Dear Pr Guillaume,
Thank you for the careful consideration you and the reviewers gave to this manuscript. I have uploaded a revised version addressing these comments. Non-trivial changes are discussed below.

I welcome the addition of the details about the genetic correlation of the robustness traits. It shows more precisely how different aspects of robustness may evolve more or less independently from each others. It thus greatly improves the quality of the paper. However, it can be argued that the way the evolutionary constraints have been assessed is far from the way networks may evolve in natural systems in the sense that we hardly expect robustness traits to be under direct selection. Indeed, as stressed at different places in the manuscript, robustness is an emergent property of the topology of the networks, which determine the expression phenotypes under selection. Therefore, as also asked by the third reviewer, I would like to see a more thorough justification for the purpose of this approach (imposed direct selection on robustness traits) to guide the reader and avoid misinterpretations about the nature of selection on robustness you expect to see in nature. My interpretation is that you use direct selection as a device to understand the evolutionary properties of the robustness traits. This should be made clearer in the text. What should also be clarified is whether and when you'd expect direct selection on robustness traits to act in natural populations.

Thanks for noticing the lack of clarity of the reasoning. As detailed in my response to reviewer #3, the text has been modified to address this issue more directly. The main reason of the unconvincing text probably lies into the fact that my original intention was to simulate “realistic” selection on robustness, before realizing that this would be inconsistent (for instance, direct vs indirect selection) and sometimes technically challenging (how to select for more sensitivity to mutations for traits ranging in a (0,1) interval?). I hope that the revised version is more straightforward and clearer on the reasons why arbitrary, but homogeneous, selection was applied.

A discussion on that very same topic is thus also awaited, especially in the light of the difference of your model with existing models. It is not yet clear from your manuscript how the genetic constraints among robustness traits bear on the evolution of the genetic networks themselves when populations are challenged by the environment but in absence of such artificial direct selection on the robustness traits. I think this is a legitimate question since it is generally expected that robustness is a trait under second-order selection. Thus, I would like to see a general discussion on that topic to help us better understand the significance of your findings in a broader evolutionary context. Such a discussion is missing at the moment but would greatly improve the impact of your paper.

I hope that the new paragraph added in the discussion addresses this concern. I totally agree that the question is legitimate, even if it is not easy to address convincingly.

Minor Comments
l58 : hypotheses -> hypothesis

Fixed.

ll122-123, 136-137 : please verify if assumptions about constitutive expression are the same as in Ciliberti et al. 2007. Seems to me that their S(0) is similar to your P_0. Also amend accordingly comments about constitutive expression in your vs other models elsewhere in the text.

Listing intentional (and unintentional) variation among the implementations of the Wagner model in the literature would almost deserve an independent paper (probably boring for most readers), so I’ll focus only on a very few “major” papers (Wagner 1994, Wagner 1996, Siegal & Bergman 2002, and Ciliberti et al 2007). In my manuscript, the constitutive expression parameter has two distinct functions: (i) scaling the regulation function, and (ii) setting the initial state of the network. As far as I know, this is not common in the literature. I have updated the manuscript for clarity, and for the record I propose below a detailed synthesis of the model setting in a very few early Wagner-model papers:
About (a) the scaling regulation function: sticking to Wagner 1994’s notation, there is a “quantitative” version of the model (hereafter: type I model):
$P(t+1) = g(W P(t) + e)$,
in which $g(x) = 1/(1+\exp(-cx))$ is a sigmoid scaling function ($0 <= g(x) <= 1$), and e is the constitutive expression (“basal transcription rate”), possibly different for each gene. In the very same paper, Wagner 1994 also proposes a “discrete” version of the model (type II):
$S(t+1) = \text{sigma}(W S(t))$,
in which sigma(x) is the sign function -> ($x < 0$: -1, $x = 0$: 0, $x > 0$: 1), constitutive expression e is explicitly set to 0 (i.e. the mid-phenotype) in this model.
As far as I can tell, both papers by Ciliberti et al 2007 (*Innovation and robustness in complex regulatory gene networks*, in PNAS, and *Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology* in PLoS Comp Biol) use the type II model of Wagner 1994, featuring a step (sign) function sigma(x) with no explicit

constitutive expression (sigma(0) = 0, i.e. the expression is exactly between -1 and 1 in absence of regulation on the gene).
About (b) the initial state of the model, P(0) or S(0): in Wagner 1994, S(0) is drawn randomly at the beginning of each simulation (the probability Prob(S(0) = 1). In Ciliberti et al 2007a, the authors write that it is set "at random", and no more details in Ciliberti et al 2007b (I assume they use the same setting as in Wagner 1994 with Prob(S(0) = 1) = 1/2). In this setting, I do not think that putting S(0) at the constitutive expression level (0) is possible, as this would be a stable state (W S(t) = 0 whatever W, and thus S(t+1) = S(t) = 0, which is in my opinion a strong argument against using this S (Wagner's type II model) setting, but this is not the purpose of my manuscript).
In Siegal & Bergman 2002, the model is "hybrid" between Wagner's models 1 and 2. They use a gene expression S(t) scaled between -1 and 1, but the sigma function is continuous (sigma(x) = 2/(1+exp(-cx))-1). The parameter c (which they call 'a', but I'll stick to Wagner's "c" notation not to confuse with my own 'a') scales the sigmoid (and the sigma function becomes the sign function when c -> infinity), but there is no explicit constitutive expression, and sigma(0) = 0, which is the "mid-point" between min and max expressions. They explicitly set the initial state as extreme expressions, randomly : *"The initial state, S(0), is constant for each simulation and is set by randomly choosing each S = -1 or +1."*, as in Wagner 1994 and Ciliberti et al 2007a and 2007b.

The above paragraph is now summarized in the manuscript (in the methods and discussion sections).

ll385-387 : is it 1000 generations or 5000 as in the legend of Figure 4?

This was indeed confusing. White panels in Figure 4 show 5000 generations of selection response, but the colored panel shows the response after 1000 generations, which is more correlated to the prediction than the 5000-generation response (Appendix 8). Since the figure mixes two evolutionary times (which is hopefully clear from the figure legend), the simpler "fix" is to delete the confusing reference to the number of generations from the text ("*... and selection responses were compared to the mutational evolvabilities computed at the beginning of the simulations.*")

Figure 4:
Caption: illstrated -> illustrated;

Sorry for the typo, fixed.

A legend for the meaning of the gray arrows in the top-right figure representing join selection on the two traits should be given to better link with the text (ie. "bivariate selection" is used in the text) (see Mullon's comment).

The legend now indicates: "*Each panel displays the selection response in eight directions, as illustrated in the legend (four univariate --- colored arrows --- and four bivariate --- gray arrows --- gradients of selection, same color code as in Figure 3).*"

The hyphenated line in the predicted-observed evolvability graph gives the wrong impression that the two evolvabilities match, while they don't. A 1-1 line would rightly show that the observed evolvability is larger than its predicted value. This should be stressed in the text as well.

Using M instead of G was indeed discussed in my last response to the reviewers, but much less in the manuscript. One should not expect a 1:1 match between the mutational evolvability and the response, as the 1:1 match is expected for G, not for M. Predicting the G/M ratio would require to derive a multivariate drift - directional selection - mutation model, which seems out of reach in this context. I have now introduced the following comment in the Methods section: *"Contrary to the genetic covariances G, mutational covariances M cannot be used directly to compute quantitative evolutionary predictions, as the relationship between M and G depend on the mutation-selection-drift equilibrium, which is notoriously difficult to handle theoretically (Bürger & Lande 1994). The following analyses thus focus on whether mutational evolvabiities are proportional to the selection responses, assuming that G is proportional to M."*
The legend of Fig 4 now reads: *"The colored inset illustrates the proportionality between the predicted mutational evolvability [...] and the observed evolvability ..."* for clarity. The reference to a proportional relationship was already present in the text.

Figure 5:
The direction of change as function of direction of selection is very hard to find, maybe split the graphs to make the patterns more discernible?

Figure 5 has been split. New fig 5 now shows only the evolution of correlations for univariate selection, and new fig S9 shows the consequences of bivariate selection. In addition, a new panel in figure 5 illustrates the fact that correlation does not respond directly to the direction of selection. Thanks for the suggestion, splitting the figure makes it easier to formulate the conclusions.

ll457-459 : avoid repetition of "cumulative"; specify that the cumulative effect is on the phenotype (=gene expression level); maybe specify that evolution is "gradual"?

Done

l468 : dfferent -> different

Done

ll507-508 : not sure which mathematical model is referred to here

The sentence has been reformulated as "*Yet, even if these measurements happen to be correlated by construction, their partial evolutionary independence highlights their potential for independent evolvability in real gene network architectures, which are substantially more complex and subtle than our gene network model.* "

l511 : catch -> capture

Done

**Review by Diogo Melo**

This review answers most of the comments I had on the previous version, and the inclusion of the M matrix exploring the complex evolution of the correlation between robustness measures is a welcome addition.

On the confusion related to correlated selection, the author is of course correct. This comment was just my (admittedly confusing) way of saying that the selection differential on one single trait could lead to a selection gradient on more traits due to phenotypic correlations. Dealing with the gradients directly solves this and the text is clear.

Thanks for clarifying. Perhaps I should add that I was referring to theoretical gradients (the slope of the fitness function). "Real" gradients (measured in the simulated population, which I haven't done), may not match exactly the intended direction (because of sampling, or if robustness components were phenotypically correlated with selected gene expressions). As suggested by the recommender, the legend of figure 4 has been updated to differentiate univariate and bivariate gradients of selection.

Some typos:
P2: refered -> referred
P14 asymetrical -> asymmetrical
P14 dfferent -> different
P15 controling -> controlling
P15 sparcity -> sparsity
P15 reductionnist -> reductionist
P15 occuring -> occurring
P15 Althought -> Although
P16 tigh -> tight

Fixed. Sorry for this, I should have of course run a spell-checker.

**Review by Charles Mullon, 14 Dec 2021 10:36**
The author has done a great job at adressing all the points previously raised, and I believe that the manuscript has much improved. I also appreciate the careful explanation provided in the response to reviewers (point 3 in particular). I only have minor comments remaining.
p. 7. To ensure clarity, please specify how response to selection R is computed.

The methods now state: "The response to direct or indirect selection was computed as the average change from generation 0; the multivariate response was stored as a 5-dimension vector R"

p. 9. Top. Here you refer to the "stability" of networks A-E presented in Appendix 6. However nowhere in Appendix 6 is the stability measure (rho_S) given. I am guessing you are referring to a more qualitative view of stability here. This is confusing as in the rest of the manuscript, stability is defined quantiatively.

Indeed, Appendix 6 illustrate the network dynamics in replicated runs used to estimate 4 robustness components. Stability is also estimated quantitatively, but this estimation does not require any resampling procedure; stability is

computed as the temporal variance of gene expressions during the last time steps. It can be visually quantified by assessing how “cyclic” are gene expressions for the five undisturbed networks (thick lines) illustrated in Appendix 6, but it does not need a column of its own.
For clarity, robustness indicators are now indicated in some of the panels.
The legend of Appendix 6 has been modified: “Four (out of five) robustness measurements rely on a resampling procedure (corresponding to the four columns of the figure). [...] The network stability can be assessed from the amplitude of the cycles in the undisturbed kinetics (thick lines), and does not rely on a stochastic algorithm.”

p.10. I found the phrase "the evolution of correlations was driven by the direction of selection (more or less robustness), and not by the orientation of the selection gradient relative to the main evolvability axis" difficult to connect with Figure 4. Could you please clarify how you come to this conclusion?

This was indeed unclear – note that the sentence was referring to figure 5, not figure 4. Figure 5 has been split into Fig 5 and sup fig 9, and a new panel has been added to figure 5. This will hopefully make it clearer how to connect the claims in the text to the results. The take-home message of the evolution of correlations is that correlations evolve, but that this evolution does not reflect directly the selection pattern. The paragraph has been reformulated for clarity:
“ The evolution of correlations was partially driven by the direction of selection (more or less robustness). Within each specific pair of robustness components, the evolution of correlation was rather consistent [...]. Yet, there was no general pattern associating the evolution of robustness and the evolution of correlation; depending on the robustness component, selecting for more or less robust networks may increase or decrease the correlations (colored inset in Figure 5). There was no effect of joint selection; selecting together two robustness components did not make them more (or less) correlated (Figure S9).“

Figure 4. This figure is a little difficult to parse. I think the main difficutly I had was to understand the meaning of "the direction of selection". I think once this is defined properly (see above), it would be good to explain in the legend of Figure 4 what the different axes correspond to (perhaps with an example).

Axis titles have been added in the small inset legend of figure 4, and the inset title has been changed to “Selection gradients”.

Figure 5. There are too many overlaying symbols in this figure. Would it be possible to separate Figure 5 into two or three subfigures (and perhaps moving some to an appendix, e.g. correlated vs antagonistic selection)?

Figure 5 has been split (the second part is now Figure S9). Thanks for the suggestion.

p. 14 "dfferent" -> different.

Fixed.

p. 15. Can you please define what you mean by "network complexity"?

Now defined as “size and number of connections”.

p. 15. "can not" -> cannot.

Done.

p. 16. "tigh" -> tight.

Done.

**Review by anonymous reviewer, 11 Jan 2022 09:11**
In this manuscript, the various facets of gene network robustness (environmental and genetic robustness) are treated as quantitative characters, and studied in a quantitative genetics framework.
The author uses a regulation network model based on the Wagner model and numerical simulations to show that four different measures of robustness (two measures of environmental robustness, two measures of mutational robustness, as well as one measure of network stability) are partially correlated. This suggests that, while not independent, gene network robustness characters can evolve towards many directions, as confirmed by numerical simulations where robustness characters are under direct positive or negative selection. An analysis of the M-matrix then confirms that mutational correlations between the robustness characters can facilitate or prevent their evolution, and that this multivariate response to selection is predictable.
By quantifying more systematically the multi-dimensional nature of gene network robustness, this work highlights the ability for a genetic network to acquire various levels of robustness depending on selective pressures and mutational

correlations between robustness components. Doing so, it provides the reader with a more synthetic vision of what is robustness, how to quantify it and how it evolves.
I appreciated reading this manuscript which nicely tackles a difficult theoretical question. Methods are clearly presented, which allows for a smooth reading. In particular, the usage of a quantitative genetics framework to describe robustness components is an important contribution.

Thank you for your careful and critical reading.

The manuscript has already been improved by a first round of review. I have a few suggestions and comments that can be easily addressed:
- The purpose of numerical simulations in section 3.3 could be better presented. In my opinion, these simulations are here to demonstrate the evolvability of robustness under a direct "abstract" selection on robustness components. As such, it should be clearly stated that stabilising selection simulations (where no perturbations are introduced except inherited mutations) are some sort of control, and should not be understood as a proof of the ability/inability of robustness to evolve indirectly in a more realistic setting.

On this point, it would have been interesting to run additional numerical simulations to study the indirect evolution of robustness against environmental and/or genetic perturbations, and to confront simulation results to the predictions made. However it is another manuscript in itself, and I certainly do not ask the author to do so.

Thanks for rising this point. This was indeed probably unclear in the manuscript, because my original intention was rather to focus on more “natural” selection pressures. However, it quickly appeared that manipulating indirect selection was very difficult. For instance, when applying stabilizing selection on traits, one generates both direct selection on environmental canalization, and indirect selection on genetic canalization. Even worse, in most cases, there is no way to perform bidirectional selection (towards more or less canalization), and up vs down selection on the same robustness component can be either direct or indirect. Therefore, even if such simulations would have been of direct relevance for the evolution of robustness, the simulation setup would have been incomplete and probably awkward.

The relevant subsection of the results now mentions: “Such direct, artificial selection pressures on robustness are not designed to reflect realistic selection on gene networks, but they might reveal evolutionary limits to the evolution of robustness due to internal constraints.”

A whole new paragraph has been added to the discussion: “*In the simulations, selection on robustness components was direct and constant both in up and down directions (i.e. towards more or less robust genetic architectures). This setting was not expected to reflect realistic evolutionary pressures on robustness, which might be more complex, overlapping, and asymmetric. Stabilizing selection, for instance, selects both directly for robustness to environment, and indirectly for robustness to mutations (Wagner et al., 1997); selection for stability also promotes indirectly robustness to mutations (Siegal and Bergman, 2002). Conversely, selecting for lower robustness through the phenotype may be difficult or even impossible: fluctuating selection does not promote decanalized genetic architectures (Le Rouzic et al., 2013), and selection for environmental sensitivity is limited by the inaccuracy of the perception of the envrionmental signal (Reed et al., 2010). Simulation results thus illustrate how robustness components may evolve independently when individually selected; whether or not there exists realistic conditions for such selection pressures is a different — and more complicated — issue.*”

Minor comments:
---
- Introduction: As far as I know, some authors have attempted to quantify robustness in multiple dimensions (see e.g. https://doi.org/10.1371/journal.pone.0000434).

Thank you for the reference. This was an interesting part of the literature I missed by focusing on the Wagner model. I included new references towards relevant reviews and models.

"In addition, in models where the genotype-phenotype association is arbitrary, [...]": please provide references

Two well-known ‘arbitrary epistatic GP map’ frameworks are now cited (the NK model by Kauffmann & Levin, and the multilinear model by Hansen & Wagner). I am not aware of any attempt to introduce plasticity in such models.

"to an element of the W matrix": Is a single weight selected at random in the matrix?

The sentence now reads “Mutations consisted in adding a random Gaussian deviate of variance sigma^2 to a random regulatory interaction of the W matrix, with a rate nu per individual and per generation.”

"By default n=6 [...] By default n'=3": Does this parametrization influence the evolution of robustness (compared e.g. to the case n'=6)? Similarly the mutation rate sigma_v^2 is probably an important factor triggering the indirect evolution of mutational robustness. It has been treated by the author in Appendix 7, but I think a few words on the choice of parameter values is needed.

As the simulation runs displayed in figures 3 to 5 are computationally-demanding, testing different parameter settings (especially larger networks) would be in practice complicated and costly. The reviewer noted two pairs of parameters that can be commented on separately.

* n and n': The network size (n) had to be limited for computational reasons: the cost of computing network dynamics is in O(n^2) (square regulatory matrix), so n is the most critical parameter in terms of simulation time. As confirmed by test simulations in Appendix 7, increasing n beyond 6 does not really change the equilibrium robustness level -- its impact on the evolution of robustness in the various scenarios of figures 3, 4, and 5 has not been tested directly. The number of selected genes n' is an arbitrary parameter (little is known about selection at the gene expression level). Earlier work with similar models (e.g. Rünneburger and Le Rouzic 2016) suggested that large n' could slow down the evolution of emergent properties of the network (when all gene expressions are constrained, evolving the network structure is more complex as no single mutation is neutral). From Appendix 7 column 6, the effect of n' on the evolution of robustness in the default setting (no direct selection on robustness) is tiny (if any).

* Mutational parameters (size and frequency of mutations): as noticed by the reviewer, they generate an indirect selection pressure on the evolution of genetic robustness. This was confirmed by simulations (Appendix 7 column 1 and 2); note that the default parameters were chosen as large as possible (to speed up simulations) but small enough so that the indirect selection has no visible effect. The manuscript already had a sentence to indicate this: "*Stabilizing selection on gene expression is expected to generate a slight selection pressure on the robustness, but this effect was apparent only for larger or more frequent mutations (Appendix 7).*"
The manuscript was updated to better justify default parameters and comment the results from Appendix 7. :
In the methods section:
"*[...] by default n=6 to limit the computational burden*"
"*Mutational parameters nu and sigma_nu were kept reasonably low to limit the strength of indirect selection for genetic robustness (Wagner et al 1997, Rünneburger and Le Rouzic 2016).*"
In the discussion:
"*In particular, the network size n and the number of selected genes n' do not alter drastically robustness components, showing that small regulatory motifs are not qualitatively different from large gene networks in terms of robustness.* "

- Table 2: Indicates 5000G. But 10,000G is used after (section 3.3).

This was indeed confusing, as simulations for different figures were run for a different number of generations. The number of generations was removed from table 2, and clearly indicated in the legend of the figures.

- Figure 1: Panel A is not so easy to read with the floating labels. Maybe use small horizontal barplots with a color code?

Done, thanks for the suggestion.

"is robust to the way random networks are generated": All the random networks are generated in the same way, the generation mode of the networks is not explored here.

The sentence has been reformulated as "The part of the variance in robustness explained by the first PC is robust to the network properties, as it remains around 80% when the mean and the variance in the regulation strengths, the network density, and the network size vary."