six particular galaxies in the model to be freely optimized, this means that their M/L is allowed to take on any value with no prior? Do the optimized M/L of these galaxies in the best-fit model appear realistic, or are some of them significantly different from the global M/L? If you were to enforce the same M/L scaling for all galaxies including these six, how would this affect the results?

Section 4.6, line 575: Why is a Kroupa IMF assumed here when Salpeter was assumed for the arc SED fit and Salpeter and Chabrier were assumed for the stellar surface mass density calculation? I think a rough estimate of the effect of different IMFs would be useful in this section (e.g., a Chabrier IMF would increase the probability by ~X times).

Section 4.7, line 603: It would be useful to show the two rectangular regions parallel to the arc in a figure somewhere.

Author Rebuttals to Initial Comments:

Referee #1 Comments:

A. Summary of the Key Results: "An Extremely Magnified Individual Star Observed in the First Billion Years" announces the discovery of a star (or small system of stars) at z = 6.2 via strong gravitational lensing.

B. Originality and Significance: Although cluster lenses have previously been been used as 'cosmic telescopes' to detect galaxies and SNe at high redshifts, and now individual stars at lower redshifts, the detection of a single star at the epoch of reionization in the first billion years of the Universe, if confirmed, would be a spectacular discovery and merit publication in Nature.

C. Data and Methodology: Validity of Approach, Quality of Data, Quality of Presentation: This result relies on HST RELICS survey data, which is solid. A major limitation is 5 sigma detection of the star in just the F110W filter, so the mass, temperature, and metallicity of the star cannot be determined. The authors need to clarify some issues with their methods as explained below and perhaps correct some of them.

D. Appropriate Use of Statistics and Treatment of Uncertainties: I have some questions about the authors' approach as discussed below.

E. Conclusions: Robustness, Validity, Reliability: Although HST photometry prevents the authors from pinpointing the properties of the star(s), the small angular diameter of the lensed object, suggesting it is just one or a few stars, is too compelling to ignore and outweighs limitations imposed by photometry. They applied a disciplined approach to ruling out other interpretations of the object. Also, models for cluster lenses are subject to well-known discrepancies even for a given cluster, so the fact that four distinct models produced similar constraints on the properties of the star reinforces the authors' conclusions. I think this result will hold up in the long run. Furthermore, the authors have successfully made the case for this object being a z = 6.2 star to the JWST allocation committee and have secured telescope time for spectroscopic followup.

F. Suggested Improvements: Experiments, Data for Possible Revision

See below.

G. References: Appropriate Credit to Previous Work?

Sufficient, no significant omissions.

H. Clarity and Context: Lucidity of Abstract/Summary, Appropriateness of Abstract, Introduction and Conclusions

Overall, the paper is well written although considerable work will be required to properly partition it into Nature format as a Letter, Supplement, and perhaps Extended Data sections that dovetail well without repetition. The language of the paper in some places needs to be made more professional by elimination of terms such as 'incredibly' and 'superlative' (in other words, no editorializing).

I therefore would be happy to reconsider this result for publication after the authors address my points below.

----

(0) Minor general comment: the authors will have to adjust most of their figures so that just the first word of each axis label is capitalized to conform with Nature format.
Figure axis labels have been updated.

Title

(1) Please change it to something more efficient along the lines of "An extremely magnified star at redshift z = 6.2". Sometimes less is More.
Title shortened to “A Highly Magnified Star at Redshift 6.2”

Abstract

(2) It's initially confusing to classify this as a massive (> 50 Msun) star and then quote temperatures ranging from 8000 - 60000 K. Please clarify that the low-temperature limit corresponds to a post-main sequence star by saying something like "... is consistent with very massive (> 50 Msun) star on the main sequence with temperatures as high as 60000 K or on the post-main sequence with temperatures as low as 8000 K". Note that 8000 K is not usually considered to be a 'hot'

star, please revise your language.
Temperature discussion has been removed from abstract/summary paragraph for clarity and brevity. Subsequent discussion has been corrected as you suggest, thank you.

(3) Later in your paper you adopt a metallicity of 0.05 Zsun for the star based on analysis that is reasonable given the information present. Please state that you find a likely metallicity of 0.05 Msun based on the properties of its host galaxy.
We softened these statements and now explore a wide range of metallicities.

(4) Please omit the point about this discovery giving the opportunity to study a low-metallicity star. It's problematic for two reasons. First, you don't really know the metallicity of the star, it could be 0.001 - 1 Zsun based on observations of metals in other high-z galaxies and a number of numerical simulations. Second, some would disagree that there are few low-metallicity stars in the local universe. You won't have any real chance of constraining the metallicity of the star without NIRspec measurements. This statement can be removed without loss of impact -- the importance of discovering a single star at z = 6.2 speaks for itself.
Discussion of metallicity has been removed from the abstract/summary paragraph, per points 3 and 4 raised here. We also removed any discussion suggesting a rare opportunity to study massive low-metallicity stars.

Introduction

(5) para 3: please define CLASH.
The Cluster Lensing And Supernova survey with Hubble (CLASH) has been defined in the text.

Results

(6) para 1: Although it's commonly understood by the lensing community, please reword "arc at z = 6.2..." to something like "arc of a lensed background galaxy at z = 6.2 ..." The arc is not at z = 6.2.
Updated to “lensed arc of a galaxy at z = 6.2”

(7) 2.1, first sentence: You state "Earendel was identified as having an extreme magnification in a model-independent way by the symmetry of two adjacent bright clumps.." How? It's not clear how you flagged

this object to be extremely magnified. Did you designate some minimum magnification as 'extreme'? Please explain.
This qualitative assessment has been clarified in the text.

(8) bottom of p.3: Please describe in more detail how you carried out your MCMC campaign, either here or on p.14. For example, what is varied between runs?
The text has been updated to clarify. The 100 iterations described were drawn from the posterior distribution generated by the MCMC in the LTM modeling software.

(9) Table 1: you state uncertainties for each quantity on the bottom line, do they apply for all the entries in a given column? I would expect them to vary with lens model. Please clarify and attach specific errors to each entry in the table if they vary.
Table 1 has been updated to include individual 1-sigma uncertainties for each quantity.

(10) Is it possible to dial down the range of magnifications in your models any further? A factor of 40 is quite large.
This is a unique case of a star directly on a lensing caustic, with much higher magnification than typical lensed galaxies or even the few previous lensed stars. The higher magnification yields a much higher uncertainty. Future observations (JWST(accepted)/MUSE(proposed)) will bring this uncertainty down considerably by better constraining the lens models and the unknown distance to the critical curve. We estimate a +/- 32% uncertainty with both JWST and MUSE data combined.

(11) section 2.3, para 1: Do you take into account absorption of flux from the star by dust in the galaxy? One of the authors, Keren Sharon, is an expert on corrections to SN Ia spectra due to local extinction by their host galaxies so you should be able to elaborate. Also, at z = 6.2 one would expect some Lya scattering due to residual neutral H in the IGM, was this folded into your analyses? If not, both will need to be done if they are required.
The HST F110W filter is redward of Lyman-alpha at z=6.2, so we do not expect any absorption from the IGM to impact our measurements in this bandpass. Our calculations of the star's luminosity assume zero dust extinction. The star resides in the outskirts of the galaxy, so we do not expect significant attenuation from the ISM there. Since we find this to be most consistent with a massive star, there may be some absorption from the star forming region in which the star was born. We do not have the data to robustly estimate the dust content of this star forming region, but we can get a ballpark estimate with some assumptions. Assuming E(B-V) ~0.1 (reasonable for young star clusters in low-metallicity galaxies, e.g. Calzetti+15) and an SMC-style extinction law with R_V ~ 2.93, we find a factor

of ~2 change in flux. While this is considerable, it is sub-dominant compared to the magnification uncertainties. This extinction discussion has been added to the text.

(12) section 2.3, para 2: You state that green shaded regions indicate uncertainties in lensing of a factor or 7, but you state earlier this uncertainty is a factor of 40 -- please clarify.
The 7x uncertainties are within a single lens model. An additional factor ~6 difference exists between models, giving the total uncertainty a factor ~40. This has been added to the text to clarify.

(13) section 2.3, para 3 / Figure 3: You quote 8000 K as a lower limit but that is not the final temperature that any of your stellar models reach in the post-main sequence phase, which appears to be 4000 - 5000 K. The transition from high MS temperatures to low post-MS temperatures happens fairly quickly, so it seems much less likely that the star would happen to be found below 20000 K. It therefore seems that the high temperatures are really all that matter. Do you factor the short time it takes for these stars to evolve to your lower limit in temperature into your analyses and Figure 3? Please clarify, I could not find any discussion of this in your text.
We note the 4000 – 5000 K stars are not bright enough in the rest UV 1600Å redshifted to 1.1μm in the observed F110W filter.
We have now added Figure 9 showing the stellar evolution vs. time. Based on this, we conclude 50 – 100 Msun stars are most likely. And thank you, these solutions do spend most of their time above 20000K. We have added this discussion to the text.
A more detailed treatment will be undertaken with JWST data.

(14) section 2.3, para 6/7: Not to be unkind, but you will have to tone down your claim that Earendel presents an opportunity to study a low-metallicity star because you have no real idea of its metallicity, just one based on a reasonable model. In reality, observations and numerical simulations have shown that stars in z ~ 6 galaxies can have metallicities ranging from 0.001 - 1 Zsun (and even have supersolar metallicities at z ~ 7). You can omit this discussion without any loss of impact.
We concur. These paragraphs have been omitted.

(15) section 2.3: Before this result can be published you have to take a good look at how Figure 3 would change with the metallicity of the star. In principle, this would require running additional grids of stellar evolution models at Z = 0.001, 0.1 and 1 Zsun and pairing them

with atmosphere models to determine spectra for low-temperature stars. However, I'm willing to be flexible on how you address this point as long as it's convincing -- for example, considering just stars at high temperatures or using precomputed stellar evolution models from open databases like Heger et al. 2010.

We have included an additional figure in the Methods section showing stellar tracks from BoOST with metallicities of Z = 0.004, 0.02, 0.1, and 1 Zsun. We find that the changes induced by varying the metallicity are notable but small compared to our magnification uncertainties, and do not change our conclusion that we are likely seeing a massive star. We also explored the variations caused by different metallicity stellar atmosphere models, and found it changes our constraint by delta(logL)/(logL) < 0.1 above 10000K, and ~0.2 below. This is insignificant compared to our other uncertainties. This discussion has been added to the text.

(16) section 2.5, para 2: You say this star formed recently but later state that it formed 900 Myr after the BB. I know what you are trying to say but instead just state that the lifetime of the lensed star is much shorter than that of its host galaxy. I think this was what you Intended.

Indeed, this was what we intended to convey. The wording has been changed to say "the lifetime of a massive star would be short relative to its host galaxy".

(17) section 2.5, para 4: Please briefly discuss the timescales on which redshifted flux from stellar mass BHs would vary and compare them to the cadence in RELICS. Can you rule out a BH based on the fact that its flux should be transient?

For this analysis, we assume a persistent BH X-ray binary fed by a companion star overfilling its Roche lobe (as in Windhorst+18). Such an object would shine for 0.6-60 Myr, with little variation in time. A transient BH X-ray binary would shine for weeks to months (e.g. Zdariarski & Gierlinski 04). Accounting for cosmological time dilation, our 3.5-year baseline rules out any transient lasting less than ~6 months in the source plane, thus ruling out a transient BH outburst.

Referee #2 Comments: (Will not review resubmission)
I first reply briefly to the questions as listed on the review form, then provide detailed comments by section and paragraph.

[A] Summary of the key results

The authors state to have observed an object that could be an object at redshift z~6. They cannot constrain much of its properties but state that it might be hot and massive. They cannot even constraint whether it is a single or multiple star, or a star at all.

[B] Originality and significance: if not novel, please include reference

Finding a single star at such high redshift is, to my knowledge, would be novel, and an observational 'jackpot', if confirmed.

They do not provide any probability estimates, so significance, in a statistical sense, of the result is zero.

[C] Data & methodology: validity of approach, quality of data, quality of presentation

There is insufficient t data for useful conclusions.

[D] Appropriate use of statistics and treatment of uncertainties

A prime example of how not to do statistical analysis with model comparison.

The text is threaded with qualitative statements such as "incredibly small" w/o quantification. Inadequate for scientific paper.

[E] Conclusions: robustness, validity, reliability

There is no scientifically useful conclusions based on the available data and theoretical models.

[F] Suggested improvements: experiments, data for possible revision

I think it is unfair to your colleagues to send such a paper for review.

[G] References: appropriate credit to previous work?

References on reviews of stellar evolution and outcomes of the stars
under consideration are thin.

[H] Clarity and context: lucidity of abstract/summary, appropriateness of abstract,
introduction and conclusions

Abstract and summary are flawed.
Pointless speculation that are wrong (summary) and/or uninteresting (abstract).

Abstract

Naming of stars is sole authority of IAU.
Disregarding this is not acceptable.

Please use the standard way to show error bars for upper and lower
limit for radius. Provide centre values plus one-sigma errors (could
be asymmetric).
We have provided a confidence level for our upper limits, as is standard.

The magnification factor is uncertain by factor 40!

The star temperature of 8kK to 60kK is ridiculously uncertain, with unclear
spectral type.

Being consistent with a hot massive star does not imply it is inconsistent
with a cold faint star.

The star being low Z is pure speculation.
On top, this is what would be expected in the first place, so it is
not even an interesting speculation.
The statement that low-metallicity stars cannot be found locally is
incorrect, e.g., in SMC and LMC have low metallicty massive stars, and
Zw 18 has low-metallicty star-forming region.
https://www.aanda.org/articles/aa/full_html/2013/05/aa20948-12/aa20948-12.html
This speculative discussion has been removed, as also suggested by Referee #1.

Introduction

The introduction section has been removed to better conform to Nature style guidelines

P1. What is the maximum angular size of infinitesimal lensed objects?
It depends on how the product of magnified region size and
magnification factor scales, and is likely a lens property and can be
solved using a modest amount of maths. Please clarify.

P2. Do cluster crossing events 'live'? Is that the correct
expression?

P3. What is "CLASH" (please define all abbreviations used, this is
customary in astronomy literature).
The Cluster Lensing And Supernova survey with Hubble (CLASH) has been defined

2 RESULTS

Similar to abstract, naming of astronomical objects is sole IAU
authority. Hence any reference to fantasy names "Sunrise Arc" and
"Earendel" need to be removed.
We have added official IAU alphanumeric designations for both objects: WHL0137-zD1 and WHL0137-LS, respectively. We also retain the nicknames.

The statement of being a high-z arc is at least wrong, if not
misleading. Arcs are apparent phenomena at the observers' side -
there is no are at high redshift; please represent astronomy correctly.
Phrasing altered, see response to Referee #1

2.1
P1. "extremely high magnification" - please quantify
This paragraph was intended as a qualitative discussion, meant to provide a “sanity check” that finding such a high magnification (as described elsewhere) is reasonable given the geometry of the system. Quantification comes in subsequent paragraphs.

P1. "a star sitting incredibly close" - sitting stars? Please
quantify "incredibly"

As above, quantification comes in later paragraphs.

P3. MCMC model - From the information provided no one can deduce whether this statement is of any significance because it is absolutely unclear what has been done. The statement is suggestive but the dependable information is zero.
This section has been rewritten to clarify - see response to Referee 1 above.

Figure 1. "LTM" is used in figure caption before it appeases in the text. Please define in caption.
The definition of LTM has been added to the figure caption.

Table 1. Please show the specific errors for each case in the table with the values rather than providing broad range multipliers below and in the caption. (common notation)
Table 1 has been updated to include 1-sigma uncertainties on each quantity individually

P4. "microlensing suppression would more likely be a transient phenomenon. Therefore we conclude the two lensed images are most likely unresolved in the current HST WFC3/IR imaging." Please quantify.
The microlensing discussion in the main text has been updated and shortened, providing a quantitative assessment of its effect on our observations along with a brief qualitative contextualization.

P4. "This is consistent with our original lens model-independent interpretation suggesting it is directly on the critical curve." But is it inconsistent with other things?
Yes, quantitative lens modeling does not support the interpretation of the lensed star as, say, an additional image of another star cluster.

Last par. "ostensibly up to infinity" (see note above) Please study mathematics and astrophysics of lensing. A star will be an object of finite size and be a at finite distance, the maximum brightness you can achieve is to have the entire 4 pi steradians - the entire sky - shine at that surface temperature. This would be very bright, but definitively less than infinite magnification. As you go to increasingly larger magnifications, the area of the source is increasingly smaller, at some point it is smaller than the object. At this point magnification has no more effect on brightness for the same image size (Liouville's theorem). Finite size of the lens plays a

role in the size of the image you can get.
The offending phrase has been removed

2.2

P2. "Our largest radius constraint is a factor of two smaller than this superlative star cluster, while our tightest constraint is nearly an order of magnitude smaller." - Why do you have different constraints? Would it not be most statistically significant to combine the different constraints, even if this results in a "systematic" uncertainty?
We now refer to a statistical factor of 7 uncertainty (due to the unknown distance to the critical curve) and a systematic factor of 6 uncertainty among the lens models.

P3. It is rather uncommon to define massive stars using a lower mass limit of 15 Msun. Please consult common textbooks or reviews.
Wording changed to “Most stars of mass M>15Msun”.

P4. The binary part is pure speculation with no observational consequence.

P4. Also the possibility of unrelated projected binaries or multiples is not discussed.
The probability of observing a massive star crossing the caustic is estimated to be a few percent. Thus the probability of an unrelated star also crossing the caustic at the same time would be a few percent of a few percent, or a few hundredths of a percent. This is quite unlikely.

2.3

P1. This remains rather unclear what has been done here. I can only speculate that the flux in F110W has been used, based on atmosphere models to derive a luminosity for a given surface temperature. If that is the case, please say so.
Indeed, flux was used to derive luminosity at a given surface temperature, assuming blackbody spectra at Teff>40000 K and using atmosphere models at lower temperatures. This explanation has been added to the text.

P1. How about extinction due to dust or absorption beyond Ly break?

Our detection in HST F110W at z=6.2 is redward of Lyman-alpha, so Lyman absorption will not be a factor. Dust extinction was ignored, however we have added discussion of its possible effects (see response to Reviewer 1).

P2. Where is the flux uncertainty represented, it seems missing.
Certainly it remains unmentioned.
Line 167(original)/471(revised) - “the photometric uncertainties are 10% (insignificant)” - flux uncertainty is not factored in due to being insignificant compared to magnification uncertainty. This has been clarified in the text.

P2. You use [M/H]=-1 atmosphere models, but the stars shown are 0.05
Zsun = [M/H]=-1.3. Is it really necessary to dial together things
apparently randomly?
Variations from different metallicities are small compared to our magnification uncertainties. We have now included a range of metallicities for both the atmosphere models and stellar evolution tracks to show this in the paper.

P3. How do you estimate this? Is there any mathematical basis? Or
do you speculate rather than estimate (which can be a scientific
method)?
Assuming the reviewer is referring to our estimate of the probability of observing a >100Msun star, we detail the calculations in the Methods section of the paper.

P3. There is no 500 Msun star in the plot. How do you get to 500 Msun?
560 Msun track is shown at the top of Fig. 3 in blue.

Figure 3. Such HRDs are not new, such track have been available since
30 yr. Additionally, there is the question of the Humphrey-Davidson
limit (
https://en.wikipedia.org/wiki/Eddington_luminosity#Humphreys%E2%80%93Davidson_limit
)

Figure 3. This is again very suggestive, but as seen in the plots, it
appears as if there was a flat prior for the different star masses and
evolution stages. But this is not so. Maybe you were misguided by
the plot showing circles where they had models - potentially based on
varying-size time steps or data files, and sizes of the circles
potentially related to star radius rather than likelihood of finding a
star there.

We have now clarified we generated the HR diagram with equal time steps of 10,000 years. We also added a new plot of evolution vs. time and improved the text as discussed in response to Reviewer #1.

Figure 3. A proper analysis would require folding the probability of a star of that mass in that evolution stage based on some IMF assumption. This is standard Bayesian analysis commonly used in current scientific literature. Standard practise.
As the reviewer states in the following point, “The constraints are so open, they are not really useful”. Given the considerable uncertainties on the observable parameters of this star, we deemed a detailed Bayesian analysis to not be warranted at this stage. The paper is intended to present the discovery only. Future observations will provide enough data to make the analysis the reviewer suggests worthwhile.

P3. The constraints are so open, they are not really useful.

P4. This is mere speculation, not results. What is the take-out here? Looking at Figure 3, I can see that a track of ~20 or 25 Msun would intersect with the light green shaded region of shifter up by 0.3 or 0.5 dex? If that is what the authors refer to, it is completely wrong. See section above about Bayesian analysis. for one, the folding needs to be done, and for second, those stars spend a very small fraction of their lifetime during that phase. This is on top of the metallicity of the stellar grid and that of the observational data shown differ.
The binary discussion is indeed speculative. However, given that previous studies cited in the paper suggest most massive stars have a companion, such speculation is warranted.

P6. The star is so distant, it can't be really studied in any meaningful way with these observations: You do not know metallicity. You estimate temperature up to about 1 dex uncertainty. You do not know multiplicity of the star by 0.6 dex (factor 4). SO you are really trying to tell the reader you can learn something about evolution of low-metallicity stars form this? This is an insult to the readers' intellect.

P6. Whereas binary BH may be formed by close binaries at low Z, it will still be only a small fraction of them that make such binary BH systems that lead to GW events. There chance that the system you observed is one of those few select is rather negligibly small.

P7. Low Z. See comment on abstract.

P8. So there is hope for data in the future of which the present paper is
devoid.

2.4

(I cannot comment on this)

2.5

P1. Various grids of Pop III and other low-Z stars exits already.
Again, a detailed Bayesian analysis would need to be done for
meaningful results.
We compare to Pop III MESA models from Windhorst+18.

P1. A statement of the kind "X is consistent with ... " has not
merit as it does not constrain what is not consistent. Not having a
bank statement to show for it is consistent with being a billionaire.

P2. Formed recently? Later it is stated it formed 900 Myr after the
Big Bang.
The wording of this statement has been updated to clarify that the star would have a short lifetime compared to that of the host galaxy.

P2. Whereas lifetimes are short what would be the effect of apparent
time dilation of lifetime by a factor 7 due to high red shift?
Time dilation would affect the star and host galaxy in the same way, so it does not change their relative lifetimes.

P2/P3. There is a history dating at last 15 years back that Pop III
stars and their IMF are not all of a kind. It also depends on the
environment and the ionisation stage of the gas. For example, from
early on, after the first SN go off, cosmic rays will propagate,
affecting the ionisation stage of the gas and potentially enriching
the gas slightly. Then there is the Lyman and Lyman-Werner radiation
from the surrounding stars. Finally, an enrichment of [Z]=-4 or less
will affect the IMF; a metallicty of [Z]=-8 or less will affect the
evolution of star so that they already differ from truly primordial
stars - such star may not be distinguishable from metal-fee stars by

their lines bit still evolve differently. They are not Pop III stars.
The present discussion on Pop III stars add no value.

P4. X-rays would also be red-shifted. What would be the variability
of such an accreting BH? Also, 200 Msun may not be a likely mass due
to the pair-SN mass gap.
X-ray emission from an accretion disk as discussed would be redshifted from ~13 keV to ~2keV, putting it within the observational capability of XMM-Newton, for which we have data available. We assume a persistent BH binary, as an outbursting BH binary is ruled out by our 3.5-year observation baseline (see response to Reviewer 1). Based on Windhorst+18, a ~230 Msun BH can form from a ~300 Msun PopIII star. This is what we base our discussion on.

P5. "the incredibly small area" - quantity. It is probably not an
area (unit cm^2) but a solid angle.
It is indeed a solid angle, and we have corrected this phrasing. Quantification comes later in the sentence, “(constrained to within 0.1 arcsec)”.

P6. "it would be highly unlikely" -quantify.
This sentence was removed in favor of other arguments.

3.
--
Section 3 (Conclusions) has been removed to match Nature formatting requirements and avoid repetition.
P1. "observed 900 Myr after the Big Bang" - I though it was a recent
observation.

P1. This also shows there basically very little constraints, hardly
any dependable data.

P1. "Likely" properties of the star are not supported by manuscript.

P2. Speculations about the future only.

P3. "Very massive low-metallicity stars are extremely rare, and thus
few are known even locally." There is 3 things in this sentence that
require quantification.

P3. very massive stars refer to those above 100 Msun which make
pulsation pair SN supernovae, or sometimes those that make pair SNe

and leave no black hole.

P3. Stars that may indeed enter the giant branch may be too far separated to become the binary BH of LIGO.

P3. On top, this is confluence of opposing speculations made in the manuscript. It is stated the star may be massive with > 40 Msun if single, or 25 Msun if binary or maybe 20 if multiple. By all good faith, beyond the low probability a ransom massive binary system making a binary BH that merges as a LIGO source, a binary of two 25 Msun stars will not result in a binary of two 30 Msun BHs.

4.
--

I will not comment on methods section.

Additional concerns and feedback
=================================

Critical

The naming of celestial objects and stars would be the
sole authority of the International Astronomical Union (IAU), and in the interest of science and standards, neither the authors nor nature should bypass this.

Minor

Typesetting has various minor flaws and should be improved for consistency and legibility. (Whereas I assume some of this will be rescued by nature editing, it is always better to start with something done carefully.)

The authors may want to look up the difference between indices and subscripts for variables.

Spacing between numbers and units is not applied consistently.

Various commas are missing.

Referee #3 Comments:
I have read the paper entitled "An Extremely Magnified Individual Star Observed in the First Billion Years" by Brian Welch et al with great interest and enjoyment, and would like to congratulate the authors on their discovery and a very nice paper. To summarise, the paper itself presents the discovery of a lensed star (system) from HST RELICS data that likely resides at high redshift, and performs detailed analyses in order to infer the source's nature as a single star or star cluster and places the result in context by aiming to establish the star's mass and metallicity. Although the data themselves are insufficient to rule out a binary star system, the discovery is one of the highest-redshift discoveries of single star system and provides a useful benchmark for future studies and follow up observations. The methodologies adopted by the authors are appropriate and thorough, with clear explanations and presentation of analyses and results. While I am convinced that this object is likely a single star (system), I would have several revisions/requests I would like the authors to perform before recommending the paper for publication:

Section 2.1:

1. Figure 1: What filters are used to create the image? Could the authors please add small circles to each component/clump of the systems used in the lens modelling (perhaps with different according to each system so they can clearly be identified), in similar fashion to the interloper? Apart from 1.1a and 1.1b, the other clumps are faint enough that they are challenging to identify in the image. Additionally, to what significance are each of the clumps (including the star) used in the lens modelling detected and in what filter?
Thank you, we have now added all these details to the text:
blue = F435W; green = F606W + F814W; red = F105W + F110W + F125W + F140W + F160W. Circles have been added around clumps used in modeling. Clump detection significances in F110W are: 1.1a = 20, 1.1b = 18, 1.1c = 7, 1.7a = 15, 1.7b = 12, 1.7c = 25, star = 12. Note the SNR for 1.7c may be slightly elevated due to blending with a neighboring clump in the detection image.

2. For the remaining 20% of the models where Earendel does *not* fall within D_Crit<0.1", at what D_crit does the star fall?
The maximum distance is D_crit ~ 0.4”. This has been added to the paper.

3. Is there a particular reason why the LTM and Glafic (c=1) results are largely different from the others?
The LTM and Glafic c=1 models each have flatter lens potentials, so the mass is less concentrated in the center of the cluster (hence the low concentration value for Glafic). This

results in a slower drop in magnification as you move further from the critical curve, and a higher magnification at a given D_crit. We added this explanation to the text.

4. Could the authors please restructure Table 1 to include the uncertainties on the individual values? At present it is difficult to evaluate the meaning on the individual values without their associated uncertaities (for instance, there's a big difference if the uncertainty on mu=8400 is x0.7 or x5.0, neither of which is explicitly associated). In particular, what are the error bars for the radius and do the tabulated upper limits include those uncertainties?
Table 1 has been updated to show +/- 1-sigma uncertainties on each quantity

Section 2.2:

5. Similarly to point 6 above, do the upper limits to the radius values quoted here include the uncertainties? And how are those uncertainties derived?
Upper limits do include magnification uncertainties. Formally, our magnification estimates are lower limits, as the distance to the caustic used in this calculation is an upper limit. So the central value quoted for the magnification is actually the 1-sigma lower limit. We then combine this lower-limit magnification with our upper limit on the size in the image plane, again at a 1-sigma confidence level, to produce our upper limits on the source plane size.

6. In addition to referencing the smallest star clusters, please add a reference to Bouwens et al. (2017; https://ui.adsabs.harvard.edu/abs/2017arXiv171102090B/abstract) to highlight that the estimated radius of the star is also smaller than very low luminonsity galaxies at high redshift.
Reference to Bouwens+17 has been added with discussion.

7. Based on the text and caption of Figure 3, it is not immediately clear to me: I assume all of the colored tracks are the stellar evolution models from BoOST? Are these all dwarf galaxy predictions with 0.05 Z_Msol or single stars? Additionally, if tracks of various metallicities are present in the Figure (as suggested in paragraph 5), this is not clear to me. Please clarify (both in the text and the Figure caption).
Yes, colored tracks are stellar evolution models for stars of various masses (indicated on the left of the plot). Each star shown has metallicity 0.1 Z_sun. We have now added a separate plot showing these tracks for different metallicities.

Section 2.4:

8. Would the results of such a simulation change significantly if a different velocity were assumed? Are there any (reasonable) star velocities that would induce significant magnification variations not observed in the photometry, accounting for the uncertainties?
We added this discussion to the text: Microlensing results would not change significantly with any reasonable change in velocity. In this case, the microlensing caustic network is dense and has many overlapping microcaustics. Thus as the star moves relative to the network, its total magnification will not change much (typically staying within a factor of 2). Velocity becomes more important as the microcaustic network thins, as then it will determine the frequency and duration of microcaustic crossing events that change the total magnification considerably.

9. Presumably the the gray bands in the top right panel of Figure 4 are the 1sig uncertainties? Please clarify.
Gray bands highlight a factor of 2 (dark) and 4 (light) variation. This has been clarified in the caption.

Section 2.5:

10. Assuming ~100 stars per arcmin2 in a 1 arcmin2 field centered on Earendel, what is the probability of a star landing close to Earendel's position if such a density was injected into the image e.g., 10000 times? Are there any that would land close?
We constrain the location of the critical curve crossing the arc to within 0.1arcsec with our lens models. So assuming an object landing within 0.1arcsec of Earendel counts as close enough that it could be mistaken for a lensed star, with a density of 100 stars per arcmin^2 we find a 0.08% chance of a local brown dwarf masquerading as our lensed star. We quote this in the paper as a probability "of order 0.01%" given the uncertainties on the density of brown dwarf stars and the vague determination of what is "close enough" to warrant a false detection.

11. Second-to-last paragraph: I would like the authors to expand on reasons for Earendel not being a Brown Dwarf via SED-fitting tests described for Section 4.2 below.
We have now tested the SpecX models: see response below.

12. Last paragraph: similarly, I am not convinced of the reasons stated by the authors to exclude a low-z solution: the discovery of Earendel is already a fortuitous event (quoted at P(100 M_sol)~2-4% in Section 4.6) and thus it is not unreasonable to assume other low-probability events could mimick the discovery. I suggest expanding this paragraph and the one prior to include mention of the tests requested for Section 4.2 below.
Our photo-z results are confidently z ~ 6.2 for the Sunrise Arc and for Earendel individually, given the clear Lyman break, as we checked and added to the text. Nevertheless, in case

these photo-z's are wrong, we added discussion about galaxies at other redshifts: most would likely be spatially resolved. While we deem them unlikely, we acknowledge confirmation requires our upcoming observations with JWST.

Section 4.2:

While I agree with the authors that Earendel's size is the strongest constraint for it being a single star or star cluster, I find details of this section of the paper somewhat lacking in establishing the high-z nature of the galaxy and Earendel, and would like to authors to clarify some aspects and perform some additional tests to ensure that Earendel in particular is not a low-z interloper or brown dwarf.

13. How are the stacks of the arc flux and uncertainties performed? Please clarify and state in the text.
Flux from each segment of the arc is summed to create the total arc flux. Uncertainties are added in quadrature to calculate total uncertainty on the full arc flux. These details have been added to the text.

14. What redshift prior and range are the authors adopting for BPZ? For the latter, presumably 0.1<=z<=13.
The redshift range is 0.001<=z<=13, with a prior based on magnitude and spectral type, described in Benitez 2000. This description has been added to the text.

15. Do the BPZ models include those of brown dwarfs (e.g., the SpecX library) and z~2 dust starburst galaxies? If they do, what are the likelihoods of (i) the best-fit brown dwarf model and (ii) a low-z interloper compared to the best-bit z~6 solution? If not, I would like the authors to please fit the SpecX brown dwarf library as well as add a model of a z~2 dusty starburst galaxy to evaluate those likelihoods.
We tested low-z dusty / old SEDs explicitly using BAGPIPES and found all yield intrinsically red spectra in the near-IR that are unable to reproduce the Sunrise Arc's observed photometry that is flat in F_nu (or AB mags) through the near-IR (1.0 -- 1.6 microns). The 5 near-IR photometric data points plus Lyman break blueward are only fit well by a galaxy at z~6. We have now added this explanation to the text.

Thank you for the SpecX suggestion. We have now used this library to model Earendel and found that a 3000 K brown dwarf could reproduce the observed photometry, with a total chi-squared of 6.4 for 9 filters. We have added this result to the text, and mention that we now rely on the low probability of chance alignment and lack of proper motion to disfavor such a star. Again, JWST will provide the final say on this.

16. What star formation history and which parameter ranges are adopted in the Bagpipes fitting, and is the redshift left as a free parameter or fixed? If the redshift is fixed, please repeat the fit allowing the same redshift range as the BPZ range and state how the z_phot and P(z) solutions compare with BPZ.
Redshift is also left free up to $z < 13$ in the BAGPIPES modeling, as now mentioned in the text. The photo-z results from BAGPIPES and BPZ are very similar as shown in Figure 4 and discussed in the text.

17. While the original HST photometry clearly does not have the power for any meaningful SED constraints on Earendel, the 3-year photometry might. Could the authors please repeat the BPZ and Bagpipes exercises (including the SpecX ad low-z interloper exercises) for only the 3-year photometric points of Earendel and compare likelihoods, z_phot and P(z) solutions.
Indeed the 3-year photometric flux measurements have smaller uncertainties and greater influence on the SED fits. But we see no reason to exclude the other data, which also contributes.

18. Please could the authors tabulate the extracted photometry (and 1sig uncertainties) of each component in a Table that get used in the fitting process.
Thank you for catching this oversight. We have now included the photometry of both the Sunrise Arc and Earendel in Table 2.

19. Have the authors checked whether Spitzer/IRAC data is available for the cluster/star? RELICS has decent Spitzer overlap (e.g., SURFS UP, PI: Bradac, Strait et al. 2021, etc)? If so, such photometry should be included in the above fitting tests/procedures and would probably be of great aid in the SED fitting.
Spitzer data does exist for this cluster. However, fluxes could not be reliably extracted given the bright cluster galaxies near the arc/star, and the faintness of the arc. This information has been added to the Methods section.

Referee #4 Comments:

This paper presents the discovery of a z~6.2 star that is highly-magnified by a foreground galaxy cluster. This discovery is the first of its kind, and its position close to the lensing critical curve enables it to be visible at high magnification for an extended period of time where it can be followed-up with future observations. The authors' analysis is thorough and convincing, and the paper clearly presents the data, methodology, and reasoning behind their conclusions. I do have some minor comments on the paper that I think would help clarify and strengthen certain points. Once these are addressed, I will recommend that the paper be accepted for publication. My comments are listed below:

Section 1, line 73: It is stated that the magnified stars (the source) are moving relative to compact objects in the foreground lensing cluster, causing microlensing variations. Is this entirely (or mostly) due to the motion of the source, or does the motion of the microlenses in the cluster matter as well? I would expect the proper motion of the foreground stars in the lens galaxy to be greater, so this statement seemed somewhat counterintuitive. This also seems to contradict a statement later in the paper (Section 2.4, line 224).

The dominant motion here is really the motion of the foreground lens stars. The velocity of stars within the background lensed galaxy will be small compared to the ~1000 km/s transverse velocity assumed for the lensing cluster. This statement has been revised accordingly.

Section 2.1, line 107: Unless it is computationally unfeasible, generating an MCMC chain of only 100 models seems too short to draw any conclusions. Can the authors run a longer chain, or explain further why this sample is sufficient?

The 100 models described are generated by sampling the posterior distribution generated by the LTM software. The actual MCMC chain used during model fitting is indeed much longer (several thousand steps, as standard for the LTM methodology). The 100 samples drawn are standard to estimate lens model variance. This description has been updated in the text.

Table 1 (and in the text where relevant): Do the uncertainties in the table apply to each lens model, the best (LTM) model, or is it the total uncertainty accounting for differences between the various lens models? If it's the former, I find it unlikely that the models all have the same level of uncertainty given the difference in methodology. If it accounts for differences among models, the authors should provide some effective weighted average.

Table 1 has been updated to include 1-sigma uncertainties on each quantity individually. Given the unique nature of a lensed star on the critical curve, we calculated magnification uncertainties not as typically done by drawing samples from each lens model's MCMC.

Rather, we determine the magnification uncertainty based on the star's unknown distance to the critical curve. We discuss the reasons for this in the text, but basically the MCMC does not provide fine enough sampling of valid solutions. Many of the lens models that reproduce the lensed multiple images well would not accurately reproduce Earendel as a single lensed image.
We now refer to this as a statistical factor of 7 uncertainty, while the various lens models add another factor of 6 uncertainty. Future observations will reduce both of these.

Section 2.1: Are there any line-of-sight foreground objects that could potentially affect the lens model?
The bright galaxy to the SW of the cluster BCG is likely a foreground galaxy (z~0.3). It is included in some models (i.e. Glafic) but does not significantly impact the lens model overall, its effect is only seen in its vicinity. There does not appear to be an overdensity of galaxies along the line of sight besides the main cluster. Smaller objects that may be cluster galaxies or may be unrelated line of sight objects appearing near the long arc have been accounted for in lens modeling (galaxies C and D).

Section 2.2, line 144: Does the statement about the typical radii of star clusters apply even at z~6? What do theory/simulations predict?
Lensed star clusters at z~6 have been observed down to <13 pc (vanzella), [text redacted related to unpublished data]. A few papers (e.g. Behrendt+19) have runhigh-resolution sim ulations, but they are limited to ~10 pc in their highest resolutions. Sothe smallest known star clusters at high-z are still considerably larger than what we see forEarendel. This disc ussion has been added to the text.

Section 2.4 and 4.6: If Earendel is a binary or multiple-star system, would a lack of microlensing variability in the future be able to indicate this? Or would this be degenerate with the angle of motion of Earendel relative to the caustic?
There is some degeneracy, but finding time-correlated flux changes would be a good indication of Earendel being a binary/multiple system. This will likely need a greater number of observations with a shorter time cadence than we currently have planned for HST and JWST follow-up, but it is something we may consider for future observing proposals.

Section 3, line 319: Alternatively, knowing the temperature of Earendel from JWST spectroscopy could provide a direct magnification constraint at its location in the lens model. Would this potentially supersede any lens model improvements from additional multiply-imaged systems?
We estimate JWST alone (assuming no further multiple image detections) to provide a factor ~4 improvement on magnification constraints. Additional multiple images would

constrain the steepness of the lens potential, significantly reducing the factor ~6 uncertainty we present in the paper. So lens model improvements will probably provide the stronger constraint alone. Also, our planned Cycle 1 observations will only be able to constrain temperature up to ~40000 K, so there is a chance that remains a lower limit if this is a hotter main sequence star. But of course, both together would be the best case scenario to reduce magnification uncertainty.

Section 4.3, line 382: Rather than taking the weighted average of all four IR bands in the stack, what if you just used the two bands closest to F110W? This would make the effective wavelength closer, even if the uncertainty is larger.
We added this calculation to the text. Using only F105W and F125W, we find an average of 34 +/- 14 nJy, similar to the F110W result (35 +/- 9 nJy). The results are similar, and it is useful to confirm the value at this wavelength.

Section 4.4.1 and 4.4.2: When allowing the relative weights of the six particular galaxies in the model to be freely optimized, this means that their M/L is allowed to take on any value with no prior? Do the optimized M/L of these galaxies in the best-fit model appear realistic, or are some of them significantly different from the global M/L? If you were to enforce the same M/L scaling for all galaxies including these six, how would this affect the results?
We now explain in the text: Priors on M/L are enforced. In LTM, priors are M/L within 0.5-3x the general scaling (effectively, the brightness is multiplied by this factor and then scaled to mass following the general relation). Lenstool has priors on velocity dispersion for each galaxy, which are set by M/L for “fixed” galaxies but set free for “free” galaxies. These were given Gaussian priors centered around the M/L-calculated dispersion, with width 15 km/s. Optimized values are all realistic (for LTM, they range from 0.8 to 2.2, for Lenstool they stay within ~2.5 sigma). Setting them all to the general scaling has a small effect overall, but makes reconstructions of the arc somewhat worse.

Section 4.6, line 575: Why is a Kroupa IMF assumed here when Salpeter was assumed for the arc SED fit and Salpeter and Chabrier were assumed for the stellar surface mass density calculation? I think a rough estimate of the effect of different IMFs would be useful in this section (e.g., a Chabrier IMF would increase the probability by ~X times).
The different IMFs would primarily impact our calculation of the SFR which is then used in our probability calculation. The largest difference would be seen using a Salpeter IMF, which would yield a factor ~1.5 larger SFR. However, this would then be effectively canceled out by the lower probability of forming a >100 Msun star. This discussion has been added to the text. Additionally, we note that the most significant change would come with a top-heavy IMF, which could be seen in low-metallicity environments. This would increase our probability, as we discuss in the paper.

Section 4.7, line 603: It would be useful to show the two rectangular regions parallel to the arc in a figure somewhere.
A figure showing the regions used has been added in support of this section.

Reviewer Reports on the First Revision:

Referee #3 (Remarks to the Author):

I firstly must offer my apologies to the authors for my delay in reviewing the modified manuscript. I thank them in turn for their efforts in answering my questions, comments and concerns, and can confirm that they have done so satisfactorily. My conclusion is that while the data cannot fully confirm the nature of the object, there is a high chance of it being a magnified star/star cluster, which will have to be confirmed with JWST. The publication of the article is timely given the launch of JWST and will provide a very interesting object for follow up. Thus, I recommend the article for publication.

Referee #4 (Remarks to the Author):

I thank the authors for revising the manuscript in light of my comments and the comments of the other referees. I am satisfied with the changes they have made. This is a novel discovery that will lead to clear follow-up analyses with upcoming data. Although the uncertainties on specific properties of Earendel are admittedly large, the authors make a convincing case that this is in fact a high-z star. With the changes made to Table 1, the results are much more clear. I recommend this paper for publication.