First, I apologize for the delay in reviewing the authors' response to the reviewers' comments, which was due to other commitments. I have carefully read the revised version and the responses to the reviewers and have decided not to send it back for further review. While the manuscript has significantly improved, there are still two critical points, along with some minor issues, that need to be addressed before I can formally recommend it for publication. I believe these final revisions are essential for the study to serve as a solid reference in the scientific literature. I fully acknowledge the substantial effort the authors have already put into revising the manuscript, but I encourage them to make this final effort to ensure the highest possible quality. I shall recommend this manuscript immediately upon receipt of these revisions.

## **CRITICAL POINTS**

## 1. On the methodological justification of using RBA

I may have been unclear when suggesting the incorporation of <u>effective</u> population size into the reasoning, and I agree with the authors that estimating effective population size is beyond the scope of this study. However, as a first step, it was necessary to establish a clear relationship between RBA and population size to ensure that a decline in RBA indeed reflects a demographic decline.

The new figures Figure S7 and S8 help clarify this relationship, showing that, for Fir in particular, the lowest median minimal relative BA corresponds to very low population size. Of course, this is partly a consequence of simulating a lower forest area for monospecific Fir stands, but nonetheless, the N/ha of fir is lower than that of Beech. By the way, these difference in census population size likely contribute the greater efficacy of response to selection in Beech as compared to Fir, resulting in the possibility of evolutionary rescue for slower evolutionary scenarios. Coming back to the relationship between BA and population size, it could be further clarified by adding a dynamic graph of population size throughout the simulation—for example, by replacing BA with N on Figure S7A; or plotting Figure S8 at the end of simulation to show final population sizes.

Regarding Figure S8, I have two specific comments: (1) The figure should indicate that the first line corresponds to Fir and the second to Beech. (2) It would be useful to clarify whether the population size results in Figure S8 directly correspond to the BA results in Figure 2B, as this would strengthen the interpretation.

Furthermore, while the modifications in the main text (section Data analysis and software) improve clarity, I am not entirely convinced by the justification provided. My main concern is more a matter of form than substance. Specifically:

1. The sentence L368 "*a decrease in population size translates mainly into a reduction in the number of individuals*" is tautological and does not add meaningful information.

2. Rather than stating this, the authors should explicitly justify why RBA is an appropriate proxy for population size in their study. This requires clarifying the underlying assumptions—namely, that in a population where all individuals have similar age and dimensions, a reduction in BA directly reflects a decline in population size, and that this relationship can be extended to populations with mixed age structures at equilibrium.

3. The potential issue arises post-disturbance, where young stands may have low BA despite a high number of small individuals. However, since all scenarios in the study undergo the same perturbations, RBA remains valid for comparative purposes.

Two additional minor comments:

L 365: "The ecological state of a forest" should be replaced by the "demographic state of a forest"

L376-381: this clarification is quite appropriate; you may just avoid the repetition "Regarding **evolutionary rescue**, we consider a population to have undergone **evolutionary rescue** ...

# 2. On evolutionary rates

I encourage you to revise your text regarding the use of Haldane's measure. You state that Haldane's measure quantifies only short-term evolutionary responses, yet the measure you used for long-term evolution (the difference between  $\overline{DrTol_{s,p,3000}}$  and  $\overline{DrTol_{s,p,y}}$  is mathematically identical to equation 7 (except you do not divide par trait standard deviation nor scale by generation time). It would therefore be more accurate to acknowledge that you are indeed using this measure (Haldane) both for "short-term" evolution (over 100 years) and "long-term" evolution (over 1000 to 1500 years) rather than introducing an unnecessary distinction.

I believe part of the confusion stems from the historical context: the evolutionary rate termed "Haldane" (H<sub>o</sub>) by Gingerich (1983) was named in tribute to J.B.S. Haldane but is distinct from the evolutionary rates in Darwins introduced by Haldane in 1949. Your text and response to the reviewer seem to conflate these two concepts. While you correctly point out that evolutionary trajectories may become non-linear over time due to changes in selection pressure and genetic drift, this does not mean that Haldane's measure cannot be used to assess long-term evolution—it simply means that the rate may vary over different time intervals. Instead of implying that Haldane's

measure is inherently inappropriate for long-term assessment, it would be clearer to explain that its interpretation must take into account shifts in selection regimes and evolutionary dynamics over time. I recommend revising your manuscript accordingly to improve clarity and avoid unnecessary confusion.

## MINOR COMMENTS

I have few other comments, that are all more a matter of form than substance, but still need to be addressed. *Description of regeneration* 

I am not fully satisfied with your response to Reviewer 2, who rightly points out that a major issue with the regeneration approach commonly used in gap models is that it constrains evolution, regardless of the multispecific context. More specifically, this approach eliminates **fecundity selection**—the fact that some adults have higher fecundity than others due to differences in functional trait values. However, the 'local regeneration' approach you implemented in ForCEEPS in this study addresses this limitation of gap models. Since individual reproductive success depends on individual BA, which in turn is influenced by heritable trait values, this approach allows for evolution.

So, I would suggest you to correct the text L 150-160 and by adding explicitly the following information (in red and bold) to your test (in black):

L150 : In individual-based forest PBMs, two main approaches are used to simulate regeneration. The **first approach explicitly links seedling establishment to the fecundity of mature trees**: the number of new seedlings is determined by the number of seeds produced within the simulation, followed by the application of a mortality filter. This method is commonly used in models that incorporate evolutionary processes. The second approach, more typical of gap models such as ForCEEPS, assumes a fixed seedling density per unit area, calibrated from inventory data, **regardless of the actual fecundity of mature trees**. Here too, a mortality filter is applied (Bugmann, 1996; Didion et al., 2009; Morin et al., 2021). With this approach, regeneration therefore consists in a constant annual "seed rain" **including** all species considered to be able to reproduce in the simulation, independently of their current abundance and density in the forest. This second approach strongly constraints evolution, first because it eliminates fecundity selection— the fact that some adults have higher fecundity than others due to differences in functional trait values. Moreover, it prevents interspecific competition from impacting regeneration.

To overcome these limitations, in this study we separated regeneration in two distinct processes: "local regeneration" and "migration regeneration". On the one hand, for local regeneration, and building on earlier work (Vallet unpublished) we multiplied the original target number of seedlings predicted by ForCEEPS for each species by its relative abundance (Fig. 1 teal box, arrow A). The relative abundance of species is assessed based on their Basal Area (BA) (m<sup>2</sup>/ha) across all patches, i.e. the sum of trunk areas at a height of 1.30 meters per hectare. Moreover, individual reproductive success explicitly depends on individual BA and thus on heritable trait values (see section evolution phase below), which allows accounting for fecundity selection.

On the other hand, migration regeneration....

#### On the dynamic nature of variances and heritability

L274-2781: I find these sentences overly complex and difficult to follow, even though I appreciate the authors attempt to clearly explain a complex issue. The key idea is here is that while heritability ( $h^2$ ) and trait variability ratio (VR) are set as initial conditions, the realized genetic and phenotypic variances evolve dynamically due to selection and drift. I would suggest making this explanation more concise and propose replacing all the sentences from L274 to L278 with: "While heritability ( $h^2$ ) and trait variability ratio (VR) are prescribed as input parameters to initialize trait values and introduce genetic noise in offspring (through  $\Sigma_G$ ,  $\Sigma_E$  and  $\Sigma_P$ ), the realized genetic and phenotypic variances (and accordingly, heritability) change dynamically throughout the simulation as selection and genetic drift act on the reproducing individuals".

By the way, I think that reviewer 2's remark that "Higher VR does not necessarily indicate higher genetic variation, right?" is not that naïve as it might seem: equation (2) and (3) indeed "say" that increasing VR mechanically increases both the genetic and environmental variation, but it may be worth clarifying this point, for example around L228.

#### On the description of climate periods

L 294-301: I find your explanation of the three period for each simulation clearer. But you may consider adding directly the simulation dates in the text to make it even clearer, i.e., replacing L294-298 by:

- Historical period (Years 1000-2000 in the simulation): Climate data from 1969 to 2000 (from the historical CHELSA dataset) was repeatedly used to represent 1000 years of past climate conditions.
- Warming period (Years 2000–2100 in the simulation): Climate data from 2000 to 2100, based on the RCP 4.5 scenario, was applied to simulate a century of climate change.

# • Stabilization period (Years 2100–3000 in the simulation): Climate conditions from 2090 to 2100 were cyclically repeated for 1900 years to represent a stabilized climate.

"We cannot find in the text where we would be suggesting that populations evolve over 2100 years.": This was not in the text, but in Figure S6, where Stabilization period is 1000 years long, please correct. Note that in this figure S6, the 'baseline simulations' are not indicated for Questions 2 and 3, which may be what confused Reviewer 2.

## On the simulation plan:

This section is confusing because you describe the simulation design used to address your main research questions (L332–363) after detailing the baseline simulations (L304–321) and certain aspects of the simulation setup (e.g., the combination of four scenarios for h<sup>2</sup> and four scenarios for V<sup>R</sup>, L321–331). While your chronological order reflects the sequence of simulations, it does not align with their relative importance, as the baseline simulations are secondary. I recommend reversing the order—starting with the major research questions and then providing the necessary details. Moreover, there is room for clarification in section L 304-321, particularly at L305 : here, please consider replacing "we carried out 50 replicates of what we call "baseline simulations" by simply "we carried baseline simulations without evolution (Years 0–1000 in the simulation)"; this would lead to several improvements because:

- those 50 replicates are confusing as less of them are used o answer questions 2 and 3;
- there is a risk of confusion with the 50 patches L308.
- It helps to understand that those baseline simulations occur without evolution and before the sequence you have described L294-298.

Please carefully revise you Figure S6 if needed, so that it fit with this description.

Note that many titles of the Material and method sections could be improved :

- **Model Description** is not precise as many of these other section deal with model description; maybe consider "odel overview and major changes brought in ForcEEPS"
- **Regeneration and Species Composition Feedback** may be replaced by "Regeneration and Species Composition Feedback module »
- **Evolution Module** may be replaced by "Evolution Module: evolvable traits" as it specifically described which traits can evolve
- Initialisation phase may be replaced by "Quantitative genetic model: initialization of traits values"
- Evolution phase may be replaced by "Quantitative genetic model: trait inheritance"
- Data analysis and software may be replaced by "Analyses of model output and implementation"

## **On traits coevolution**

Please consider replacing L 473 "In this context, trait coevolution between ShTol and DrTol and interspecific competition played a more decisive role in the fate of forests" by "In this context, **the trade-off between ShTol and DrTol, along with interspecific competition**, played a more decisive role **in shaping the eco-evolutionary dynamics** of forests",

The original sentence suggests that the observed relationship between ShTol and DrTol results from co-evolution, but since a negative correlation between these traits was imposed at initialization, it is unclear whether the observed pattern emerges from true co-evolutionary processes or simply reflects the initial constraint. To avoid misleading interpretations, I suggest rewording the sentence to emphasize the trade-off as a key driver of eco-evolutionary dynamics, rather than implying a co-evolutionary process that may not have occurred.

## Regarding the "Model and simulation plan limitations" section

The sentence "On a different topic, the arbitrary choice of stabilizing the climate after 100 years of warming (at the end of the predictions available in climate projections), limits conclusions on any long-term effects." Should not be removed, but you may consider softening it, for instance;

On a different topic, the arbitrary choice of stabilizing the climate after 100 years of warming (at the end of the predictions available in climate projections), may limit our conclusions, as this scenario is both arbitrary and likely unrealistic. Different post-warming trajectories (e.g., continued warming or cooling) could have led to different outcomes, highlighting the importance of halting climate change to allow species to adapt. In particular, the long-term effects observed in our study are influenced by the assumption of climate stabilization after 2100, as we noticed that...

Regarding the trade-off between growth traits and drought stress at the interspecific level, see Fririon, V., H. Davi, S. Oddou-Muratorio, F. Lebourgeois, and F. Lefèvre. 2023. Within and between population phenotypic variation in

growth vigor and sensitivity to drought stress in five temperate tree species. For. Ecol. Manage. 531. [Note that this is for your information, and not for inclusion in this manuscript, due to potential conflict of interest here]

L565please replace "assisted gene flows are also advocated" by "assisted gene flow IS also advocated