Co-ordinating Editor:  
Comments to the Author:  
Two very helpful reviews of your manuscript have been received. Both highlight that this is an interesting / promising study, but as noted by both reviewers, there are substantial problems in the design, analysis and interpretation of the work. Referee #1 provided very detailed descriptions about these limitations - and makes some excellent suggestions for re-analysis. The lack of plot level environmental condition assessment effects on fire outcomes seems to be particularly worrying. Care also needs to be taken in the interpretation of the work as replication is very limited (particularly the high fire severity scenario).  At this stage, I cannot accept the paper for publication but I encourage you to re-examine the data and its analysis, and consider submitting your work for publication to a more local journal. JOHN  
  
  
Referee(s)' Comments to Author:

JTS comments in blue

Referee: 1  
  
Comments to the Co-responding Author  
General comments on Stevens et al. ms  
I like the general concept of the paper, nice idea to transfer things that the first author did in California to Colorado (although I’m a bit aghast at the idea that Raven and Axelrod’s [fire-denying] work is applicable to all of the western US without major new research), and nice to have prefire data set to work with. However, there are some serious shortcomings of the dataset and the analysis and I come away with the feeling that the authors are pushing these data a bit farther than they can reasonably go.

Not sure what “fire-denying” means

One issue is the very small sample size, especially in the high severity plots (only 4), this is a big issue in interpreting all of the results. Secondly, there is no information in the model about local sampling environment (local water balance/insolation variables driven by slope and aspect, slope position, e.g.), which will have a major effect on the results of a study looking at mesic-habitat vs xeric-habitat species. Third, it is not clear how the very important annual precipitation numbers are being taken into account in the analysis. To outline the issues: the first partial sample was made 6 years before the fire during a very dry year and then combined with a second partial sample 5 years before the fire in a super wet year (could the difference in prefire floras among the plots have to do with the sampling year?); then the subsequent years until the fire year were notably wetter on average than the long-term average, then comes the fire year, the driest ever, and finally every one of their sampling years happened in below-average precip years. The effect of interannual precip changes and the drying trend after the fire are probably big in this study, but these variables don’t make their way transparently into the analysis and they certainly aren’t mentioned in the Discussion.

Paula, these are good points.

1. Limited N in high-severity areas; nothing we can do about this except maybe note that this dataset has been published elsewhere with this limited sample size?
2. Yes I think it would be good to incorporate environmental variables, could you share this data or possibly incorporate it into “tc understory species cover hayman 1997-2012 for jens.xlsx” in the github directory?
3. Curious your thoughts here. Do you know which plots were sampled in 1996 vs 1997? We could do a comparison of the species’ origins in the two years, and show no significant difference hopefully. If the plots that eventually burned at low severity were all the ones sampled in the drier year 1996, we are going to have a problem. I think the reviewer is otherwise off base with their criticism here a little bit… It’s important to note that we didn’t use abundance for any of our analyses, merely presence/absence. It is unlikely that drier than average years alone (without fire) would explain the complete extirpation of a species from a plot, especially since (I think) this is largely perennial flora… But perhaps if we could show no differences in presence/absence of the two biogeographic groups between 1996 and 1997 that would actually help us make this point? And on some level we have to just accept that all the post-fire sample years were below average precipitation and there’s no way to really control for that, but it’s also noteworthy that the ratio in Figure 3 rebounds in the low-severity plots but not in the moderate or high severity plots, which could be a proxy for the changes driven by precip alone (since there would have been minimal canopy alteration). I guess we should flesh this out more in the discussion though.

So I think the analysis needs major work. With respect to the interpretation, the authors don’t do a very credible job of explaining the high severity pattern. They point out that the proportion of xeric to mesic taxa increases, but they suggest it is mostly due to a loss of mesic taxa. However that change only happens over a one year period (and it is actually a change from a sample taken 6-7 years previous, not right before the fire), and then over the next three years there is the steepest increase of species numbers in either biogeographic group for any period of time. So what really happens in the mesic group is a big turnover of species, with a short-term loss followed by long-term gains, not a catastrophic long-term loss. The only sustained change in the high severity dataset is a continuous increase in the xeric taxa.

Reviewer is a bit off base here, their comment appears to be mostly based on looking at Figure 2, while our speculation about long-term loss of individual mesic species is based mostly on Figure 4, which DOES show that in high-severity plots mesic species are more likely to disappear and not return. I’m not too worried about this but in the discussion we should make sure we refer to Figure 4 whenever we talk about “catastrophic long-term loss” type stuff.

Also, Figures 3 and 4B show that in year 2012 the highest mesic group proportion is in the moderate severity plots and the proportions are lower (and not different from one another) in the low and high severity plots.

I think the reviewer is only referring to Figure 3B here, and they are right, though I think this effect is largely driven by where the low-severity plots started, with fewer north temperate species overall. Still I suppose we should note this in the discussion where appropriate.

The general gist of the Discussion is reasonable from the standpoint of the theory that the authors are working from, but I’m not convinced that the data actually show the patterns that they want them to show. What will the results show after more explicitly incorporating plot-level data on the local environment and more transparently incorporating the annual precip values? Like the authors, I believe that the general nature of what they describe is likely to be true in Rocky Mountain forests, but I don’t think the current analysis is providing the unarguable proof they seek, at least partly because of underlying deficiencies in the dataset.

Specific comments  
  
91              Remove comma after (2010)

Done

95-96           Change to “…a large fire that occurred in dry conifer forests of the Colorado Rocky Mountains, USA…”

Done

104             Sierra Nevada just pops up here out of the blue. What is this about the Sierra Nevada?

Added Sierra Nevada to earlier citation of Stevens et al. 2015

110-200 Methods, in general: readers need to know a lot more about your plots, and you can’t run this type of analysis across such a large landscape without explicitly including differences in sampling environment in your models. You must clarify that you only had four high severity plots, this is a very small number, and there was heavy sample size imbalance among the classes. You must also incorporate plot-level measures of the environment in your analyses, as the local environment could be playing a major role in driving the patterns you see, especially when you have such small sample sizes which couldn’t have allowed even coverage of the variation across the landscape within the severity classes (or did you sample only flat ground?). What were the slopes and aspects of the plots, why didn’t you use TCI or TRMI or something like that in your models? Basically you have a study here which is taking advantage of a pre-fire dataset that wasn’t very big – still a nice use of prefire data – but you are straightjacketed by the plot locations. You need to be much more upfront about this in the Methods and come clean with the readers about the challenges this presents. Given all of this, I don’t see how you can do these analyses without incorporating info on the local plot environment.

Relates to general comment above and is reasonable. Jens will run analyses once Paula can share data.

135-136 Hmmm. Combining plots from different years? Denver area precipitation in 96 and 97 differed hugely (nearly 2x more precip in 97 than in 96). How did this affect your data? You should provide a graphic in the paper showing the precipitation for each year you sampled, with the long term mean shown in the graphic (and/or some measure of the first snow-free date on the site for each year, or maybe the April 1 snowpack measurement?).

Relates to general comment above and is reasonable. Paula, what are your thoughts on data sources available to illustrate long-term trends in precip? I’m hoping this is something you would be able to put together, possibly as an appendix figure.

149-152 Up to 50% mortality isn’t low severity by anyone’s reckoning. Why don’t you use more standard classes?

Paula, thoughts? Presumably there is methods justification in other Hayman papers?

187             When you say you accounted for repeat sampling, how exactly did you do that in the model? How is the year effect included? It is very important in this dataset because precip has been very variable in central Colorado over the last few decades. I went into the climate record and found that every single year you sampled was a below average precipitation year (you missed the two wet years in the decade after the fire), whereas the years between your first sample and the fire were notably wetter on average than the long-term mean. I’m not seeing evidence that you took this variability and the postfire drying trend into account

Jens: I added clarifying language about how random intercepts control for repeated measures sampling; the defense of fire being the main cause of floristic shifts and not a drying trend will be more nuanced but could in part rely on interpretation of the “rebound” in mesic flora in the low-severity plots but not in the high-severity plots, which I mentioned in response to the reviewer’s general comment above.

191-200 Need to better explain this. I think you mean that species had to disappear and then not reappear in any of the subsequent sampling years. Right now it sounds like if a species was present in 1997 but absent in 2012 (i.e. it could have been also present in other years between), you call it “permanently extinct”, and the same for spp absent in 1997. It is just semantics but please make this more clear to avoid confusion among your readers

I think the language is fairly clear but let me know what you think and feel free to suggest additional changes. I added a sentence to the end of this paragraph acknowledging that “permanent” is only with respect to the 2012 sample.

212-213 I would call this marginal significance, not non-significance. Also, it seems to me that the chances that the non-significance of the pattern is driven by small sample size are very high.

Done, added that high-severity stands had “high variance due to a small sample size in this burn severity class”

223             Add “of” between “proportion” and “northern”

Done

231-239 I disagree with this conclusion. The graphic in Fig 3 shows clearly that the major change over time is the increase in the xeric component, not the ephemeral decrease in the mesic component. The mesic component recovers over time and even exceeds its prefire number (sure the P value was 0.08, but you had only 4 plots). Also you assume that the fire caused all of the extinctions you saw in 2003, but there were 6-7 years between the original sampling and the first postfire sampling, some of these species could have easily also dropped out locally before the fire (although most of those years were relatively wet, but that info needs to make its way into your argument).

I assume the reviewer means Fig. 2, and the reviewer makes a good point. There are two things happening concurrently. First, the initial decrease in absolute richness of mesic flora in high-severity plots is in fact followed by a recovery to somewhere near pre-fire levels, as the reviewer rightly points out. Second, our statement was driven by the fact that we saw more northern species disappear and not return in any of our sample years in high severity plots than in low or moderate severity plots as shown in Figure 4). We agree that we should make this second point more clearly and not use it as the primary cause for the shift in the ratio. In fact, the recovery of richness of northern flora in high-severity plots is actually due to the slow but consistent colonization by new northern species (around one species per year) as shown in Figure 4, but the fact remains that there are a significant number of northern species that do not return in the ten-plus years after the fire..

Figure 4                Forgive me, but I don’t understand the two sets of values shown for each sampling year. Need to clarify how to interpret this in the figure caption. This figure shows that the long-term colonization of the plot by mesic taxa more than compensates for the ephemeral loss the year after fire

It’s clear that the reviewer’s not understanding Figure 4 contributed to some of their criticism (including the above comment). We might want to think about how to revise this figure, I am open to suggestions if either of you have any.

255             replace semi-colon with comma

Done

266             Did this sentence get finished? We all know that lichens are another taxonomic group. Did you mean to note that they are characteristic of mesic habitats?

Jesse? This was your addition…

275-276 Use FVS (or the formulas therein) to estimate the canopy cover. You have the tree data, right?

Paula, how easy would this be for you? You mentioned the possibility of it in your email…

281-282 You actually don’t provide any direct evidence of lack of composition change, just lack of species abundance change and lack of change in the two biogeographic components. Need to stop using the term composition throughout, or at least qualify that you mean the relative balance of two groups of species.

A fair point, but I would argue that Figure 4 does show lack of compositional change in low severity sites, because there are very few species extinctions. We need to make this link more explicit though.

289-290 Actually you don’t demonstrate anything about heterogeneity. Need to change this. You are just talking about it, you don’t actually report any heterogeneity measurements

Fair point, we could probably say “our data suggests that this heterogeneity… appears to be important…”

309-310 I wouldn’t make that stretch. Raven and Axelrod also nearly completely discounted the role of fire in driving evolution and species assemblages, which was clearly hugely wrong on their part

Fair, we can take this out.

444             Table 2: remove “going”  
Done  
  
Referee: 2  
  
Comments to the Co-responding Author  
This ms used a unique data set including pre-fire data to evaluate a hypothesis that species associated with warmer and drier climates to the south disproportionately increased after wildfire in Colorado.  This study examined the hypothesis of thermophilization, whereby disturbance produces a community of species affiliated with warmer and drier conditions.  I found the premise of the study to be interesting and the data set to be unique.  I did have some concerns, however, noted below where the findings of thermophilization might not be as pronounced as expected (?), uncertainty with how species richness was presented and analyzed, and difficulty in understanding the figures.  I hope that the comments below are helpful in revising the ms.      
  
Main comments  
1) The presentation of species richness was confusing throughout the ms and would particularly benefit from clarification.  What area was species richness associated with?  Was this per 0.1 ha plot?  Or was this in total?  This should be clarified in the Methods and also in locations such as Results line 204.  Does the degrees of freedom correspond with n=sampling units (over time) or n=total number of species assessed?  Suggest clarifying throughout using notation such as species/0.1 ha or species richness (0.1 ha) or something similar so that people have meaning for what area species richness refers to, as richness only has meaning if tied to a particular area (or point along a species-area curve).  Given that the number of 0.1-ha plots differed among fire severity classes, presumably species richness MUST be standardized to be per 0.1 ha, to avoid a situation of comparing "apples to oranges" whereby the greater number of plots in low-severity areas would automatically thus be expected to have the most species, owing to the species-area relationship?  In short, this needs clarification throughout.

Good points

* Yes this is species per 0.1ha plot. Jens will update throughout.
* Jens will double check df but I think it might be Nplots x Nyears – 1 in this instance

2) There is some evidence that southern species did appear to increase (relative to northern species) on the high severity areas, but this evidence might not be as strong as it seems.  For example, in Fig. 3, note how the one point is even seemingly an outlier, low compared to the rest.  The P value on line 213 for the increase in northern species is 0.08 for high severity (for raw richness).  If that one seemingly outlying point was removed, how might this affect the analysis of raw richness and proportions?

Good point, that one outlier probably does matter and we could remove it and report the result without in the interest of transparency. The increase in northern species in high-severity areas might be significant, and the increase in southern species will probably still be significant.

Moreover, in looking at the graphs, the trends for northern and southern species after fire really are quite similar…..and we might also automatically expect that the proportional increase in northern (as compared to southern) species would be smaller, owing to the greater number of northern species to start with?  In other words, if the southern pool increased by 2 species, and the northern pool increased by 2 species, then that is a huge proportional increase for the southern pool given the few species to start with but not a large proportional increase in the northern pool because there are already many species.  It seems like this needs consideration.

Not sure I totally understand this, help? We’re not comparing proportional increase in northern species to proportional increase in southern species, we are comparing change in the ratio of the two groups over time…

3) It very well could be that the underlying mechanism for increases in southern species (with due consideration to #2 above) could be warming and drying affinity.  And regardless, if that correlation exists, it would be of interest.  But is that the actual main mechanism?  How do we know that, for example, it might so happen that the southern species have better mechanisms for seed dispersal, soil seed banking, or triggering of germination via heat or smoke?  I would actually think those sorts of mechanisms could be as or even more important as climate affinity.  The reason being that to the south are the frequent-fire pine forests of Arizona, New Mexico, and Mexico.  To the north of Colorado, there are forests where fires are not thought to have been as frequent.  The warm-dry affinity is implied throughout and could be true, but what about these other mechanisms that cannot yet be ruled out and might actually be paramount?  I believe other possibilities should be discussed.

I talk about some of these in my JEcol paper; we have trait data in that paper to suggest it’s more a microclimate effect than a fire effect, and it’s probably worth a rehash here. One point from our data is that we don’t see as strong shifts in the ratio in low severity plots despite the fact that they burned, and (presumably) there was litter removal and soil disturbance here just as there was in the high-severity plots. The post-fire environmental data will help us make conclusions on this.

4) Canopy disturbance is emphasized, starting with the Intro (lines 59-62), but other known impacts of fire severity, such as removal of the O horizon and changes in nutrient availability, are not mentioned.  It would seem they deserve at least mention, because it is not convincing that understory changes are solely controlled by the tree canopy, at least not in lieu of data showing that soils and soil seed banks are not important.

See above comment

5) The figures have potential but I did not understand them.  See specific comments below.           
Will comment below

Minor comments  
Abstract  
24 there is no hyphen on words with ly ending, so should be severely burned (no hyphen)

Done

Intro  
55 what is the number "2" here?  Should this be "too" or "also"?

Typo, fixed

60 extra space after microclimate, and appears this should actually be a semicolon to make this a compound phrase?

Done

64 Don't all species have resource-grabbing strategies?  Otherwise, how survive?

I was thinking weedy/ruderal type species but this wasn’t very clear, I’ve deleted it.

Methods

121 and 143 can delete "very"

Done

184 at what scale does this richness correspond to?  Is this average per 0.1-ha plot (should be stated to tie richness to area).

Yes and Jens will update throughout.

186 What were the actual factors in this model?  Was this a standard repeated measures design consisting of year, or was burn severity also included as factor, what of interactions?

Jens will clarify.

191 I like the assessment of colonization and turnover.  Were these analyzed statistically in any way?  If yes, please state.  If no, that is fine, could then mention that this assessment focused on \*estimation\* of the values rather than applying inferential statistics.

I didn’t analyze statistically, I probably could. Lower priority. Thoughts?

Results

202 Again, here, I'm confused as to what species richness?  Is this in total, or is per 0.1 ha?  Species richness must always be tied to a unit area (e.g., 0.1 ha) or sampling domain (e.g., a group of islands) to have meaning.  It would seem that this should be per 0.1 ha?

Jens will update throughout.

Discussion

264 It could be canopy removal, but why only canopy removal in the high-severity area?  I would actually think that fire effects on mineral soils and the O horizon, as well as on the soil seed bank and aboveground seed sources, could be just as important (maybe more so?) than canopy removal.

See response to general comment #3.

275 data are plural in this context, so I think "is" should be "are"

Done

280 The definitions of resistance and resilience in the ecological literature have become confusing, as authors have deviated from standard definitions in the English language.  What is referred to on line 280 is actually "resistance" by definition - the ability to resist changing after disturbance.  If and only if resistance fails, resilience can come into play - returning to some previous condition after changing to a varying degree following disturbance.  If there is no change (ie high resistance), then resilience is actually not possible.

Thoughts? Can update.

289-295 This is a little confusing.  Did the present study demonstrate this (if so, where are the data?) or are those citations cited being referred to as demonstrating that canopy disturbance heterogeneity affects understories?  In the present study to demonstrate this, wouldn't this require a data set of fine-scale canopy cover with data on understory communities across canopy cover gradients to examine possible relationships?  Moreover, wouldn't an experimental approach to tease out the relative influence of canopy cover with other co-varying variables, such as fire effects on soils, be needed to determine cause-effect?  I worry that in the absence of these types of data sets, these statements might be a little misleading....as I myself actually think that factors like distance to seed source, soil seed bank and aboveground seed production, soil heterogeneity etc might be just as important as canopy cover.

Good point, see comment on Reviewer 1 lines 289-290. Can tone down language since we didn’t test.

Table 1: given that the number of plots varies, would this be better to show as percentages?  There is a problem with formatting of the species.  The species should be italicized and scientific names are not presented using underscores? (same for Table 2).

Reasonable suggestion and could do easily. Thoughts? Jens will fix formatting.

Fig. 2 Please provide the area associated with "plot" so that this figure is stand alone.  What are the error bars?  Should there be letters or other statistical comparisons so readers can see what is statistically significant?

Area and error bar explanation an easy fix, Jens will do. Letters/statistical comparisons not trivial since there are a massive number of potential pairwise comparisons. Instinct is to leave as-is and let error bars give indication of significance. Thoughts?

Fig. 3 What is this showing in the upper panel?  is this the proportion of TOTAL species recorded on all plots within a burn class?  For the 2nd panel, what are the error bars?  Please also refer to species richness as species/0.1 ha or at last provide what area this corresponds to.

Easy fixes to caption, Jens will address

Fig. 4 This looks like it could be a neat graph, but I don't understand what it is actually showing?  What are the error bars?  (SD's, SEMs?).  How do we tell which one is the gray bar associated with the northern species and the black bar associated with the southern species?  For each column bar (ie fire severity class, green orange red), per year, there is one error bar for colonization and one error bar for extinction.  That is necessary as each mean needs one error bar.  So how are both the northern and southern ones being shown when there is only one error bar per mean?

Reviewer 1 had a similar problem. Seems like the outline (gray vs black) is giving problems. Maybe just breaking up into a four-panel figure (top row colonizations, bottom row extinctions, left column northern species, right column southern species) would make this easier to interpret. I may have gotten too clever for my own good with this one. Thoughts?