nature portfolio

Peer Review File

Unifying framework explaining how parental regulatory divergence can drive gene expression in hybrids and allopolyploids

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Janko and colleagues, building on earlier work by Bottani et al, introduce a mechanistic physiochemical model of gene regulation in the context of hybrids and polyploid organisms, incorporating regulatory divergence and asymmetry in cell size and euchramatic states to investigate an array of empirical observations like transcriptome downsizing, gend dosage effects, expression-level dominance, and subgenome dominance, that are currently major focuses of research on polyploid genomics and evolution. They model expression in hybrids and polyploids under multiple situations that assess the impact of unbound transcription factors, cross-regulation, different underlying processes that make TF-promoter binding more specific to a given subgenome, and identify several interesting relationships between these factors and expression patterns in hybrids and compare the results that suggest the patterns largely stem from from their model to several empirical observations, highlighting concordance between the two as well as areas where the models provide interesting predictions for subsequent empirical studies.

I am extremely excited about this paper and the framework it introduces and think the improvements on past work in this area will enhance the applicability and impact. Genomic and evolutionary research of newly formed hybrids and allopolyploids has been a field that has produced rich amounts of data and interesting observations but with little structure or guidance from robust quantitative theory to guide experiments or synthesize observations. The presented models appear to provide just such a foundation and I think have the potential to make a tremendous impact on this field and our understanding of genome evolution. It already seems to show with even more clarity how regulatory divergence of parental lines can provide at least a partial unified explanation for transcriptome downsizing, subgenome dominance, and expression-level dominance, as well as how specific factors like promoter vs TF adaptation to asymmetries in euchromatin and asymmetry in cell size may provide some taxon-specific factors that can explain heterogeneity in observed patterns of gene expression across hybrid/polyploid taxa. I think this work would be a great addition to the journal, but think there are some minor, but important, changes required before acceptance, which I describe below.

If there is one thing that I imagine can hinder this paper from being taken up by the field with great impact it is that it might appear quite dense or incomprehensible to many biologists who do not work frequently with these kinds of physiochemical models, as will be the case for most polyploid genomics researchers. In my opinion writing or presenting the data with that audience in mind will be very important for this paper. As the manuscript is, while I found the text to generally be easily understandable it took me quite a bit of time looking at the figures to identify the results/conclusions that I read in the text.

I hope this does not come across as nitpicking, I'm simply imagining that the difficulty I experienced interpreting the figures may be shared by other biologists. As a reviewer I had no problem spending many minutes breaking down the figures until I could understand them but as a reader giving the paper a skim, it would likely be passed over as too opaque.

Some of these concerns might be simply addressed, for example, it was not immediately obvious to me that both the shapes, line patterns, and colors applied to all figures/subfigures for Figures 2, and 3. Reconfiguring the figure legends may help this.

It's also likely that the axis labels will be difficult to interpret for some biologists as f_A, f_B, fA_H, and fB_H might not be immediately understood as the expression in parents and hybrids and so it might not be clear that the axes are the expression fold-change in parents against the expression foldchange in hybrids. I believe some of this confusion might be that the model is really predicting promoter occupancy and not expressions directly, but for the purpose of interpretability, it may be worth either providing a clearer explanation early in the text to guide readers or changing the label to more clearly

communicate the intended measures.

The figures also present several dimensions of information on top of each other (shapes, lines, colors). I wonder if it's even necessary to include the different shapes in Figures 2A,C Figures 3A,C,E, and Figure 4 considering the shapes are just the x-axis without log2 transformation. This could make the figures less overwhelming at first look.

Alternatively, the authors may want to think about a conceptual figure to demonstrate how specific patterns are expected to look in the plots, I found Figure S2 to be quite helpful in this regard.

While this might be frustrating, I think paying extra attention to making this paper as easy to interpret as possible will help it make the impact I believe it deserves.

Reviewer #2 (Remarks to the Author):

Janko et al. present a thermodynamic model to estimate transcription factor (TF)-promoter binding and investigate genome dominance in allopolyploids. They show that even if the hybrid preserves the same number of TF molecules per cell, the concentration of TFs with respect to their homologous promotor sites is halved and the fractional occupancy of promotors on each subgenome is lower than original parental value. They also explore scenarios where the overall gene expression in hybrids and allopolyploids deviate from additivity. Overall, I found that this manuscript is interesting and presents a systematic exploration of parameter space of a thermodynamic models of genome dominance in hybrids.

Main concerns:
The authors make the following claim: "By setting both $C1$ and $C2$ to one, Bottani et al. implicitly assume that each specific binding site binds with its characteristic energy to all TFs, regardless of their parental origin. This energy may be specific for sites of A- or B- types (i.e., possibly $E1 \neq E2$), but it does not differentiate between TF molecules from both parental species (i.e., $E1 = E21$ and $E2 = E12$). This assumption contradicts their original claim that adaptation to higher non-specific site concentrations occurs through modification of TFs themselves. Instead, the model in fact assumes that the higher binding affinity of specific sites in parent B occurs through modification of the binding site."

In Bottani et al. it is mentioned: "More specifically, we assume that the two merging genomes have different sizes. Homoeologs targets G1 and G2 in each genome will be regulated by the paralogous TFs, TF1 and TF2, with different affinities (TFi regulates/interacts with Gi in subgenome i with a dissociation constant Ki). Depending on the scenarios, allowing cross-regulation or not, TF1 will also bind to the promoter of G2 and TF2 to the promoter of G1. For simplicity, in the following we deal with target promoters having only one binding site for the TFs."

Maybe I am missing something, but I am not sure what the authors want to say. I think they misinterpreted our statement from Bottani et al., where the authors explored different limiting scenarios "This coefficient is such that C1 = 1 for full cross-regulation, and C1 = 0 in the unrealistic limit of no cross-regulation (see the Supplemental Information online). "

Dear Editor and both reviewers,

Thank you for thorough reading and commenting of our MS. We appreciate the relevancy of sent comments and in the following text, we address each point and describe how and where your suggestions were incorporated to the revised MS. For your easy orientation, the original text from the decision letter is left in standard font, while our responses are introduced by "Reply" and indicated in *italics*. Where possible, we always refer to pages and lines of the MS where appropriate change may be found (we use format p. X: l. Y-YY for pages and lines, respectively).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Janko and colleagues, building on earlier work by Bottani et al, introduce a mechanistic physiochemical model of gene regulation in the context of hybrids and polyploid organisms, incorporating regulatory divergence and asymmetry in cell size and euchromatic states to investigate an array of empirical observations like transcriptome downsizing, gene dosage effects, expression-level dominance, and subgenome dominance, that are currently major focuses of research on polyploid genomics and evolution. They model expression in hybrids and polyploids under multiple situations that assess the impact of unbound transcription factors, cross-regulation, different underlying processes that make TF-promoter binding more specific to a given subgenome, and identify several interesting relationships between these factors and expression patterns in hybrids and compare the results that suggest the patterns largely stem from their model to several empirical observations, highlighting concordance between the two as well as areas where the models provide interesting predictions for subsequent empirical studies.

I am extremely excited about this paper and the framework it introduces and think the improvements on past work in this area will enhance the applicability and impact. Genomic and evolutionary research of newly formed hybrids and allopolyploids has been a field that has produced rich amounts of data and interesting observations but with little structure or guidance from robust quantitative theory to guide experiments or synthesize observations. The presented models appear to provide just such a foundation and I think have the potential to make a tremendous impact on this field and our understanding of genome evolution. It already seems to show with even more clarity how regulatory divergence of parental lines can provide at least a partial unified explanation for transcriptome downsizing, subgenome dominance, and expression-level dominance, as well as how specific factors like promoter vs TF adaptation to asymmetries in euchromatin and asymmetry in cell size may provide some taxon-specific factors that can explain heterogeneity in observed patterns of gene expression across hybrid/polyploid taxa. I think this work would be a great addition to the journal, but think there are some minor, but important, changes required before acceptance, which I describe below.

Reply:

We deeply appreciate your encouraging feedback, which motivates us to continue our research in this field. Thank you for recognizing the potential impact of our work.

If there is one thing that I imagine can hinder this paper from being taken up by the field with great impact it is that it might appear quite dense or incomprehensible to many biologists who do not work frequently with these kinds of physiochemical models, as will be the case for most polyploid genomics researchers. In my opinion writing or presenting the data with that audience in mind will be very important for this paper. As the manuscript is, while I found the text to generally be easily understandable it took me quite a bit of time looking at the figures to identify the results/conclusions that I read in the text.

Reply:

We concur that our initial manuscript focused intensely on presenting our findings, perhaps at the expense of broader readability. To address this, we have undertaken several conceptually important changes. We then had our MS checked by several biologists without chemical/mathematical background and also by a chemist without biological background to make sure it is understandable to both types of readers. Addressing their questions, we believe that performed changes make our revised MS much more accessible to most readers from biological disciplines, in line with your suggestions.

Details of changes are provided in the point-by point reply below but we provide their summary here for clarity:

1) *We have realigned the opening paragraph of the introduction to mirror the structure of the Results & Discussion section more closely, thus enhancing coherence and readability. It now clearly distinguishes transcription-wide effects from allele-specific (subgenome-specific) ones (p. 3, l. 45 - 69).*
2) *We emphasized already in the introduction how the fractional occupancy may be used as a proxy for gene expression in our and related models (p. 4, l. lines 110 - 112 & 118).*
3) *We substantially modified the figures to achieve much greater clarity, also taking into account their "didactic" value in order to allow the readers easy interpretation. Specifically, we now provide:*
    a. *Clear summary of the concept of transcription categories and gene expression patterns, which are discussed and used subsequently throughout the text (Fig. 1a and further explanation in the appendix 1). Simplified example of results is also provided on Fig. 1c, which provides guidance to readers to link our Figures 3-5 with simulation results and transcription profiles of hybrids of various ploidies.*
    b. *Clear description of philosophy behind allele-specific expression and type of cis-/ trans- regulation (Fig.1b), which provides guidance to interpret simulated results presented on Figs 6&7.*
    c. *Visual link between employed mathematical equations and their biological meaning is now provided on Fig. 2,*
    d. *Simpler figures with less symbols and lines.*
4) *We also made some changes to the Results & Discussion sections to make the statements less complicated and easier to interpret.*

I hope this does not come across as nitpicking, I'm simply imagining that the difficulty I experienced interpreting the figures may be shared by other biologists. As a reviewer I had no problem spending many minutes breaking down the figures until I could understand them but as a reader giving the paper a skim, it would likely be passed over as too opaque.
Some of these concerns might be simply addressed, for example, it was not immediately obvious to me that both the shapes, line patterns, and colors applied to all figures/subfigures for Figures 2, and 3. Reconfiguring the figure legends may help this.

Reply:

Acknowledging your suggestion, we have introduced a new Figure 1. This figure visually explains transcription categories (Fig. 1a) and offers a simplified demonstration of our results for diploids only (Fig. 1c). This gives the reader an easy insight into the type of data presentation, which are later shown in Figures 3-5 in full complexity.

As you suggested, on the full Figures 3-5, we reduced the number of symbols leading to greater clarity and less symbol overlap. Furthermore, an insert of Fig. 3a now clearly demonstrates how our simulation results align with expression level categories.

We are aware that our simplification on Fig.1c causes a bit of redundancy with Figure 3a, but we believe that this adequately addresses your concern and makes useful service to readers. All these changes have also been described concisely in the figure legends.

It's also likely that the axis labels will be difficult to interpret for some biologists as f_A, f_B, fA_H, and fB_H might not be immediately understood as the expression in parents and hybrids and so it might not be clear that the axes are the expression fold-change in parents against the expression foldchange in hybrids.

Reply:

In agreement with your feedback, we have revised the axis labels across all relevant figures (Figs. 3-7) to enhance clarity. These labels now include both the technical and common terms, aiding in interpretation, such as "interparental expression divergence:", "total expression:" and/or "RAE:", depending on the context.

Also, we made new Fig. 1 panels b and c, where the link between common sense biological terms, like allele A_{hyb}, and models' output, like as f_A *occupancies are provided, to give the readers better guidance for easier interpretation of results.*

I believe some of this confusion might be that the model is really predicting promoter occupancy and not expressions directly, but for the purpose of interpretability, it may be worth either providing a clearer explanation early in the text to guide readers or changing the label to more clearly communicate the intended measures.

Reply:

As you suggest, we now provide clearer explanation for the link between promoter occupancy and gene expression. It was mentioned in the original submission, but we now explicitly added that our models stem from thermodynamic suite of models, also used by Bottani et al., which describe TF-promoter binding as a proxy to gene expression. This is now emphasized in the introduction (see p. 3, l. 45 - 69), Discussion (p. 13, l. 459 - 467) and Methods (p. 15, l. 493).

In addition, Fig. 2 indicates visually how promoter-TF binding is linked to gene expression, which is now also demonstrated on Fig. 1b, which indicates correspondence between the classical log-fold change and the fractional occupancies used in our model (please see modified axes labels)

The figures also present several dimensions of information on top of each other (shapes, lines, colors). I wonder if it's even necessary to include the different shapes in Figures 2A,C Figures 3A,C,E, and Figure 4 considering the shapes are just the x-axis without log2 transformation. This could make the figures less overwhelming at first look.

Reply:

In line with your suggestion, we have streamlined the figures to reduce their complexity. We retained for clarity only key symbols at extreme log2 values -1 and 1, to provide a link between left and right panels, and we removed others that contributed to visual overload, please see Figures 3-7. However, please note that due to the standard use of the log2 transformation in RNAseq analyses, we used it in our figures too for the sake of clarity.

Alternatively, the authors may want to think about a conceptual figure to demonstrate how specific patterns are expected to look in the plots, I found Figure S2 to be quite helpful in this regard.

Reply:

We appreciate this excellent suggestion and have implemented it through two key changes. Firstly, we provide a new Figure 2 for a clearer linkage between biological representations and applied equations. For that purpose, we introduced a simplified set of equations describing a situation in a single species (eq. 4-6) into this figure to show how all model parameters are linked with the biological scenarios.

Second, we prepared a new Figure 1, which is now used to link standardly applied categories of gene expression sensu Shi et al. 2012 with our simulation results, please see Fig. 1 a & c. This way, the reader may easily get a concept of all subsequent figures used to depict changes of gene expression levels with regard to divergence of trans-regulatory elements. (Note that here we explicitly keep in mind the "loose" link between fractional occupancy and gene expression, please see above for explanation). Finally, we took into account your suggestion and combined Figure S2 with past Figures 5&6 (Now Figures 6&7), because it represents same type of information and makes clear link between expectation and results. To achieve this, we constructed a panel B in the Fig. 1 to deliver such an information to the reader early in the text.

While this might be frustrating, I think paying extra attention to making this paper as easy to interpret as possible will help it make the impact I believe it deserves.

Reply:

We are thankful for your constructive suggestions. Far from finding them frustrating, we believe they have significantly enhanced the accessibility and enjoyment of our text for a wider audience.

As mentioned above, we had our revised version and graphics checked by "naive" readers from both the biological and chemical fields, making us feel confident that we improved the clarity of our MS.

Reviewer #2 (Remarks to the Author):

Janko et al. present a thermodynamic model to estimate transcription factor (TF)-promoter binding and investigate genome dominance in allopolyploids. They show that even if the hybrid preserves the same number of TF molecules per cell, the concentration of TFs with respect to their homologous promotor sites is halved and the fractional occupancy of promotors on each subgenome is lower than original parental value. They also explore scenarios where the overall gene expression in hybrids and allopolyploids deviate from additivity. Overall, I found that this manuscript is interesting and presents a systematic exploration of parameter space of a thermodynamic models of genome dominance in hybrids.

Reply:

We sincerely appreciate your positive evaluation of our manuscript and your recognition of its systematic exploration in the field of genome dominance in hybrids. Your insightful comments listed below have not only been constructive for refining our study but also led us to surmise that you might be one of the authors of the influential Bottani et al. 2018 paper.

If this is indeed the case, we would like to express our deep appreciation for the work in Bottani et al. 2018, as well as in previous paper Veitia et al. 2013, both of which were pivotal in shaping our research, particularly in understanding the role of regulatory divergence among parental species in gene expression in hybrids and poly(aneu)ploids. While our initial investigations were empirical focusing on our model organisms, the foundational insights from these papers motivated us to explore a theoretical approach. This shift was crucial in formulating the hypotheses of our current study, for which we are grateful.

Main concerns:
The authors make the following claim: "By setting both $C1$ and $C2$ to one, Bottani et al. implicitly assume that each specific binding site binds with its characteristic energy to all TFs, regardless of their parental origin. This energy may be specific for sites of A- or B- types (i.e., possibly $E1 \neq E2$), but it does not differentiate between TF molecules from both parental species (i.e., $E1 = E21$ and $E2 = E12$). This assumption contradicts their original claim that adaptation to higher non-specific site concentrations occurs through modification of TFs themselves. Instead, the model in fact

assumes that the higher binding affinity of specific sites in parent B occurs through modification of the binding site."

In Bottani et al. it is mentioned: "More specifically, we assume that the two merging genomes have different sizes. Homoeologs targets G1 and G2 in each genome will be regulated by the paralogous TFs, TF1 and TF2, with different affinities (TFi regulates/interacts with Gi in subgenome i with a dissociation constant Ki). Depending on the scenarios, allowing cross-regulation or not, TF1 will also bind to the promoter of G2 and TF2 to the promoter of G1. For simplicity, in the following we deal with target promoters having only one binding site for the TFs."

Maybe I am missing something, but I am not sure what the authors want to say. I think they misinterpreted our statement from Bottani et al., where the authors explored different limiting scenarios "This coefficient is such that C1 = 1 for full cross-regulation, and C1 = 0 in the unrealistic limit of no cross-regulation (see the Supplemental Information online). "

Reply:

We apologize if our commentary on Bottani et al.'s work appeared unclear or was misinterpreted. Our intention is not to detract from the value of Bottani et al.'s model but to emphasize that certain conclusions drawn from it may be more context-specific than initially presented, which we have discussed more broadly in our paper.

We recognize that Bottani et al. indeed explored various scenarios of cross-regulation in their modeling, including full cross-regulation and no cross-regulation. For the purpose of our model, our interpretation of "no cross-regulation" differs slightly. Contrary to Bottani et al., who assumed no binding between TFs and orthologous promoters in such a scenario, we propose a situation where a transcription factor (TF) binds to orthologous promoters with the same affinity as to any other non-specific sites, a variation we consider biologically plausible (please see p. 10, l. 339 – 344 and p. 20, l. 679 - 683).

The main point of our discussion centers on the assumptions in Bottani et al.'s model regarding the evolution of regulatory components (TFs and binding sites) in species with a higher euchromatic ratio. Specifically, Bottani et al. suggest that species with higher level of euchromatin may evolve TFs with generally higher binding efficiency (this is mentioned on Page 400, 2nd paragraph and page 401 1st paragraph of the original publication), leading to subgenome dominance in hybrids, even under full cross-regulation. Our analysis, however, indicates that subgenome dominance under full cross-regulation would not occur if only TFs were modified. In our opinion and simulation results, this is because the modified TFs would bind to both sets of promoters with the same, albeit higher, affinity. We demonstrate by simulation that subgenome dominance under full cross-regulation is achievable only when binding sites in one parental genome are modified, leading to their higher fractional occupancies in a hybrid nucleus (see especially the section 2.2.a., p. 10, l. 329 - 336 and Figures 6D vs.6E).

As outlined in our methods section and in Appendix 2, we elaborated on model used in Bottani et al. and find that used equations may actually also be interpreted as simulating a scenario where the species with higher euchromatic ratio modifies its promoter sequences, not the TFs themselves. This might have caused some sort of discrepancy between the verbal description in Bottani et al. 2018 (Page 400, 2nd paragraph and page 401 1st paragraph) and the actual modeling in the equations. We now discuss how such model assumptions relate to biological predictions for hybrid and polyploid organisms.

To clarify, our aim is to provide an analytical perspective, showing how different modelling assumptions yield varying outcomes. We have now detailed these points in p. 10, l. 339 – 344 and p. 20, l. 679 - 683 where we discuss our interpretation of binding under no-cross-regulation and in p. 5, l. 140 – 146 (introduction), p. 8 – 9, l. 268-286 & p. 10, l. 326 - 336 (results), p. 15 – 16, l. 524-557 (Methods) and in Appendix 2, where we elaborate on why we consider important to investigate both types of the assumed modification of binding sites versus TF molecules in case of asymmetric euchromatic ratios.

We trust that these clarifications in our revised manuscript effectively address your concerns.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I appreciate the authors' thorough incorporation of previous suggestions. I find the changes made markedly improve the clarity of the manuscript and am satisfied with the revisions. I'm excited to see this work published!

One minor change to consider, if there is time, is to alter the color palette of the graphs to be colorblind friendly or at least avoid red/green combinations given the prevalence of red/green colorblindness. However, this is far from a reason to greatly delay publishing the manuscript

Dear Editor and reviewers,

Thank you for your decision and recommendations to our MS. We appreciate the relevancy of sent comments and in the following text, we address each point and describe how and where your suggestions were incorporated to the revised MS. For your easy orientation, the original text from the decision letter is left in standard font, while our responses are introduced by “Reply” and indicated in *italics*.

We are grateful for your constrictive criticism and hope we addressed it well.

Sincerely, Karel Janko

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I appreciate the authors' thorough incorporation of previous suggestions. I find the changes made markedly improve the clarity of the manuscript and am satisfied with the revisions. I'm excited to see this work published!

One minor change to consider, if there is time, is to alter the color palette of the graphs to be colorblind friendly or at least avoid red/green combinations given the prevalence of red/green colorblindness. However, this is far from a reason to greatly delay publishing the manuscript

Reply:

- *We are grateful for your kind words and appreciation of our work.*
- *Thank you for noting the issue with color blindness, we were not aware of that. To address your comment, we have changed the green color by the black in relevant figures where red-green contrast was. As such, we believe that potential issues most frequent type of color blindness was addressed satisfactorily.*