

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

Received: 20 July 2023

Accepted: 12 September 2024

Published online: 08 October 2024

Karel Janko<sup>1</sup> , Jan Eisner<sup>2,3,6</sup> , Petr Cigler<sup>4,6</sup>  & Tomáš Tichopád<sup>3,5</sup> 

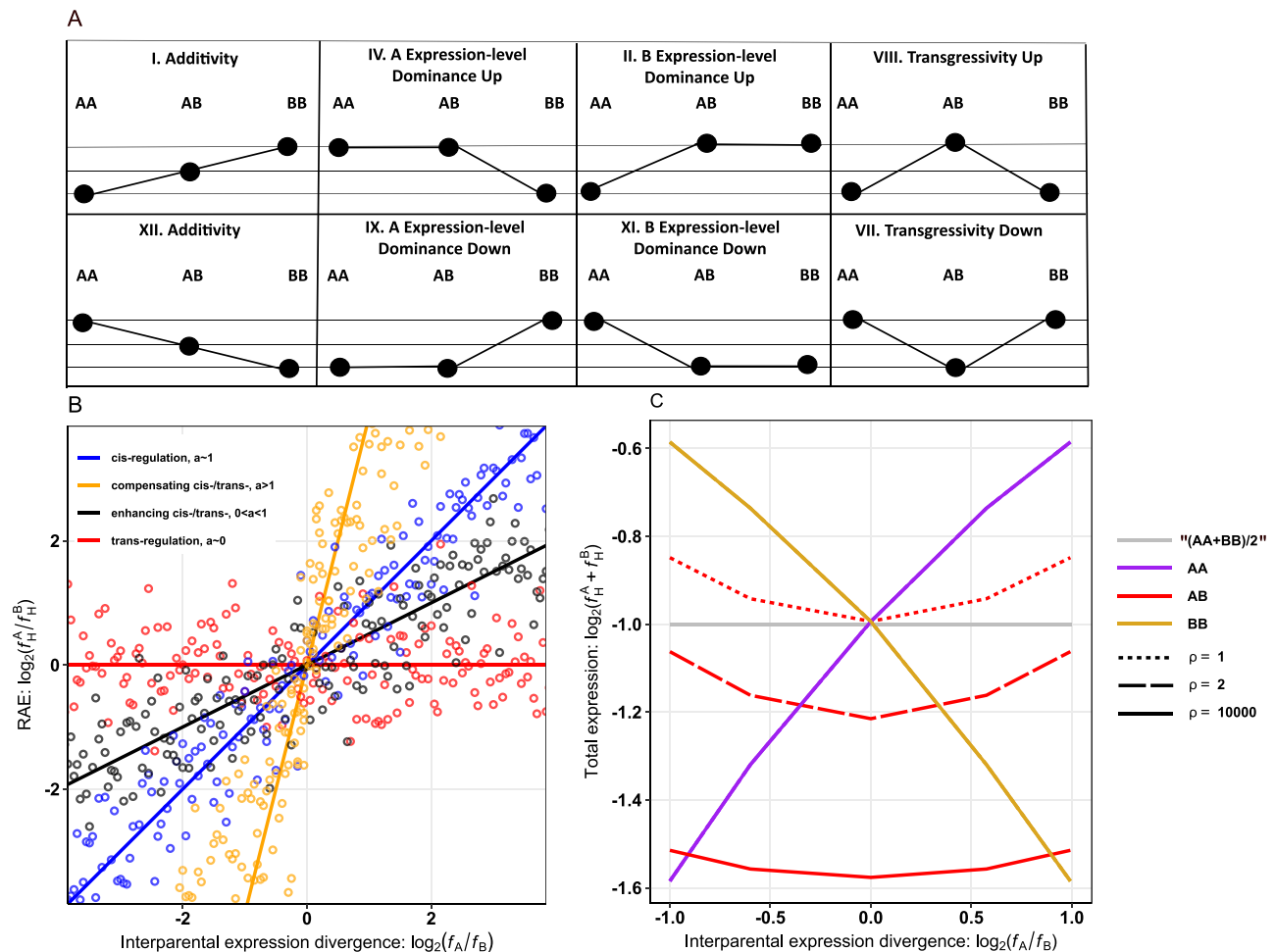
Hybridization and polyploidy are powerful evolutionary forces, inducing a range of phenotypic outcomes, including non-additive expression, sub-genome dominance, deviations in genomic dosage, and transcriptome downsizing. The reasons for these patterns and whether they are universal adaptive responses to genome merging and doubling remain debated. To address this, we develop a thermodynamic model of gene expression based on transcription factor (TF)-promoter binding. Applied to hybridization between species with divergent gene expression levels, cell volumes, or euchromatic ratios, this model distinguishes the effects of hybridization from those of polyploidy. Our results align with empirical observations, suggesting that gene regulation patterns in hybrids and polyploids often stem from the constrained interplay between inherited diverged regulatory networks rather than from subsequent adaptive evolution. In addition, occurrence of certain phenotypic traits depend on specific assumptions about promoter-TF coevolution and their distribution within the hybrid's nucleoplasm, offering new research avenues to understand the underlying mechanisms. In summary, our model explains how the legacy of divergent species directly influences the phenotypic traits of hybrids and allopolyploids.

Hybridization and polyploidization have long intrigued researchers as powerful drivers of evolutionary change, which may lead to a range of phenotypic effects, from beneficial adaptations to detrimental imbalances in allelic products, e.g. refs. 1–4. Empirical studies of various hybrid and polyploid organisms identified some notable patterns in how these organisms modify gene expression levels as well as the relative expression of orthologous alleles inherited from their parental species.

Specifically, cases of non-additive expression have frequently been documented, where expression levels of hybrid's genes resemble the expression levels of one of the parents, but do not match the

expected average expression between the parents (i.e. the expression in a hybrid deviates from the mid-parental values). This leads to the so-called expression-level dominance pattern<sup>1,5,6</sup>, Fig. 1A, Supplementary Note 1–Fig. S1. Several studies that investigated absolute transcriptome sizes also reported overall transcriptome downsizing when gene expression in hybrids and in polyploids was not necessarily scaled to the ratio of parental cell sizes, but may be systematically modified<sup>7–10</sup>. Studies comparing relative allelic expression (RAE), noted that divergence in cis/trans-regulatory elements between parental species may drive the allelic expression in hybrids and allopolyploids. Specifically, in hybrids between closely related (sub)species,

<sup>1</sup>Laboratory of Non-Mendelian Evolution, Institute of Animal Physiology and Genetics of the Czech Academy of Sciences, Rumburská 89, 277 21 Liběchov, Czech Republic. <sup>2</sup>Department of Mathematics, Faculty of Science, University of South Bohemia in České Budějovice, Branišovská 1760, 370 05 České Budějovice, Czech Republic. <sup>3</sup>Department of Biology and Ecology, Faculty of Natural Sciences, University of Ostrava, Chittussiho 10, Ostrava, Czech Republic. <sup>4</sup>Institute of Organic Chemistry and Biochemistry of the CAS, 166 10, Prague, Czech Republic. <sup>5</sup>University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Zátěží 728/II, 389 25, Vodňany, Czech Republic. <sup>6</sup>These authors contributed equally: Jan Eisner, Petr Cigler. ✉ e-mail: [k.janko@yahoo.com](mailto:k.janko@yahoo.com); [petr.cigler@uochb.cas.cz](mailto:petr.cigler@uochb.cas.cz)



**Fig. 1 | Gene expression inheritance categories, types of cis- and trans-regulation and their models through fractional occupancies of specific binding sites.** **A** Gene expression inheritance categories sensu Yoo et al.<sup>35</sup>; the pictograms indicate expression levels in parental species A and B whose diploid genomes are denoted as AA and BB, respectively, while hybrid is denoted as AB. Simplified figure contains 8 out of 12 total categories; details provided in Supplementary Note 1. **B** Types of cis- and trans-regulation of hybrids' allele expression. For each hypothetical gene (represented by a randomly chosen individual dot), the scatterplot demonstrates its expression divergence between parental species A and B on the x-axis (i.e. the  $\log_2$  fold change difference between the gene's expression in parents corresponding to  $\log_2(f_A/f_B)$  ratio in our model notation) and relative expression of hybrid's alleles  $A_{hyb}$  and  $B_{hyb}$  on the y-axis (i.e. RAE corresponding to  $\log_2(f_H^A/f_H^B)$  in our model notation, see Eq. 20 ad 21). Four types of cis-/trans-regulatory interactions are depicted, depending on the slope of correlation  $a$  between  $\log_2(f_H^A/f_H^B)$  and  $\log_2(f_A/f_B)$  ratios: (1) pure cis regulation ( $a \sim 1$ ; blue dots); (2) pure trans regulation ( $a \sim 0$ ; red dots) when the RAE is balanced

irrespective of parental divergence implying that both alleles are equally regulated by a common suite of TFs (i.e., when  $\log_2(f_H^A/f_H^B) = 0$ ); (3) compensating cis-/trans-interaction ( $a > 1$ ; yellow dots) and (4) enhancing cis-/trans-interaction ( $0 < a < 1$  black dots). Details are provided in Supplementary Note 1. **C** Fractional occupancies of specific binding sites (of A- and/or B-origin) under the reference scenario 1. Fractional occupancies are weighted by their genomic dosages; scenario 1 assumes single specific site per chromosome and negligible concentration of free TFs. X-axis shows the expression divergence between parental species in terms of fractional occupancies  $\log_2(f_A/f_B)$ , y-axis indicates the occupancies of all specific sites in an individual. Note: Types of individuals and strength of cross regulation are denoted by line colours, styles and symbols. Species A is indicated by cyan line, species B by yellow line, their average expression (i.e. [expression of AA+ expression of BB]/2) are denoted by grey line, diploid hybrid AB by red lines. Dotted lines show results for  $\rho = 1$  (full cross-regulation), slashed lines show for  $\rho = 2$  (limited cross-regulation) and full lines show  $\rho = 10000$  (no cross-regulation).

orthologous alleles  $A_{hyb}$  and  $B_{hyb}$  derived from parental species A and B tend to be subject to joint control by both sets of transcription factors (TFs). By contrast, in cases of distantly related parental species, cis-regulation prevails due to the divergence in promoters that respond less to orthologous TFs<sup>11–13</sup>. Transcriptional bias between subgenomes, known as the subgenome dominance, is also common with hybrids and allopolyploids being phenotypically more similar to one of their parental species, e.g. ref. 14. Finally, deviations from genomic dosage are also commonly observed when expressions of homoeologous alleles in allopolyploids do not match their ratio in genomic DNA (gDNA). For instance, allopolyploid hybrids with asymmetrical composition of parental genomes, like the AAB triploid hybrids combining two genomes of parental species A and single

genome of species B, may disproportionately underexpress the minority alleles ( $B_{hyb}$ ) derived from the parent B compared to the majority alleles ( $A_{hyb}$ ) derived from the parent A<sup>5</sup>.

It is a matter of hot debate in contemporary evolutionary biology as to which of these patterns are due to some case-specific mechanisms intrinsic to particular taxa, which stem from some common background, and what molecular mechanisms underlie the gene expression patterns in hybrids and polyploids. Three questions are particularly debated<sup>2</sup>. Namely, do parallel patterns arise “instantaneously” in unrelated organisms as a direct consequence of genome merging and multiplying (i.e. the “genomic shock”) or do they evolve subsequently in response to selective pressure acting on established hybrid and polyploid strains? Which effects may be assigned as a

consequence of genome mixing (hybridization) and which may be attributed to increased genome copy number (polyploidization)? Finally, to what extent do similar patterns among independently arisen lineages follow some general processes or remain driven by taxon-specific mechanisms?

Several general explanations have been proposed and for instance, the subgenome dominance may arise as a consequence of differential load and distribution of transposable elements (TEs) between the two parental species. The subgenome having its TEs generally closer to the genic regions is supposedly more prone to epigenetic silencing its expression attenuation, e.g. ref. 15. Adverse effects may also result from the mixing of separately evolving species-specific regulatory networks<sup>16,17</sup>. Unfortunately, finding generalizable conclusions is challenging as it is often difficult to discern whether observed patterns evolved as adaptive optimization of the ratio between cell size and gene production in hybrids and polyploids (e.g. ref. 2), or whether they have been induced by genome merging or multiplication<sup>18</sup>. Moreover, regulatory elements of most genes are either unknown or their detection is extremely challenging, labour and cost intensive<sup>19</sup>, making them difficult to directly investigate especially in non-model organisms.

Nonetheless, recent advances in nucleic acid sequencing offered unprecedented insight into the mechanisms of gene regulation. One of the most common approaches has been to perform allele-specific RNA sequencing of parents and their hybrid progeny. In such a way one can evaluate how much the expression divergence between parental species (i.e. fold change between expression of the gene in species A and B) is conserved in the relative allelic expression (RAE) of their hybrids (i.e. the cis-regulation inducing allele-specific effects) and how much the trans-acting regulatory mechanisms affect both of the hybrid's alleles  $A_{hyb}$  and  $B_{hyb}$  derived from species A and B, respectively<sup>20</sup>; see Fig. 1B and Supplementary Note 1. Such methods have been applied to investigate the variation in gene expression across intra- as well as inter-species levels, including polyploids (e.g. refs. 5,11,21–25). However, they offer only indirect insight into gene regulation and their application to empirical cases may be confounded by hidden effects imposed by differentiation between parental species<sup>12</sup>, ploidy level and genomic dosage<sup>5</sup>.

To understand how alleles and traits are expressed in organisms that have merged or multiplied genomes, scientists therefore need a robust theoretical framework that can make testable predictions. A promising approach to understanding the gene regulation under various circumstances may be to model the interactions between DNA and regulatory molecules that bind to it. This is usually modelled through thermodynamic and kinetic principles using statistical physics, Markov-chain models, or computational simulations<sup>26</sup>. In concordance with empirical data (e.g. reviewed in ref. 27), these models assume that TFs bind to both non-specific and specific sites on DNA, including gene promoters, and that gene expression into mRNA can occur when a TF molecule is physically bound to its promoter of a given gene. Some of these models thus approximate the transcriptional activity of a gene from the fractional occupancy of its promoter, which is mathematically tractable by applied equations<sup>26</sup>. Despite the utility of these models in systems biology to understand phenomena like allelic dominance<sup>28</sup> or gene regulatory networks<sup>29</sup>, there is still much to learn about how transcription patterns in organisms with merged or multiplied genomes relate to interactions between TFs and their binding sites.

Recently, Bottani et al.<sup>30</sup> proposed a thermodynamic model that assumes that promoters derived from different subgenomes are exposed to a common set of TFs within an allopolyploid organism (Fig. 2) and assumed that transcriptional activity of a promoter may be modelled by its fractional occupancy by TFs ( $f$ ). By modelling the affinity of TFs to binding sites through corresponding dissociation constants, they proposed how subgenome dominance may emerge as

a consequence of different expression levels of homoeologous alleles (genes with different ancestry but the same function) within an allopolyploid. Specifically, Bottani et al.<sup>30</sup> investigated several scenarios of divergence in trans-regulatory elements, cell volume and open chromatin conformation between parental species. They assumed that transcription factors need an open chromatin conformation to bind and this can happen also on non-specific binding sites in both parental genomes<sup>31</sup>. The authors concluded that parental genomes with a larger euchromatic ratio should have evolved TFs that bind more strongly, to make up for the larger number of potential binding sites that are accessible but not actually functional. Consequently, TFs from that parental genome with more euchromatin should bind with higher affinity to their homologous promoters. This in turn should bias their expression upwards, thus explaining the pervasive subgenome dominance in many allopolyploids.

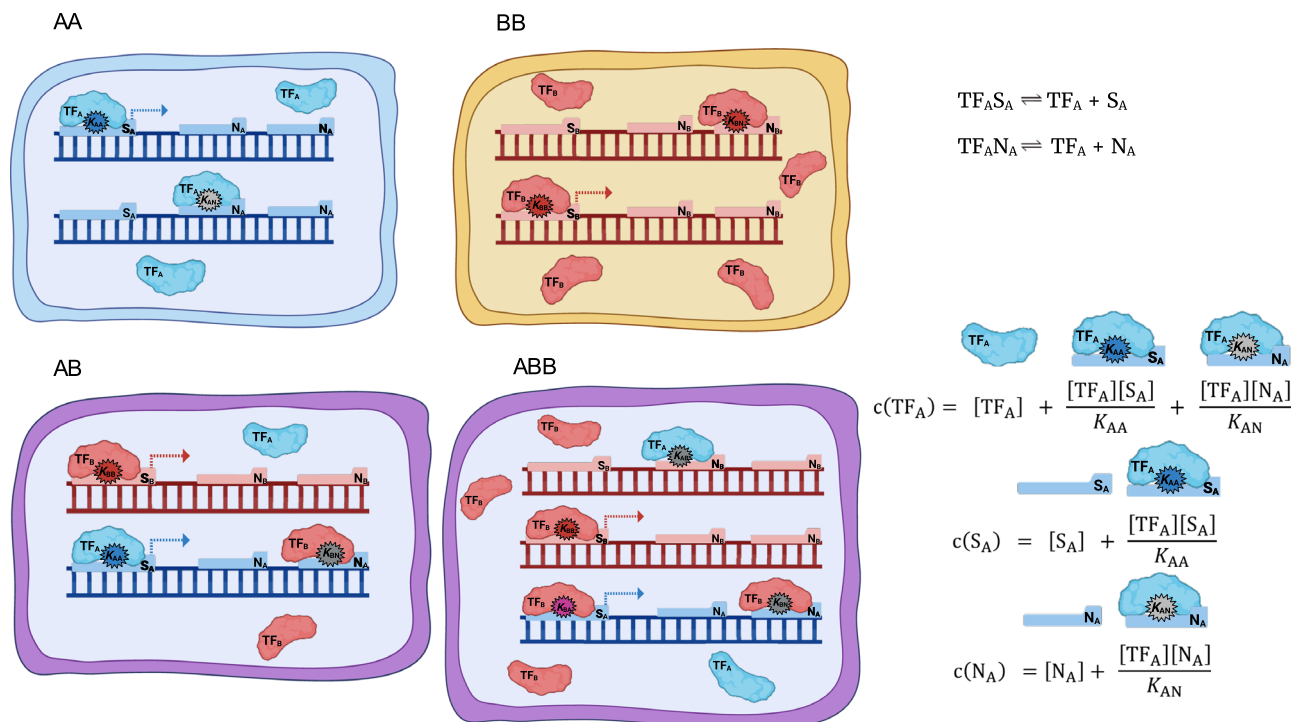
Bottani et al.'s study<sup>30</sup> represents a crucial conceptual shift, illustrating how differences between parental genomes, after merging and doubling into allopolyploids, may directly induce phenomena like dosage compensation and subgenome dominance, a trend also recently explored by An et al.<sup>32</sup>. However, to achieve mathematical tractability of complex systems combining two diverged regulatory networks in a single organism, Bottani et al.<sup>30</sup> used certain mathematical simplifications, which can make some conclusions of the study more context-specific than initially apparent. For instance, the model was applied to allotetraploids with a symmetric subgenome composition, such as AABB, which exhibit similar properties to AB diploid hybrids in terms of the equations used. This similarity raises questions about whether the observed mechanisms are solely attributable to genome merging or also to polyploidization. Furthermore, the model predicts that subgenome dominance would occur even in scenarios of full cross-regulation, a prediction that may not completely align with empirical data. Full cross-regulation theoretically should ensure equal occupancy of both sets of promoters, resulting in balanced gene expression from both subgenomes<sup>24</sup>, even if one set of TFs binds more strongly to specific sites. As we elaborate below and in Supplementary Note 2, our analysis suggests that such a scenario would only be feasible if the species with a higher euchromatic ratio modified its binding sites, rather than the TF molecules themselves.

To address these emerging questions, the present study introduces an enhanced thermodynamic model to explicate TF-promoter interactions, which resolves some of the constraints present in previous approaches. This flexible model enables us to separately identify the impacts resulting from genome merging (hybridization) and genome multiplication (polyploidy). We apply our model to a range of biologically feasible hybridization and polyploidization scenarios, thereby illustrating its capacity to accurately predict numerous observed patterns.

Our work posits that the broad array of gene and allele expression patterns witnessed in both experimental and natural diploid and polyploid hybrids can be primarily attributed to the restricted inter-communication among the diverged regulatory networks of their parent species.

## Results and discussion

In this study, we advanced our understanding of gene regulation in hybrid and polyploid organisms by developing a novel thermodynamic model that encompasses five mass-balance equations. These equations delineate the equilibrium concentrations of transcription factors (TFs) bound to specific and nonspecific sites in relation to the total concentrations of all regulatory components inherited from parental species. We assume that the binding sites on DNA bind the TFs independently, so formation of each complex of TF with its binding site proceeds as an individual molecular event. Example equations are visually described in Fig. 2 with their link to biological scenarios and



**Fig. 2 | Scheme of gene regulation via binding of transcription factors (TF) to specific sites (gene promoters).** Four types of nuclei are drawn, representing the two diploid species A (blue) and B (beige) with genomic constitutions AA and BB, respectively, and their diploid (AB) and triploid (e.g., ABB) hybrids visualized as violet cells (symmetric tetraploids of AABB states are not shown). Nuclei may contain variable volumes and two or more chromosomes, depending on their ploidy. Each chromosome contains variable number of non-specific sites (N) and one or more specific binding sites (S<sub>A</sub> or S<sub>B</sub>), whose ancestry is visualized in colours reflecting the original parental species. TFs of A- or B-origin occur in characteristic concentrations, depending on parental gene expression levels, and may exist in four states, i.e. free, bound to a non-specific site or bound to either gene promoter. The last type of binding may be of “conspecific” form, when TF binds to the promoter from homologous subgenome (e.g., A-type TF binds to A-type promoter) or of “heterospecific” form when A-type TF binds to B-type promoter. The strength of

binding is denoted by the dissociation constant  $K_{ij}$  where indices  $i$  and  $j$  indicate the A, B or N and corresponds to conspecific, heterospecific or nonspecific bindings, respectively (see Eqs. 12–17). The gene is assumed to be expressed (arrow) when its promoter site (S<sub>A</sub> or S<sub>B</sub>) is bound by any TF molecule. Example chemical equilibrium equations (top-right) are presented for a single species A (for simplicity). In this situation, TF binds in two manners only: strongly to the specific regulatory sites S<sub>A</sub> (the corresponding dissociation constant  $K_{AA}$  has a low value) and weakly to many non-regulatory, non-specific sites N<sub>A</sub> (the corresponding dissociation constant  $K_{AN}$  has a high value). The insert in bottom-right visualizes the link between the biological model and mass-balance equations Eqs. 4–6 for TF<sub>A</sub>, S<sub>A</sub> and N<sub>A</sub> used to calculate their equilibrium concentrations in the species A (for a full set of equations see the model description in Methods section, Eqs. 4–17). Created in BioRender. Tichopád, T. (2023) BioRender.com/z05d991.

they are described in detail in the Methods section and in Supplementary Note 2.

We have ensured the adaptability of our model to a variety of assumptions about the types of hybridizing species and the properties of their regulatory networks. By tweaking parameters within the model, we have simulated six different scenarios aimed at investigating the impact of factors such as the number of specific binding sites, variations in cell volume between parental species, limitations in cross-regulation between subgenomes, and alternate assumptions of TF movement between binding sites on gene expression in hybrids (as detailed in the Methods section). Furthermore, we have scrutinized the influence of asymmetric euchromatic ratios on the outcome, specifically assessing the scenarios where the parental species with an increased number of non-specific sites has adaptively adjusted the binding attributes of its TFs molecules or its gene promoters.

In this section, we will discuss the results of our model and how it predicts trends in hybrid and allopolyploid organisms based on empirical studies. We will first explore overall gene expression patterns, regardless of the allelic origin of a given transcript. Then, we will look at RAEs and how they relate to the parental species.

### Overall gene expression patterns

At the level of overall gene expression patterns, we observed two prominent patterns in our simulations, namely the general downsizing

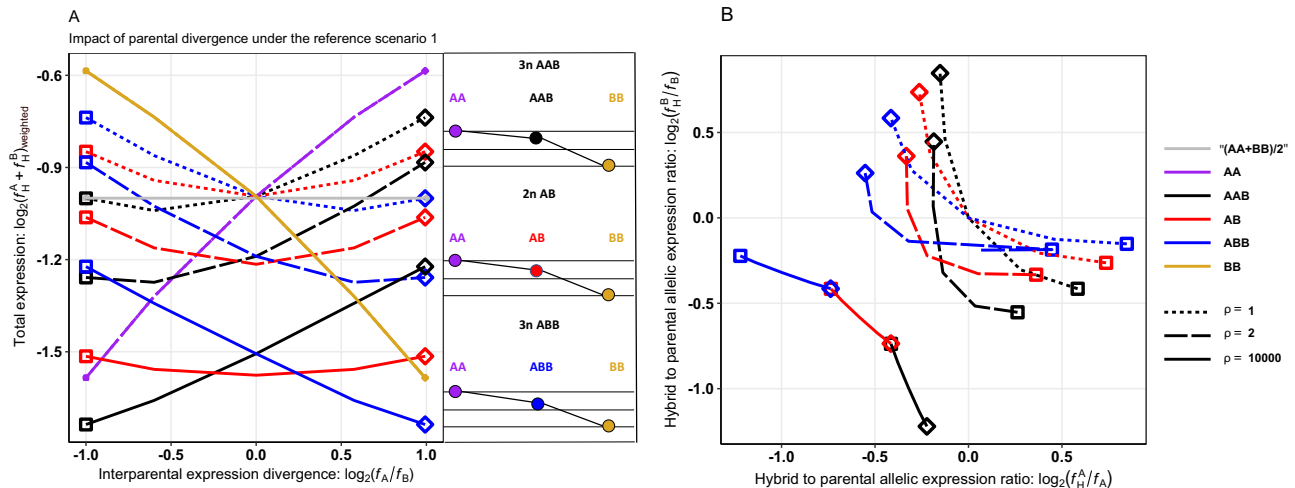
of hybrids’ transcriptomes and deviation from average “mid-parent” values in hybrids’ expression with a trend towards the expression-level dominance pattern:

### Transcriptome downsizing

The applied thermodynamic model of TF binding onto promoter sites implicitly predicts that as soon as the cross regulation between hybrids’ subgenomes becomes limited due to regulatory divergence between parental species, the downsizing of total transcriptome in a hybrid is observed (see Fig. 1C for details on diploid hybrids and Fig. 3A for all biotypes). This effect of *limited cross regulation* is caused by the fact that even if the hybrid preserves the same number of TF molecules per cell, the concentration of ‘conspecific’ TFs with respect to their homologous promoter sites is halved. Consequently, the fractional occupancy of promoters on each subgenome is lower than original parental values (Fig. 3B). We observed such a downsizing under all simulated scenarios and conditions, irrespective of number of specific binding sites, proportion of bound vs free TFs, ploidy level or asymmetries in cell volumes (Figs. 3–5).

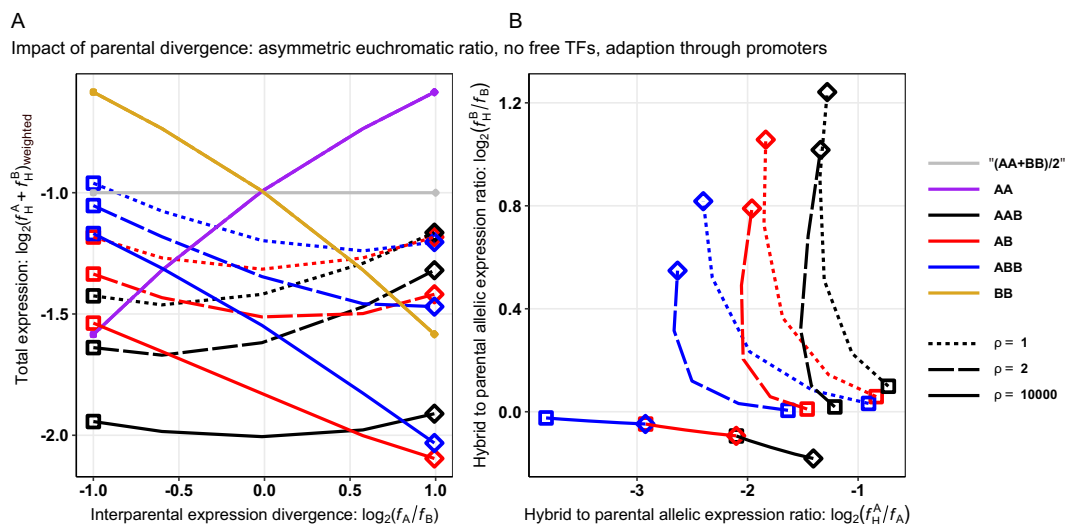
We also found that transcriptome downsizing may occur even under *full cross regulation* between subgenomes in a hybrid, but only under certain conditions. Namely, it occurred only when both parental species differ in nonspecific DNA content and the adaptation to a higher euchromatic ratio was achieved via evolution of promoter sites,





**Fig. 3 | Fractional occupancies of genes and alleles under the reference scenario 1.** **A** Fractional occupancies of specific binding sites (of A and/or B-origin) weighted by their genomic dosages (note that A-derived binding sites have twice higher absolute numbers in species A with genotype AA than in AB hybrid and identical to AAB triploid). X- and y-axes are analogous to Fig. 1C. Types of individuals and strength of cross regulation are denoted by line colours and styles, respectively. Parental species with genotype AA is indicated by cyan, species with BB by yellow, their average expression (i.e. [expression of AA + expression of BB]/2) are denoted by grey line, diploid hybrid AB by red, triploid hybrid AAB by black and triploid hybrid ABB by blue. Dotted lines show results for  $\rho = 1$  (full cross-regulation), slashed lines show for  $\rho = 2$  (limited cross-regulation) and full lines show  $\rho = 10000$  (no cross-regulation). The insert with coloured schemes demonstrates how the simulation results correspond to expression level categories similarly as in

Fig. 1A. Specifically, when fractional occupancy of parent A is higher than of parent B (right side of x-axis;  $f_A > f_B$ ), the levels of fractional occupancies in all hybrid forms (AB, AAB and ABB) are higher than average of their parents. **B** Comparison of fractional occupancies of alleles in hybrids ( $f_H^A$  and  $f_H^B$ , respectively) relative to occupancies of their homologous alleles in parents ( $f_A$  and  $f_B$ , respectively). X-axis shows the  $\log_2(f_H^A/f_A)$  ratio of  $A_{\text{hyb}}$  allele's occupancy in hybrid relative to that in parental species A, y-axis shows the  $\log_2(f_H^B/f_B)$  ratio of  $B_{\text{hyb}}$  allele's occupancy in hybrid relative to that in parental species B. Line colours and styles are analogous to panel A. Note: Reference scenario 1 assumes single specific site per chromosome and negligible concentration of free TFs. Square and diamond symbols along the lines provide the link between panels A and B by showing two reference values of  $\log_2(f_A/f_B)$  expression divergence between parental species; square:  $\log_2(f_A/f_B) = -1$  and diamond:  $\log_2(f_A/f_B) = 1$ .



**Fig. 4 | Fractional occupancies of genes and alleles under the scenario 6.a.** **A** Fractional occupancies of specific binding sites (of A- and/or B-origin) weighted by their genomic dosages (note that A-derived binding sites have twice higher absolute numbers in AA species than in AB hybrid and identical to AAB triploid). X- and y-axes are analogous to Fig. 1C. Line colours and styles correspond to the types of individuals and strength of cross regulation analogously to Fig. 3. **B** Comparison of fractional occupancies of alleles in hybrids ( $f_H^A$  and  $f_H^B$ , respectively) relative to occupancies of their homologous alleles in parents ( $f_A$  and  $f_B$ , respectively). X-axis shows the  $\log_2(f_H^A/f_A)$  ratio of  $A_{\text{hyb}}$  allele's occupancy in

hybrid relative to that in parental species A, y-axis shows the  $\log_2(f_H^B/f_B)$  ratio of  $B_{\text{hyb}}$  allele's occupancy in hybrid relative to that in parental species B. Line colours and styles are analogous to panel A. Note: Reference scenario 6.a assumes single specific site per chromosome, negligible concentration of free TFs and 10 times higher euchromatic ratio in species B with adaptation through modification of binding sites. Square and diamond symbols along the lines provide the link between panels A and B by showing two reference values of  $\log_2(f_A/f_B)$  expression divergence between parental species; square:  $\log_2(f_A/f_B) = -1$  and diamond:  $\log_2(f_A/f_B) = 1$ .

as in Bottani et al.'s<sup>30</sup> model (Fig. 4A). In such a situation, the hybrid's promotor sites derived from the species with larger euchromatic ratio attract the TFs with higher affinities, leading to their higher fractional occupancies. However, the decrease in occupancy of the other

parental's promotor is more substantial, causing the overall down-sizing of the hybrid's transcriptome (Fig. 4B).

These findings have interesting implications for existing literature, which predicted that hybrids and especially polyploids would

have their transcription level modified beyond the values expected by simple summing or averaging of their parents<sup>7</sup>. However, empirical tests of such predictions are scarce since typical RNAseq experiments measure only relative, rather than absolute levels of expression<sup>33,34</sup>. Our model predicts that as soon as the regulatory networks of parental species diverged to a level where full cross regulation is not possible, the hybrids should possess generally lower expression levels per cell.

### Non-additive gene expression

The utilized model also predicts that overall gene expression in hybrids and allopolyploids deviates from additivity, i.e. hybrid's expression levels do not match the “mid-parent” values given by averaging the expression levels in their parental species (see Fig. 1). The occurrence of this pattern is caused by several mechanisms, which are discussed below.

In general, we found that when the parental species have different expression levels (i.e. when fractional occupancies differ between parental species so that  $f_A \neq f_B$ ), the hybrids approach the *expression-level dominance-UP* pattern and tend to have above-average occupancy of promoters, closer to that of the parents with higher expression levels (see Fig. 1C for details on diploid hybrids and Fig. 3A, B for all biotypes). To understand this effect, let us point at Fig. 3B, which compares the occupancies of alleles in hybrids and in parental species. It appears that when two species differ in expression of a particular gene, their hybrids tend to upregulate the promoters of the less expressing species and downregulate those of the more expressing species. However, due to different concentration of both types of TFs, such effects are asymmetric and the allelic upregulation is more prominent than allelic downregulation (Fig. 3B). This is consistent with empirical observations (e.g. refs. 5,35,36) and such an asymmetry ultimately causes higher-than-average gene expression in hybrids, trending towards the *expression-level dominance-UP* pattern (Fig. 3A).

We observed such non-additive expression patterns in all simulated model designs, but the particular type of non-additivity (see Fig. 1A and Supplementary note 1 for explanation of the term) depended on the concentration of free TFs, the genome dosage within hybrids, on asymmetries of cell volumes and on euchromatic ratios between parents. Three notable patterns emerged from our simulations.

*Genome dosage* has a considerable effect on non-additive expression, depending on the asymmetry of a subgenome's composition in hybrids. Namely, when the expression levels in species A are higher than in B (i.e. when  $f_A > f_B$ ), the expression levels in triploids with an AAB subgenome composition match species A's levels more closely than do symmetric hybrids (AB or AABB). This indicates the *expression-level dominance UP* pattern (Fig. 3A; right side of  $x$ -axis). In contrast, the ABB triploid's expression level approaches mid-parental values or even slightly drops below them (Fig. 3A, right side of  $x$ -axis). This is due to the fact that in a common nuclear environment, the changes in concentration of more abundant A-derived TFs had larger impact on B-derived promoters, than vice versa (Fig. 3B). Consequently, when expression of species A is greater than that of species B ( $f_A > f_B$ ), the downregulation of A-derived promoters is much weaker than upregulation of B-derived promoters. This leads to overall higher expression levels of AAB hybrids than of AB or AABB hybrids.

*Asymmetries in cell volume* in combination with the variation in proportion of free vs. bound TFs also had a considerable effect on non-additive expression patterns in hybrids. Specifically, when most TF molecules are bound and the concentration of free TFs is minimal, the asymmetries in cell volumes have negligible effect and hybrids possess above-average expression levels (Fig. 5A). Expression patterns thus do not differ from the case of symmetric cell volumes (compare with Fig. 3A). However, when cells contain a large fraction of free TFs, asymmetries in cell volume start to matter. In such cases, above-average expression levels may only be observed when the species with

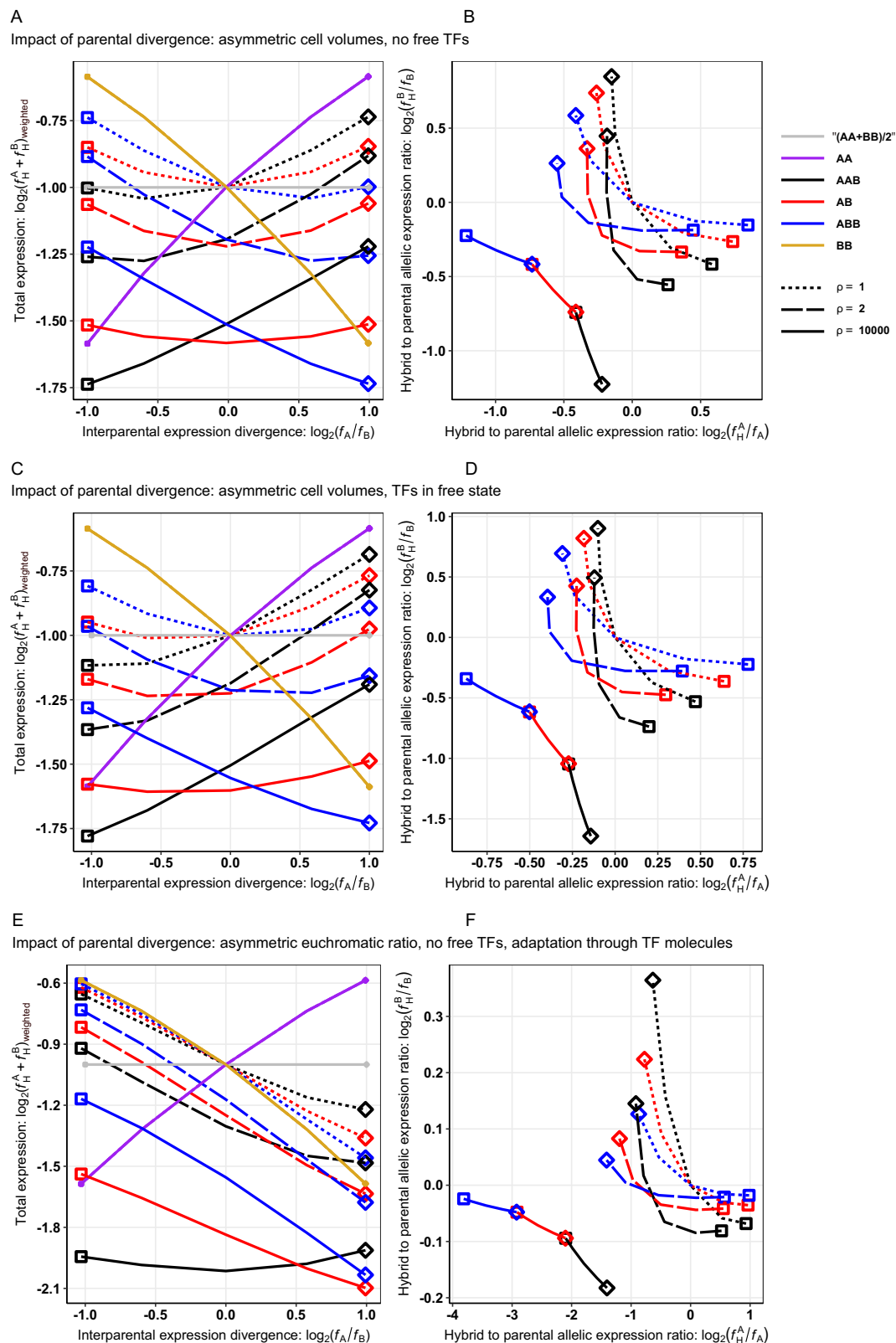
larger cell volume also has stronger expression (i.e. when cell volume (AA) > cell volume (BB) and  $f_A > f_B$ ; Fig. 5C right side of  $x$ -axis). In the opposite cases when cell volume (AA) > cell volume (BB) but  $f_A < f_B$ , hybrids tend to mid-parental values or even lower occupancies; Fig. 5C left side of  $x$ -axis.

The combined effect of cell size asymmetry and proportion of free TFs may be understood from comparing the allelic expression in hybrids and parental species (Fig. 5B and D). Namely, as in Bottani et al.<sup>30</sup>, our model assumed that the change in cell volume in parental species is accompanied by an adequate change in the absolute number of TFs in its cells in order to ensure the same fractional occupancy. Consequently, the parental species with larger cells (AA) generally contributes more TF molecules to a hybrid than the other species (BB). Hence, when the species with larger cells has higher gene expression, the upregulation of B-derived promoters is more significant than downregulation of A-derived promoters (Fig. 5D upper left quadrant). In the opposite case when the species with the larger cell has lower expression (Fig. 5D lower right quadrant), the upregulation of promoters is less prominent than in the previous case while the downregulation is stronger. This causes the tendency of hybrids towards expression level dominance-UP patterns in genes where the parental species with a larger cell volume has higher expression, and towards mid-parental values in genes where it has lower expression. Interestingly, this mechanism is almost undetectable when the concentration of free TFs is minimal, so that hybrids generally tend to expression level dominance-UP (compare Fig. 5B and 5D).

Finally, we noted the effect of *asymmetric contents of non-specific binding sites*. This mechanism is described in greater detail in the next section about subgenome dominance. Here we just note that the impact of interparental difference in euchromatic ratio depends on whether the adaptation to increase the amount of non-specific sites occurs via modification of promoter sequences or via modification of TF molecules. When promoters are modified, the subgenome from the species with higher euchromatic content (species B in our simulation) always has higher fractional occupancy because it has higher affinity for all TFs (Fig. 4A, B). The other subgenome is downregulated in such a case, conforming to the subgenome dominance pattern (Fig. 4B). As a side effect of such asymmetry, hybrids' genes are generally downregulated (Fig. 4A).

On the other hand, when TF molecules are modified, the hybrid's fractional occupancies match the species with a higher euchromatic ratio and conform to the expression-level dominance pattern. Specifically, they conform to the expression-level dominance-UP in genes where the species with a higher euchromatic ratio has higher expression (Fig. 5E, left side of  $x$ -axis), and to the expression-level dominance-DOWN in genes where the species has lower expression (Fig. 5E, right side of  $x$ -axis). This is because under full cross-regulation, TFs from the species with the higher euchromatic ratio have a higher affinity towards any specific sites and consequently have a higher impact on promoter occupancies than the other parental's TFs. Consequently, the fractional occupancies of promoters from the species with the higher euchromatic ratio are modulated to a lesser extent than the other parental's promoters (Fig. 5F).

Overall, our simulated results align well with numerous empirical patterns. Like in our model, hybrid and allopolyploid organisms often exhibit deviations from “mid-parent” values, leading to non-additive gene expression. This frequently conforms to the expression-level dominance pattern when expressing their genes at levels that resemble one of the parental species<sup>2,5,6</sup>. Consistent with empirical data, our model indicates that hybrid expression levels predominantly match the parent with higher expression, reflecting an expression-level dominance UP pattern e.g. refs. 5,6,35,36, but see refs. 37,38. Interestingly, our model's predictions regarding the effect of gene dosage



finds empirical support in a recent study by Bartoš et al.<sup>5</sup>, which examined gene expression in diploid and triploid hybrids of spined loach species, *Cobitis elongatoides* and *C. taenia*. The authors reported that triploids with double the genomic dosage of *C. elongatoides* had a higher proportion of genes exhibiting *elongatoides*-expression-level dominance patterns compared to diploid hybrids and triploids with

double the genomic dosage of *C. taenia*, where genes with *taenia*-expression-level dominance were more prevalent.

### Allele specific expression

At the level of allelic-specific expression patterns, three major patterns were revealed:

**Fig. 5 | Fractional occupancies of genes and alleles under combination of reference scenarios 3 and 5 and 6.b.** **A, B** Combination of scenarios 3 and 5, i.e. single specific site per chromosome with negligible concentration of free TFs and cell volumes of parental A twice as large as the parental B. **C, D** Different combination of scenarios 3 and 5, i.e. single specific site per chromosome with significant concentration of TFs in a free state and cell volumes of parental A twice as large as the parental B. **E, F** Scenario 6.b, i.e. single specific site per chromosome, negligible concentration of free TFs and 10 times higher euchromatic ratio in species B with adaptation through modification of TF molecules. **A, C, and E** Fractional occupancies of specific binding sites (of A- and/or B-origin) weighted by their genomic dosages (note that A-derived binding sites have twice higher absolute numbers in AA species than in AB hybrid and identical to AAB triploid). X- and y-axes are

analogous to Fig. 1c. Line colours, styles and symbols correspond to the types of individuals and strength of cross regulation analogously to Fig. 3.

**B, D, F** Comparison of fractional occupancies of alleles in hybrids ( $f_H^A$  and  $f_H^B$ , respectively) relative to occupancies of their homologous alleles in parents ( $f_A$  and  $f_B$ , respectively). X-axis shows the  $\log_2(f_H^A/f_A)$  ratio of  $A_{hyb}$  allele's occupancy in hybrid relative to that in parental species A, y-axis shows the  $\log_2(f_H^B/f_B)$  ratio of  $B_{hyb}$  allele's occupancy in hybrid relative to that in parental species B. Line colours and styles are identical to **A, C, and E**. Note: Square and diamond symbols along the lines provide the link between **A, C, E** and **B, D, F** by showing two reference values of  $\log_2(f_A/f_B)$  expression divergence between parental species; square:  $\log_2(f_A/f_B) = -1$  and diamond:  $\log_2(f_A/f_B) = 1$ .

### Subgenome dominance

A prominent feature of many hybrids and allopolyploids is that one ancestor's subgenome dominates the transcriptomic and phenotypic patterns of the other. According to performed simulations, this subgenome dominance arises as a direct consequence of genome merging when cross-regulation between subgenomes is limited. Three mechanisms triggering this pattern are noteworthy:

First, the *genome dosage* significantly impacts the subgenome dominance when cross-regulation is limited. In unbalanced polyploids, like e.g. AAB triploids, the promoter sites of the species contributing higher genomic dose (A-derived promoters) tend to have systematically higher fractional occupancies than promoters of the other species (Fig. 6A). This is because in a common cellular environment of an AAB hybrid, there is generally a higher concentration of A-derived TF molecules than of B-derived ones. Consequently, the A-derived promoters are downregulated to a lesser extent than the B-derived ones (Fig. 3B). In hybrids with balanced genome dosages, like AB diploids or AABB tetraploids, the subgenome dominance does not occur due to this mechanism.

Second, *asymmetries in cell volumes* between parental species and the type of movement of TF molecules (i.e. the variation in proportion of free vs. bound TFs) also affects the allelic ratios. Specifically, the genome of the parental species with larger cells provides a higher amount of TF molecules and consequently, its promoters tend to have higher fractional occupancies when cross-regulation is limited (Fig. 6B). This pattern occurs in all ploidy levels and genome dosages, however, as mentioned in the section about non-additive gene expression, it was observed only when a significant proportion of TFs existed in a free state. In the case when TFs are (almost) fully bound to binding sites, this type of subgenome dominance vanished (Fig. 6C).

Third, *asymmetries in the euchromatic ratio* generally caused higher fractional occupancy of promoters derived from the species with higher concentrations of non-specific binding sites. This phenomenon typically occurred under limited cross-regulation between parental subgenomes, but in some situations it was also possible under full cross-regulation (Fig. 6D). However, the subgenome dominance would not occur under full cross-regulation when the parental species adapted to a higher euchromatic ratio solely by increasing the specificity of its TFs. This is because all promoter sites would be equally occupied by both types of TFs even if one has generally higher specific binding energy (Fig. 6E). The only one situation when full cross-regulation allowed subgenome dominance to occur was when adaptation to increased euchromatic ratio was realized via evolution of promotor binding sites (compare Fig. 6D, E). In such a case, the higher affinity of its promotor sites to any TFs attracted both conspecific and heterospecific TFs with higher likelihood.

### The impact of regulatory divergence on trans-regulation and on the classification of genes in a hybrid

The applied thermodynamic model also explains the deviations of RAE in hybrids and polyploids from values predicted by the expression

divergence between parental species. To understand this pattern, let us first consider the behaviour of RAE under two extreme conditions; the full cross-regulation when orthologous TFs do not differentiate between homologous and orthologous promoters, and no cross-regulation, when TFs do not differentiate between allospecific (orthologous) promoters and other non-specific DNA motifs.

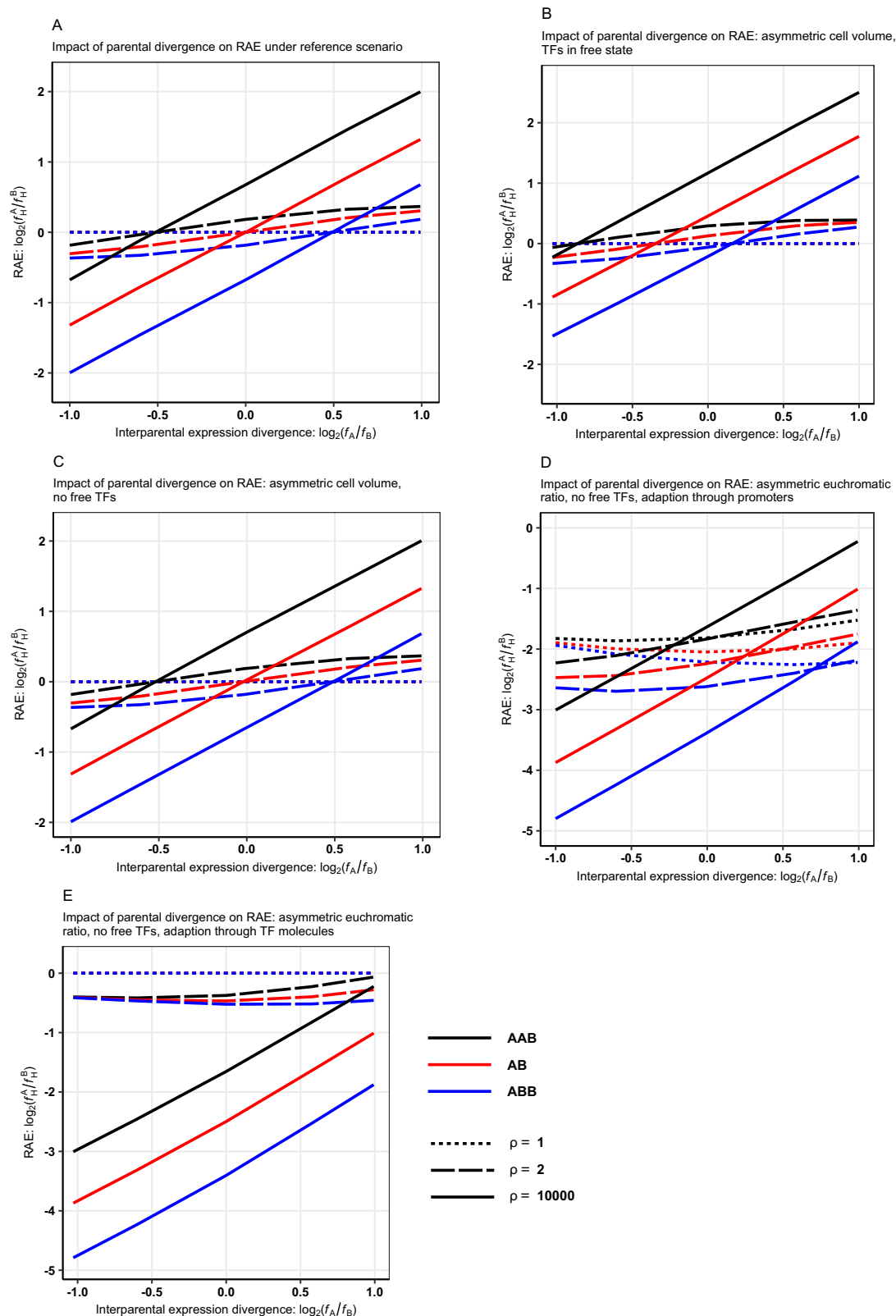
In the case of *full cross-regulation*, the values of RAE, normalized per chromosome in order to compare diploids and polyploids, always equals 1. This is because TFs from both subgenomes bind with equal affinity to homologous and orthologous copies of promoters, thereby enforcing their equal expression (Fig. 6A; dotted line). Genes in such hybrids would thus be categorized as “only trans” according to the nomenclature used in refs. 20,22,24, (see Fig. 1B and Supplementary Note 1 for explanation of categories), because both parental alleles in a hybrid ( $A_{hyb}$  and  $B_{hyb}$ ) have equal expression (i.e. the fractional occupancies of conspecific binding sites on these alleles are the same,  $f_H^A = f_H^B$ , where the lower H index stays for AB, BA, AAB or ABB hybrid type, see model description in the Methods section for details) irrespective of the expression divergence between parental species A and B (i.e.  $\log_2(f_A/f_B)$ ). In other words, the slope of correlation between RAE in hybrids and gene expression divergence between parental species equals zero (i.e.  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B) = 0$ ).

By contrast, with *no cross-regulation*, the RAE tends to exceed the expression divergence of parents and the slope of correlation is steeper than 1, i.e.  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B) > 1$  (Fig. 6A; solid lines). A hybrid's genes would thus be categorized as under “compensating cis and trans effects” according to nomenclature used in refs. 20,22,24, see Fig. 1B. A combination of two processes is responsible for this effect. First, the limited cross-regulation generally causes lower fractional occupancies of orthologous promoters, leading overall to a downregulation of the hybrid's transcriptome (see the section about transcriptome downsizing above). Second, and simultaneously, the promoters from the parental species with lower gene expression are downregulated to a greater extent than the promoters of the parent with higher expression (Fig. 3B). This is because in a common cellular environment, the TFs from the parent with lower expression are “diluted” to a greater extent than TFs from the species with stronger expression. A combination of these two effects leads to the slope of correlation  $> 1$ , i.e.  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B) > 1$  (Fig. 6A; solid lines).

At intermediate situations with existing *but limited cross-regulation*, the RAE deviates from 1 and the slope of the correlation takes a value greater than zero, i.e.  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B) > 0$ . According to nomenclature used in refs. 20,22,24 (Fig. 1B) a hybrid's genes may thus conform to “enhancing cis and trans effects” when the slope is between 0 and 1, “only cis regulation” when the slope equals 1, or even “compensating cis and trans effects” when the slope is  $> 1$ .

We noted that strength of cross-regulation affected the correlation between RAE and parental expression divergence ( $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B)$ ) under all simulated scenarios and assumptions (Fig. 6A-E). In our opinion, it has considerable implications for studies about gene





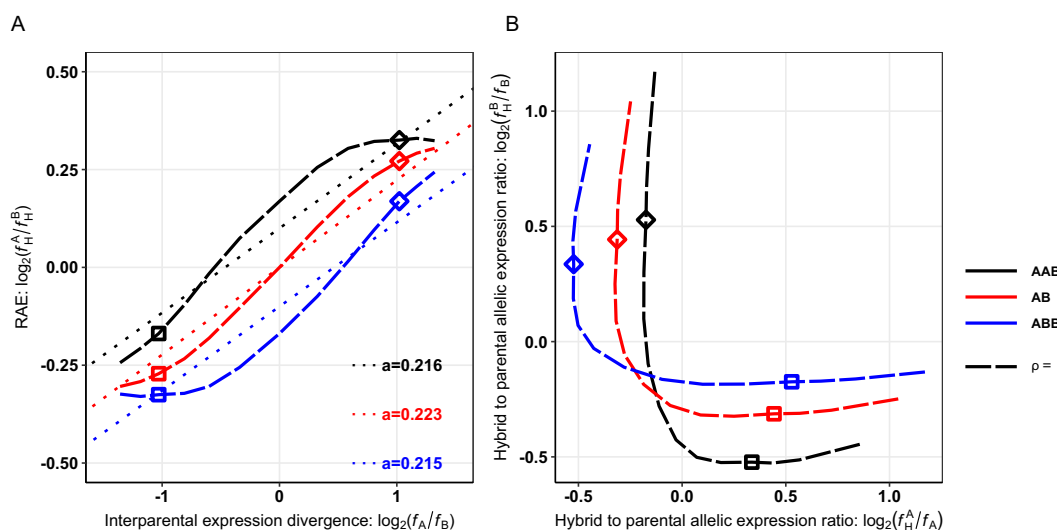
regulation. Effectively, it suggests that even if the expression of a given gene is fully under trans- regulation, as assumed by our model, its RAE in a hybrid may apparently conform to pure- cis-, pure-trans or combined cis-/trans- regulation, depending on the cross-talk between the two merged regulatory networks. This is in accordance to previous empirical data<sup>11–13</sup>.

### Non-linear distribution of RAE and the impact of ploidy and genomic dosage on efficiency of trans-regulation

Finally, our model has revealed another intriguing finding that challenges conventional analyses of cis-/trans- regulation of genes which are based on relative allelic expression (RAE) (e.g. refs. 20,22,24, see Fig. 1B). Specifically, we have observed that the relationship between

**Fig. 6 | Relative allelic expression (RAE) in hybrids and polyploids with respect to parental expression divergence.** **A** Simulation results under reference scenario 1, i.e., assuming single specific site per chromosome and negligible concentration of free TFs. **B** Simulation results under the combination of scenarios 3 & 5, i.e., assuming single specific site per chromosome, significant concentration of TFs in a free state and asymmetric cell volumes with the parental A twice as large as the parental B. **C** Simulation results under the combination of scenarios 3 & 5 assuming single specific site per chromosome, negligible concentration of free TFs and asymmetric cell volumes with the parental A twice as large as the parental B. **D** Simulation results under the combination of scenarios 3 & 6.a, i.e. assuming single specific site per chromosome, negligible concentration of free TFs and

10 times higher euchromatic ratio in species B with adaptation through modification of binding sites. **E** Simulation results under the combination of scenarios 3 & 6.b, i.e. assuming single specific site per chromosome, negligible concentration of free TFs and 10 times higher euchromatic ratio in species B with adaptation through modification of TF molecules. *X*-axis demonstrates the expression divergence between parental species in terms of their fractional occupancies  $\log_2(f_A/f_B)$ , *y*-axis indicates the relative allelic ratios  $\log_2(f_H^A/f_H^B)$  of orthologous (parental) alleles  $A_{hyb}$  and  $B_{hyb}$  in a hybrid, normalized by their genomic dosage (note that A-derived binding sites have twice higher absolute numbers in AA species than in AB hybrid and identical to AAB triploid). Line colours and styles correspond to the types of individuals and strength of cross regulation analogously to Fig. 3.



**Fig. 7 | Relative allelic expression (RAE) and fractional occupancies of alleles in hybrids and polyploids with respect to parental expression divergence.** **A** RAE in hybrids and polyploids under the limited cross-regulation ( $\rho = 2$ ). *X*-axis demonstrates the expression differentiation between parental species in terms of their fractional occupancies  $\log_2(f_A/f_B)$ , *y*-axis indicates the relative allelic ratios of orthologous (parental) alleles  $A_{hyb}$  and  $B_{hyb}$  in a hybrid, normalized by their genomic dosage  $\log_2(f_H^A/f_H^B)$  (note that A-derived binding sites have twice higher absolute numbers in AA species than in AB hybrid and identical to AAB triploid). Coefficients *a* of regression slopes are estimated upon the assumption of a linear fit of each curve for diploid AB (red), triploid AAB (black) and triploid ABB (blue).

**B** Comparison of fractional occupancies of alleles in hybrids ( $f_H^A$  and  $f_H^B$ , respectively) relative to occupancies of their homologous alleles in parents ( $f_A$  and  $f_B$ , respectively). *X*-axis shows the  $\log_2(f_H^A/f_A)$  ratio of  $A_{hyb}$  allele's occupancy in hybrid relative to that in parental species A, *y*-axis shows the  $\log_2(f_H^B/f_B)$  ratio of  $B_{hyb}$  allele's occupancy in hybrid relative to that in parental species B. Line colours and styles are identical to the panel A. Note: Square and diamond symbols along the lines provide the link between panels A and B by showing two reference values of  $\log_2(f_A/f_B)$  expression divergence between parental species: square:  $\log_2(f_A/f_B) = -1$  and diamond:  $\log_2(f_A/f_B) = 1$ .

the relative expression of hybrid's alleles  $A_{hyb}$  and  $B_{hyb}$ ,  $\log_2(f_H^A/f_H^B)$ , and the expression divergence between the two parents,  $\log_2(f_A/f_B)$ , was non-linear when the cross regulation of the subgenomes was limited (refer to Fig. 7A). When parental strains differ in expression levels (i.e.  $f_A \neq f_B$ ), the TFs derived from the more highly expressing parent exhibit higher concentrations within the hybrid's nucleus compared to those from the less expressing parent. As a result, the TFs from the more expressing parent exert a stronger influence on the hybrid's promoters and disproportionately strongly modulate the fractional occupancies of promoters derived from the less expressing parent (see Fig. 7B). The size of this effect increases as the degree of expressive divergence between the parents increases. Consequently, as the expression divergence increases, the first derivation of  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B)$  correlation decreases and the curve becomes "flatter" towards the extremes of the *x*-axis of Fig. 7A.

Such nonlinearity is more pronounced in polyploid organisms with asymmetric genome dosage, such as AAB or ABB triploids, compared to symmetric hybrids like AB diploids or AABB tetraploids.

To understand why asymmetric genomic dose matters, let us consider the case where expression in species A is greater than in species B ( $f_A > f_B$ ; Fig. 7A–right part of the *x*-axis). In such cases, the

proportion of A-derived TFs is higher in AAB triploids than in symmetric hybrids and consequently, the A-derived TFs have an even greater influence on the fractional occupancies of B-derived promoters (Fig. 7B, upper left quadrant). Similarly, the downregulation of A-derived promoters in AAB triploids is less pronounced than in symmetric hybrids (Fig. 7B, upper left quadrant). A mirror effect is observed in cases of ABB triploids when  $f_A < f_B$  (Fig. 7B, lower right quadrant). Therefore, when  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B)$  correlations are fitted with linear regression as in refs. 20,22,24, one observes a smaller slope in asymmetrical AAB or ABB triploids compared to AB diploids or symmetric AABB polyploids, which has been empirically shown e.g. in ref. 5.

These findings suggest that interpretations of cis/trans-regulation types, which are based on RAE should take into consideration the non-linear nature of the  $\log_2(f_H^A/f_H^B) \sim \log_2(f_A/f_B)$  correlation, requiring alternative approaches. Furthermore, our findings provide an explanation for the impression of greater efficiency of trans-regulatory crosstalk between subgenomes in asymmetrical polyploids compared to diploid hybrids, as reported e.g. by Bartoš et al.<sup>5</sup>, who used Shi et al.'s<sup>24</sup> model to compare the distribution of RAE in diploid and triploid fish hybrids.

## Utility and shortcomings of equilibrium physicochemical models for predicting hybrid's and polyploid's expression and future directions

A growing amount of data on hybrid and polyploid organisms brings us closer to understanding the molecular processes underpinning the wide diversity of phenotypic traits possessed by these organisms. Bottani et al.<sup>30</sup> made a pioneering step by demonstrating that the subgenome dominance, a prominent pattern, may arise instantaneously as a direct consequence of genome merging from differently adapted parental species onto an allopolyploid, rather than as a result of subsequent evolution. However, while the assumptions in their model provided valuable insights, they also somewhat limited the exploration of more complex interactions, particularly in distinguishing the roles of genome merging from those of genome multiplying. The present versatile thermodynamic model (Eqs. 7–11 in section Methods below) indicated that even wider array of empirical observations may be explained as a direct and immediate consequence of the mixing of regulatory networks with diverged cis- and trans- elements through interspecific hybridization. In addition to subgenome dominance, other well-known patterns may be depicted by thermodynamic model, like the effects of trans-regulatory divergence between parental species on the cross-regulation between subgenomes in hybrids or the non-additive expression with frequent expression-level dominance. It also depicted some less well studied patterns, like the overall transcriptome downsizing in hybrids, and indicated how the polyploidization may affect the efficiency of trans-regulation of a hybrid's alleles via modifications of the subgenome dosage rather than via mere multiplication of the subgenomes.

The presented model also highlights three important aspects that have implications for assessing gene expression regulation pathways. Firstly, it suggests a need to revisit some commonly used measures of cis-/trans-regulation based on RAE (e.g., refs. 20,24, see Fig. 1B) since the correlation between RAE values and parental expression divergence is not necessarily linear (Fig. 7A). The model also indicated that even if the gene is under strict trans-regulation, it may appear fully cis-regulated, depending on the interparental divergence. Finally, third, our findings indicate that certain patterns observed in hybrids and polyploids rely on specific assumptions about how components of regulatory network function and evolve. For example, unequal cell volumes of hybridizing species may trigger the subgenome dominance but only when a significant proportion of TF molecules are in a free state (see the section about subgenome dominance above and compare Fig. 6B with Fig. 6C). Similarly, subgenome dominance induced by an unequal euchromatic ratio may occur under full cross-regulation between subgenomes, but only if the parental species has adapted to increased euchromatic ratios by modifying specific binding sites rather than TF properties themselves (see the section about subgenome dominance above and compare Fig. 6D with Fig. 6E).

These results imply that understanding how traits are expressed in hybrids and polyploids critically depends on the available knowledge and assumptions regarding the distribution of TF molecules within a nucleus and their coevolution with other components of the regulatory network. For instance, while mechanisms of protein to DNA binding are extensively studied and TF binding sites have recently been shown to be under increased rates of adaptive evolution<sup>39</sup>, it remains unclear how organisms adapt to changes in the euchromatic ratio in their genomes. Moreover, models investigating controlled TF binding commonly assumes that trans-acting factors create an environment common to all cis-regulatory elements (reviewed in ref. 26, Fig. 2). However, TF movement can take different forms, ranging from sliding along chromosomes to free movement through nucleoplasm<sup>40,41</sup>. This variation affects the presumed concentration of free TF molecules within a cell, and as our model demonstrates, this parameter significantly impacts the predicted effects of hybridization and polyploidization.

The obtained results underscore the utility of TF-promoter binding models in improving our understanding of trait expression in hybrid and polyploid organisms. However, it is important to acknowledge that there are still limitations in thermodynamic models that may have unknown impacts on their applicability to real-world scenarios<sup>42</sup>. Theoretical models, including the one presented here, often use various simplifying assumptions, such as using TF-promoter binding as a proxy for gene expression or not considering the complexity of TFs as multimeric molecular complexes, e.g. refs. 1,26,30. The binding events at nucleic acid chains are thus considered as independent (i.e. not affecting each other either in cooperative or repulsive ways) and the presence of other DNA-binding proteins is neglected, but more complex models may take cooperativity into account, as made e.g. in ref. 42. Additionally, thermodynamic models inherently assume equilibrium states. In reality, however, cellular environments are non-equilibrium and the link to RNA or protein expression is more intricate than simple TF binding.

The agreement of our model with empirical observations should also be interpreted with caution. It indeed aligns with numerous empirical data suggesting that when expression-level dominance occurs, hybrids generally match the more expressing parent, i.e. expression-level dominance UP prevails over DOWN<sup>5,6,35,36</sup>. However, opposite patterns or prevalent expression-level dominance DOWN patterns have also been observed in certain organisms<sup>38</sup>, and UP versus DOWN patterns may also reflect the ancestry of particular subgenomes<sup>37</sup>.

These observations indicate that the mechanisms underlying expression in hybrids and allopolyploids are influenced by multiple factors, necessitating further refinement of theoretical models. Hybrids and allopolyploids combine two sets of trans-acting factors, which may interact in various ways within a common nucleus, ranging from redundancy to diverse oppositional or compensatory effects. Hu and Wendel<sup>12</sup> further argue against assuming that trans-acting regulators and cell architecture are additively inherited in a polyploid nucleus since the same regulatory circuits may not necessarily be shared between parental species, even when their expression outputs are equivalent<sup>43</sup>.

Despite these limitations, the observed correspondence with empirical observations and the inferred dependency of patterns on the type of regulatory network function suggests that integrating models of TF-promoter binding in systems biology holds promise for enhancing our understanding of the evolutionary consequences of hybridization and polyploidization. Further investigation of their strengths and weaknesses, specifically regarding hybrids and polyploids, could provide insights into how divergence in cis-/trans-regulatory networks among parents directly translates into the observed patterns in hybrids and polyploids.

## Methods

In this section, we briefly outline the Bottani et al.<sup>30</sup> model and we then present our improved and more complex model along with six modelling schemes we considered.

### Models of TF-promoter binding and application to hybrids and polyploids

Bottani et al.<sup>30</sup> modelled the fractional occupancy ( $f$ ) of the promoter (which can be considered as a proxy of transcriptional activity) in the parental species such as

$$f_i = \frac{[TF_i] e^{-E_i/k_B T}}{[TF_i] e^{-E_i/k_B T} + [NS_i] e^{-E_{NS}/k_B T}} \quad (1)$$

where  $[TF_i]$  refers to concentration of a TF in the parental species,  $[NS_i]$  refers to a concentration of nonspecific sites in a genome,  $e^{-E_i/k_B T}$  describes the binding energy between TF and specific site, and

$e^{-E_{NS}/k_B T}$  describes binding energy between TF and nonspecific site. To describe the fractional occupancies ( $f'_i$ ) of orthologous promoters from both parental subgenomes in the allopolyploid (Eqs. 3 and 4 in their Appendix), Bottani et al.<sup>30</sup> used the equation for one orthologue:

$$f'_1 = \frac{[TF_1]'e^{-E_1/k_B T} + [TF_2]'e^{-E_{21}/k_B T}}{[TF_1]'e^{-E_1/k_B T} + [TF_2]'e^{-E_{21}/k_B T} + [NS]e^{-E_{NS}/k_B T}} \quad (2)$$

and an analogous formula for  $f'_2$  (their Eq. 5 in the Appendix) for the other orthologue. Here, the indices “1” and “2” refer to the origin from parent A and B, respectively, so that  $e^{-E_i/k_B T}$  describes the binding energy between TF and specific site, both originating from the species  $i$ , while  $e^{-E_{ji}/k_B T}$  refers to the binding energy between TF originating from species  $j$  and a specific site originating from species  $i$ .

In essence, Bottani et al.’s<sup>30</sup> approach stems from a mechanistic model as one can find in Chu et al.’s<sup>26</sup> Eq. 13, which assumes the interaction of single specific binding site with TFs that are in a large excess. Binding of a TF molecule to a specific site has negligible effect on their overall concentration and hence fractional occupancies  $f'_1, f'_2$  in a common nucleus may be calculated independently of each other.

Such an approach is useful as it reduces the number of equations necessary to solve the complex thermodynamic system describing allopolyploid organism, making it therefore mathematically more tractable. However, it brings along several simplifications, which may affect the interpretability of the model with respect to real biological scenarios. Namely, by using their simplified equations, Bottani et al.<sup>30</sup> implicitly considered two independent specific binding sites in the hybrid’s cell. This may not sufficiently differentiate between diploid hybrid of AB genomic constitution and an allotetraploid of AABB constitution that in reality should have four copies of each gene. Consequently, in our opinion, Bottani et al.’s<sup>30</sup> model may not distinguish the effect of symmetric genome multiplication (e.g., AABB tetraploid) from mere hybridization. This poses no problem if one wants to evaluate the effect of mixing diverged regulatory networks on the expression of alleles. However, it may complicate disentangling the effects of pure polyploidization from those of genome mixing (hybridization) and may not necessarily address the situation in asymmetric polyploids, like triploids.

We further explore the implications of Bottani et al.’s<sup>30</sup> model, particularly regarding hybridization between species with unequal euchromatic ratios. Bottani et al.<sup>30</sup> suggest that the parent with higher euchromatic ratio may modify its TF molecules to increase their affinity for specific binding and consequently, allopolyploids should express subgenome dominance even under full cross-regulation between subgenomes from both parental species. However, our analysis indicates that their model might also be interpreted as if the species with higher non-specific site concentration (species B) is modifying its promoter sites to increase affinity for any TFs, without any changes to the properties of the TF molecules themselves. This interpretation appears from the definition of the coefficient  $C_1$  controlling the cross-regulation of specific binding sites in Eq. 4 in Appendix of ref. 30:

$$f'_1 = \frac{[TF_1]' + [TF_2]'C_1}{[TF_1]' + [TF_2]'C_1 + [NS]K_1} \quad (3)$$

Here,  $e^{-E_{21}/k_B T} = C_1 e^{-E_1/k_B T}$  and  $e^{-E_{NS}/k_B T} = K_1 e^{-E_1/k_B T}$ . Coefficient  $K_1$  describes how much lower is the affinity of A-type TFs for non-specific binding sites relative to their affinity for A-type specific sites and coefficient  $K_2$  is defined analogously for B-type TFs. Meanwhile, the coefficient  $C_1$  describes the relative affinity of heterospecific TFs from species B to specific binding site A compared to the affinity of A-type TFs for their conspecific A-type site and  $C_2$  is similarly defined for B-type specific binding sites.  $C_1$  and  $C_2$  can range from zero (no cross-regulation) to one (full cross-regulation). When cross-regulation is

limited, the subgenome dominance would occur, irrespective of how the parental species adapted to increased euchromatic ratio. However, under full cross-regulation, setting both  $C_1$  and  $C_2$  to one implies 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  $E_1 \neq E_2$ ), but it does not differentiate between TF molecules from both parental species (i.e.,  $E_1 = E_{21}$  and  $E_2 = E_{12}$ ). This view diverges from the verbal interpretation provided by Bottani et al.<sup>30</sup>, where adaptation through modification of TFs was suggested. Our analysis suggests that if this was the case, the equality between energies would be  $E_1 = E_{12}$  and  $E_2 = E_{21}$ . Thus, our findings propose that Bottani et al.’s<sup>30</sup> model aligned with a scenario assuming modification of the binding sites rather than TF molecules.

This nuance is significant as it opens new avenues for understanding how subgenome dominance can manifest in relation to the evolutionary adaptations of binding sites versus TF molecules. While binding sites of TFs are known to undergo accelerated evolution<sup>39</sup>, it remains an open question whether such adaptations would occur simultaneously in all or most gene promoters or in fewer genes coding for the respective TF molecules and their binding motifs, which presumably represent greater mutational target size, relative to cis-regulatory mutations<sup>44</sup>. Our study aims to contribute to this ongoing discussion, offering insights that might refine our understanding of regulatory adaptations in hybrid and polyploid species.

## Presentation of generalized thermodynamic model

**Description of the model.** To alleviate limitations inherent in previous models, we introduce a thermodynamic model of mass-balance equations relating the equilibrium and total concentrations of the regulatory components involving the TF molecules of types A and B (i.e.  $TF_A$  and  $TF_B$ , respectively), the specific binding sites of type A and B (i.e.  $S_A$  and  $S_B$ , respectively) and non-specific binding sites (N). Indices ‘A’ and ‘B’ describe the origin from parental species A and B (Fig. 2). We assume that TFs can occur in four states, i.e. free (i), bound to either specific binding site of type A (ii) or type B (iii), or bound to non-specific binding sites N (iv). Analogously, each binding site  $S_A$ ,  $S_B$  or N can be either empty or occupied by conspecific or heterospecific TF molecule. Although the binding sites are formally present on one macromolecule of DNA, we consider them to interact with TFs individually.

In a single parental species A, the model reduces to three mass-balance equations (Fig. 2) with 3 parameters of total concentrations of TFs,  $c(TF_A)$ , specific binding sites,  $c(S_A)$ , and non-specific binding sites  $c(N)$ :

$$c(TF_A) = [TF_A] + \frac{[TF_A][S_A]}{K_{AA}} + \frac{[TF_A][N]}{K_{AN}}, \quad (4)$$

$$c(S_A) = [S_A] + \frac{[TF_A][S_A]}{K_{AA}}, \quad (5)$$

$$c(N) = [N] + \frac{[TF_A][N]}{K_{AN}}. \quad (6)$$

This model outputs the concentrations of free TFs ( $[TF_A]$ ), empty specific ( $[S_A]$ ) and non-specific ( $[N]$ ) binding sites and is applicable analogously for the parent B.

In hybrids, however, the situation is more complex since the model requires 5 total concentrations of both types of TFs,  $c(TF_A)$  and  $c(TF_B)$ , both types of specific binding sites,  $c(S_A)$  and  $c(S_B)$ , and of non-specific binding sites  $c(N)$ . Hence, the cross-terms describing the bindings of orthologous sets of TFs and promoters must be



considered, giving rise to a system of five equations:

$$c(\text{TF}_A) = [\text{TF}_A] + \frac{[\text{TF}_A][\text{S}_A]}{K_{AA}} + \frac{[\text{TF}_A][\text{S}_B]}{K_{AB}} + \frac{[\text{TF}_A][\text{N}]}{K_{AN}}, \quad (7)$$

$$c(\text{TF}_B) = [\text{TF}_B] + \frac{[\text{TF}_B][\text{S}_A]}{K_{BA}} + \frac{[\text{TF}_B][\text{S}_B]}{K_{BB}} + \frac{[\text{TF}_B][\text{N}]}{K_{BN}}, \quad (8)$$

$$c(\text{S}_A) = [\text{S}_A] + \frac{[\text{TF}_A][\text{S}_A]}{K_{AA}} + \frac{[\text{TF}_B][\text{S}_A]}{K_{BA}}, \quad (9)$$

$$c(\text{S}_B) = [\text{S}_B] + \frac{[\text{TF}_A][\text{S}_B]}{K_{AB}} + \frac{[\text{TF}_B][\text{S}_B]}{K_{BB}}, \quad (10)$$

$$c(\text{N}) = [\text{N}] + \frac{[\text{TF}_A][\text{N}]}{K_{AN}} + \frac{[\text{TF}_B][\text{N}]}{K_{BN}}. \quad (11)$$

Such a set of five equations describes the scenario of hybridization between two parental species A and B and the formation of their hybrids of various ploidy levels (Fig. 2), which are specified by the concentration vector  $C$ ; see Supplementary note 2. The model is then calculated independently for each type of hybrid.

The affinities of binding sites and TFs are characterized by six dissociation constants  $K$ , which describe conspecific interactions

$$K_{AA} = \frac{[\text{TF}_A][\text{S}_A]}{[\text{TF}_A\text{S}_A]}, \quad (12)$$

$$K_{BB} = \frac{[\text{TF}_B][\text{S}_B]}{[\text{TF}_B\text{S}_B]}, \quad (13)$$

heterospecific interactions

$$K_{AB} = \frac{[\text{TF}_A][\text{S}_B]}{[\text{TF}_A\text{S}_B]}, \quad (14)$$

$$K_{BA} = \frac{[\text{TF}_B][\text{S}_A]}{[\text{TF}_B\text{S}_A]}, \quad (15)$$

and non-specific interactions

$$K_{AN} = \frac{[\text{TF}_A][\text{N}]}{[\text{TF}_A\text{N}]}, \quad (16)$$

$$K_{BN} = \frac{[\text{TF}_B][\text{N}]}{[\text{TF}_B\text{N}]}. \quad (17)$$

(in our notation of  $K_{ij}$ ,  $i$  stays for the origin of TFs,  $j$  for the origin of binding sites).

By setting the values of the five total concentrations (left hand sides of Eqs. 7–11) and the six dissociation constants (Eqs. 12–17), the model finds for each type of hybrid a numerical solution that provides concentrations of free TFs and of empty binding sites. The model then allows for the calculation of fractional occupancies of both types of specific binding sites in a hybrid, which are normalized per number of chromosome copies and represent the ratio of occupied binding sites to the total number of binding sites of a particular type.

In particular, fractional occupancy  $f_A$  of the specific binding sites in the parental species A is expressed as

$$f_A = 1 - \frac{[\text{S}_A]}{c(\text{S}_A)} = \frac{[\text{TF}_A][\text{S}_A]}{c(\text{S}_A)K_{AA}} \quad (18)$$

since both alleles originate from the same parental species A.

Analogously, the fractional occupancy  $f_B$  in the parental species B is expressed as

$$f_B = 1 - \frac{[\text{S}_B]}{c(\text{S}_B)} = \frac{[\text{TF}_B][\text{S}_B]}{c(\text{S}_B)K_{BB}}. \quad (19)$$

In the case of hybrid (H) or allopolyploid, the situation is more complex since its nucleus contains at least two alleles originated from different parental species, each potentially having different occupancies. Hence, in the diploid hybrid AB and triploid hybrids AAB and ABB, the fractional occupancy of A-type specific sites  $f_H^A$  is calculated as:

$$f_H^A = f_{AB}^A = f_{AAB}^A = f_{ABB}^A = 1 - \frac{[\text{S}_A]}{c(\text{S}_A)} = \frac{\frac{[\text{TF}_A][\text{S}_A]}{K_{AA}} + \frac{[\text{TF}_B][\text{S}_A]}{K_{BA}}}{c(\text{S}_A)} = \left( \frac{[\text{TF}_A]}{K_{AA}} + \frac{[\text{TF}_B]}{K_{BA}} \right) \frac{[\text{S}_A]}{c(\text{S}_A)} \quad (20)$$

and the fractional occupancy of B-type specific sites as:

$$f_H^B = f_{AB}^B = f_{AAB}^B = f_{ABB}^B = 1 - \frac{[\text{S}_B]}{c(\text{S}_B)} = \frac{\frac{[\text{TF}_A][\text{S}_B]}{K_{AB}} + \frac{[\text{TF}_B][\text{S}_B]}{K_{BB}}}{c(\text{S}_B)} = \left( \frac{[\text{TF}_A]}{K_{AB}} + \frac{[\text{TF}_B]}{K_{BB}} \right) \frac{[\text{S}_B]}{c(\text{S}_B)}. \quad (21)$$

Although fractional occupancies  $f_{AB}^A, f_{AAB}^A, f_{ABB}^A$  and  $f_{AB}^B, f_{AAB}^B, f_{ABB}^B$ , respectively, are expressed by the same formulas, their values in particular hybrid types depend on concentration of free TFs ( $[\text{TF}_A]$  and  $[\text{TF}_B]$ ) and free binding sites ( $[\text{S}_A]$  and  $[\text{S}_B]$ ) as calculated by Eqs. 7–11. These values differ among AB, AAB and ABB hybrids depending on particular genomic doses controlled by the concentration vector  $C$ ; see Supplementary Note 2. The resulting fractional occupancies are presented in Figs. 3–7, utilizing the R statistical language and its ggplot2 library<sup>45</sup>.

**Versatility of the model and simulated scenarios.** Our model differs from the model used by Bottani et al.<sup>30</sup>, as it relaxes some of their assumptions allowing us to simulate a range of biologically relevant scenarios. We thus explicitly test how the subgenome dominance arises upon hybridization between species with different euchromatic ratios, depending on whether adaptation to an increased amount of non-specific sites occurs via modification of TFs or via modification of promotor sites. We also explore additional scenarios, such as the effects of asymmetric cell volumes between hybridizing parental species, restricted cross-regulation between subgenomes, as well as alternative assumptions of existence of free TFs affecting their movements between binding sites. All these scenarios can be simulated using simple changes in model parameters, which are summarized in Supplementary Note 2 and described in the following paragraphs. In total, we simulated six different modelling schemes:

- (1) To begin our simulations, we establish a model setup similar to Bottani et al.<sup>30</sup> as a reference point. In this setup, we assume that regulatory divergence between conspecific and heterospecific TF-promotor bindings arises from evolution of promotor sequence, rather than from changes in TFs. In addition, we initially assume that only a negligible proportion of TFs exists in a free state, and virtually all TFs are bound to specific or non-specific sites, but this

condition will be tested in alternative setups. For each model, we initially set the fractional occupancies of specific sites to be the same for both parental species ( $f_A = f_B = 0.5$ ), resulting from the interplay between the concentrations and dissociation constants. Next, we simulate situations where parental species differ in their gene expression levels, resulting in  $\log_2(f_A/f_B)$  ratios varying from  $-1$  to  $1$ . We achieve this change in expression solely by modifying the concentrations of TFs in both parental species.

Having set the reference set up, we explored the effects of two distinct modelling assumptions:

- (2) Number of specific binding sites: To account for the fact that particular TF may regulate many genes simultaneously<sup>1</sup>, we compared the reference scenario assuming a single specific binding site (gene promoter) per subgenome, with a situation where TFs may find more specific promoter sites. We modelled this alternative scenario assuming 150 binding sites per subgenome, as in Veitia et al.<sup>1</sup>.
- (3) Type of TF movement among binding sites: To address various types of TFs trafficking throughout the cell<sup>26</sup>, we compared the reference scenario which assumes that TFs do not occur in free states, with a situation where a considerable portion of TFs is free in the cytoplasm. We achieved this alternative model by balancing the concentration of TFs and dissociation constants, simulating a situation where TFs may be present in a nucleoplasm when not bound to specific or non-specific sites<sup>26,31</sup>.

We then modified model coefficients to test four biological scenarios relevant to understanding how parental expression differentiation affects hybrids and allopolyploids.

- (4) Firstly, we investigated how the divergence in trans-regulatory elements between parental species impacts on the fractional occupancy of promoters in hybrids. Such divergence was simulated by modifying parameter values of corresponding dissociation constants that ultimately control the cross-specificity of affinities between transcription factors and promoters (TF-S). To do so, we kept fixed the ratio of specific to non-specific affinity constants ( $K_{ii}/K_{in} = 10^4$ ; where  $i$  stands for origin from A or B species). However, we controlled the cross-regulation between orthologous sets of TFs and promoters through the parameter  $\rho$ , which characterizes the ratio of the corresponding “cross-dissociation constants”. Depending on whether the trans-regulatory divergence is driven by changes of binding sites as assumed in Bottani et al.’s<sup>30</sup> model, or by TF molecules (see modelling assumptions 6.a and 6.b below), the parameter  $\rho$  equals the ratios  $K_{BA}/K_{AA}$  and  $K_{AB}/K_{BB}$  or the ratios  $K_{AB}/K_{AA}$  and  $K_{BA}/K_{BB}$ , respectively (see Supplementary Note 2 for details).

Namely, we investigated three scenarios:

- a. *full cross-regulation* was simulated by setting TF-S affinity uniform irrespective of the ancestral origin of molecules (here,  $\rho = 1$ ), suggesting that  $K_{BA}/K_{AA} = K_{AB}/K_{BB} = 1$  or, respectively, that  $K_{AB}/K_{AA} = K_{BA}/K_{BB} = 1$ .
- b. *limited cross-regulation* where the binding affinity of hetero-specific TF-S interactions is two times lower than of conspecific ones ( $\rho = 2$ , and hence,  $K_{BA}/K_{AA} = K_{AB}/K_{BB} = 2$  or, respectively,  $K_{AB}/K_{AA} = K_{BA}/K_{BB} = 2$ ).
- c. *no cross-regulation* where the affinity between TFs and heterospecific copies of promoters (e.g.,  $TF_A S_B$ ) is as low as the binding of a TF to a non-specific site. In other words, in the no cross-regulation scenario, TFs in a hybrid do not bind to orthologous regulatory elements with higher affinity than to nonspecific sites. In such a case,  $\rho = 10^4$  and  $K_{BA} = K_{AB} = K_{AN}$ ; see Supplementary Note 2).
- (5) Next, we investigated the role of unequal cell volumes in parental species. We assumed a scenario where one parental species had a cell volume twice as large as the other, resulting in lower concentrations of specific and non-specific binding sites. To maintain

the same fractional occupancy of its promoters, the species increased the concentration of its TFs without modifying their binding affinities. After genome merging, hybrids and allopolyploids had average cell sizes between the respective ancestors, and the concentrations of their TFs were calculated from their parental values as in Bottani et al.<sup>30</sup>; see Supplementary Note 2.

- (6) Finally, we examined the effects of unequal amounts of non-specific binding sites among parents. This scenario simulates asymmetries in euchromatic ratios among hybridizing species, and assumes that the parental species with a higher euchromatic ratio had adaptively evolved its TF-promotor interactions to mitigate the higher concentration of non-specific binding sites<sup>30</sup>. Two types of such adaptation were tested.

- a. Firstly, we simulated the scenario assuming that adaptation to intracellular conditions is driven by evolution of promotor sites towards binding any TFs with higher affinity (the scenario actually modelled by Bottani et al.<sup>30</sup>).
- b. Secondly, we employed the alternative scenario suggesting that adaptation occurred via evolution of properties of TF molecules towards higher binding affinities.

These contrasting scenarios 6.a and 6.b were modelled by modifying how the value of the cross-regulatory parameter  $\rho$  controls the trans-regulatory interactions between conspecific and heterospecific TF-promotor binding. Specifically, the effect of evolution of promotor sites was modelled by controlling the  $K_{BA}/K_{AA}$  and  $K_{AB}/K_{BB}$  ratios while the effect of TF evolution was modelled by controlling the  $K_{AB}/K_{AA}$  and  $K_{BA}/K_{BB}$  ratios; see Supplementary Note 2 for details and precise use of parameter  $\rho$ .

## Data availability

No biological data were generated in this paper; all results are presented.

## Code availability

The code processed by octave software is described in Supplementary Code 1.

## References

1. Veitia, R. A., Bottani, S. & Birchler, J. A. Gene dosage effects: non-linearities, genetic interactions, and dosage compensation. *Trends Genet.* **29**, 385–393 (2013).
2. Yoo, M.-J., Liu, X., Pires, J. C., Soltis, P. S. & Soltis, D. E. Nonadditive gene expression in polyploids. *Annu. Rev. Genet.* **48**, 485–517 (2014).
3. Comeault, A. A. & Matute, D. R. Genetic divergence and the number of hybridizing species affect the path to homoploid hybrid speciation. *Proc. Natl Acad. Sci. USA* **115**, 9761–9766 (2018).
4. Stöck, M. et al. Sex chromosomes in meiotic, hemiclinal, clonal and polyploid hybrid vertebrates: along the ‘extended speciation continuum’. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200103 (2021).
5. Bartoš, O. et al. The legacy of sexual ancestors in phenotypic variability, gene expression, and homoeolog regulation of asexual hybrids and polyploids. *Mol. Biol. Evol.* **36**, 1902–1920 (2019).
6. Li, M., Wang, R., Wu, X. & Wang, J. Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid *Brassica napus*. *BMC Genom.* **21**, 330 (2020).
7. Gianinetti, A. A criticism of the value of midparent in polyploidization. *J. Exp. Bot.* **64**, 4119–4129 (2013).
8. Matos, I., Machado, M. P., Schartl, M. & Coelho, M. M. Gene expression dosage regulation in an allopolyploid fish. *PLoS ONE* **10**, e0116309 (2015).
9. Zhang, M. et al. Effects of parental genetic divergence on gene expression patterns in interspecific hybrids of *Camellia*. *BMC Genom.* **20**, 828 (2019).

10. Wang, X., Morton, J. A., Pellicer, J., Leitch, I. J. & Leitch, A. R. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.* **107**, 1003–1015 (2021).
11. Goncalves, A. et al. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.* **22**, 2376–2384 (2012).
12. Hu, G. & Wendel, J. F. Cis-trans controls and regulatory novelty accompanying allopolyploidization. *N. Phytol.* **221**, 1691–1700 (2019).
13. Mattioli, K. et al. Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biol.* **21**, 210 (2020).
14. Ren, L. et al. The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* × *Culter alburnus*. *Genome Res.* **29**, 1805–1815 (2019).
15. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).
16. Tulchinsky, A. Y., Johnson, N. A., Watt, W. B. & Porter, A. H. Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. *Genetics* **198**, 1155–1166 (2014).
17. Botet, R. & Keurentjes, J. J. B. The role of transcriptional regulation in hybrid vigor. *Front. Plant Sci.* **11**, 410 (2020).
18. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
19. GTEx Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
20. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
21. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**, 346–350 (2008).
22. Tirosh, I., Reikhav, S., Levy, A. A. & Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**, 659–662 (2009).
23. Emerson, J. J. et al. Natural selection on cis and trans regulation in yeasts. *Genome Res.* **20**, 826–836 (2010).
24. Shi, X. et al. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat. Commun.* **3**, 950 (2012).
25. Osada, N., Miyagi, R. & Takahashi, A. Cis- and trans-regulatory effects on gene expression in a natural population of *Drosophila melanogaster*. *Genetics* **206**, 2139–2148 (2017).
26. Chu, D., Zabet, N. R. & Mitavskiy, B. Models of transcription factor binding: sensitivity of activation functions to model assumptions. *J. Theor. Biol.* **257**, 419–429 (2009).
27. Mueller, F., Stasevich, T. J., Mazza, D. & McNally, J. G. Quantifying transcription factor kinetics: at work or at play? *Crit. Rev. Biochem. Mol. Biol.* **48**, 492–514 (2013).
28. Porter, A. H., Johnson, N. A. & Tulchinsky, A. Y. A new mechanism for mendelian dominance in regulatory genetic pathways: competitive binding by transcription factors. *Genetics* **205**, 101–112 (2017).
29. Okubo, K. & Kaneko, K. Evolution of dominance in gene expression pattern associated with phenotypic robustness. *BMC Ecol. Evol.* **21**, 110 (2021).
30. Bottani, S., Zabet, N. R., Wendel, J. F. & Veitia, R. A. Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci.* **23**, 393–402 (2018).
31. Spivakov, M. Spurious transcription factor binding: non-functional or genetically redundant? *Bioessays* **36**, 798–806 (2014).
32. An, H., Pires, J. C. & Conant, G. C. Gene expression bias between the subgenomes of allopolyploid hybrids is an emergent property of the kinetics of expression. *PLOS Computat. Biol.* **20**, e1011803 (2024).
33. Coate, J. E. & Doyle, J. J. Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. *Genome Biol. Evol.* **2**, 534–546 (2010).
34. Coate, J. E. & Doyle, J. J. Variation in transcriptome size: are we getting the message? *Chromosoma* **124**, 27–43 (2015).
35. Yoo, M.-J., Szadkowski, E. & Wendel, J. F. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171 (2013).
36. Combes, M.-C. et al. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol. Evol.* **7**, 1110–1121 (2015).
37. Bell, G. D. M., Kane, N. C., Rieseberg, L. H. & Adams, K. L. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* **5**, 1309–1323 (2013).
38. Ren, L. et al. Determination of dosage compensation and comparison of gene expression in a triploid hybrid fish. *BMC Genom.* **18**, 38 (2017).
39. Zhang, X., Fang, B. & Huang, Y.-F. Transcription factor binding sites are frequently under accelerated evolution in primates. *Nat. Commun.* **14**, 783 (2023).
40. Wunderlich, Z. & Mirny, L. A. Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.* **36**, 3570–3578 (2008).
41. Stoof, R., Wood, A. & Goñi-Moreno, Á. A model for the spatio-temporal design of gene regulatory circuits †. *ACS Synth. Biol.* **8**, 2007–2016 (2019).
42. Bottani, S. & Veitia, R. A. Hill function-based models of transcriptional switches: impact of specific, nonspecific, functional and nonfunctional binding. *Biol. Rev. Camb. Philos. Soc.* **92**, 953–963 (2017).
43. Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006).
44. Metzger, B. P. H., Wittkopp, P. J. & Coolon Joseph, D. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* **9**, 843–854 (2017).
45. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* 1st edn, Vol. 213 (Springer, New York, NY, 2009).

## Acknowledgements

Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic. K.J. and J.E. were supported by the Czech Science Foundation Project No. 21-25185S and by the Ministry of Education, Youth and Sports of the Czech Republic (grant No. 539 EXCELLENCE CZ.02.1.01/0.0/0.0/15\_003/0000460 OP RDE. T.T. was supported by the Ministry of Education, Youth and Sports of the Czech Republic—for the project Reproductive and Genetic Procedures for the Conservation of Fish Biodiversity and Aquaculture (CZ.02.1.01/0.0/0.0/16\_025/0007370). P.C. was supported by the Ministry of Education, Youth, and Sports of the Czech Republic (project No. CZ.02.01.01/00/22\_008/0004558, co-funded by the European Union) and the Czech Academy of Sciences—Strategy AV21 (VP29). Institute of Animal Physiology and Genetics receives support from Institutional Research Concept, Grant/Award Number: RVO67985904.

## Author contributions

Definition of hypothesis—K.J.; study conceptualization—K.J., J.E. and P.C.; funding acquisition—K.J. and P.C.; Mathematical modelling and investigation of model parameters—J.E. and P.C.; data and model visualization—T.T.; literature reviewing—K.J. and T.T.; writing initial draft—K.J.; MS finalization—K.J., J.E., P.C. and T.T.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52546-5>.

**Correspondence** and requests for materials should be addressed to Karel Janko or Petr Cigler.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024