Supplementary Note 1

Gene and allele expression types and their inference in hybrid organisms

I. Types of gene expression levels in hybrids:

To investigate how the genes in hybrids are expressed relative to their levels in parental species, Yoo et al. [1] proposed classification of individual genes into the so-called ***expression inheritance categories*** (see Figure S1). These assume ***additivity*** (categories I and XII) if the given gene's expression significantly differs between both parental species, whereas the hybrid's expression is intermediate, ***transgressivity*** when the hybrid's expression is significantly inferior (categories III, VII, and X) or superior (categories V, VI and VIII) to both parental species or ***expression-level dominance*** (categories II, IV, IX, and XI) when both parental species significantly differ from each other and the hybrid significantly differs from one but not the other parent. The last categories may be of the type ***expression-level dominance UP***, when the hybrid matches the more expressing parent, or ***expression-level dominance DOWN***, when the hybrid matches the less expressing parent.

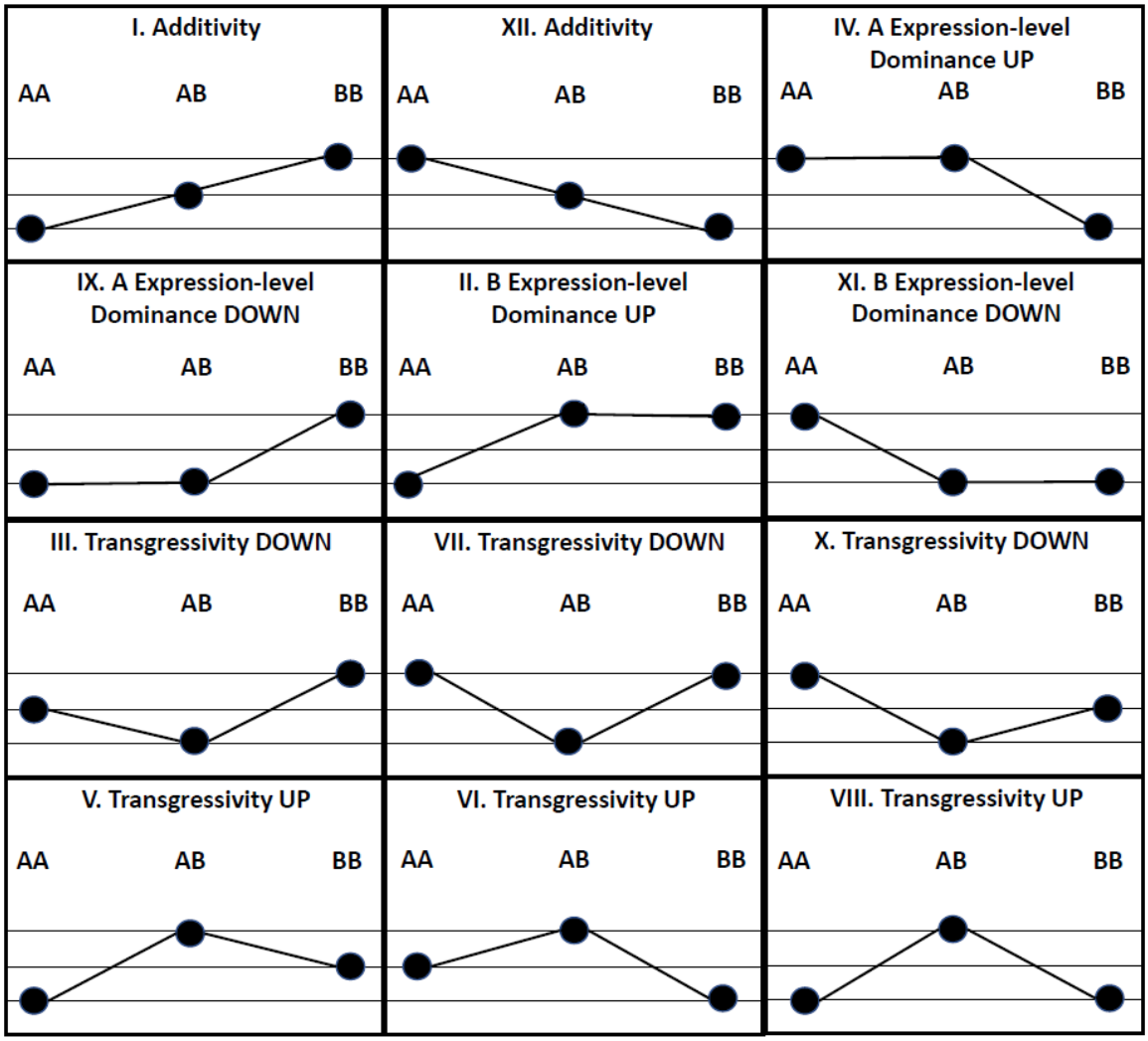

Figure S1: Gene expression inheritance categories sensu Yoo et al. [1]. For each of the twelve original categories the pictograms indicate expression levels in parental species A and B whose diploid genomes are denoted as AA and BB, while hybrid is denoted as AB.

II. Inference of cis-/trans-regulatory effects on allelic expression:

There are numerous variants of methods how allele-specific expression data may be used to evaluate the relative contribution of cis- and trans-regulation to a gene's expression (rev. e.g. in [2]). A commonly used method has been proposed [3–5] to classify genes into regulation-type categories depending on how the expression divergence between parental species is conserved in the relative allelic expression (RAE) of orthologous and paralogous gene copies in hybrids (see Figure 1B in the main text). In other words, the classification stems from the correlation between the relative expression of hybrid's alleles A_{hyb} and B_{hyb} (i.e. RAE) in a hybrid and the expression divergence between the two parental species A and B (i.e. $\log_2$ fold change between expression of this gene in species A and B).

Specifically, there are four categories.

1) A gene is supposedly ***under trans- regulation*** by both sets of orthologous transcription factors (TFs) when its expression differs between parental species A and B (i.e. when expression in genotype AA ≠ expression in genotype BB, what corresponds to $f_{\mathrm{A}} \neq f_{\mathrm{B}}$ in our model notation) but the A_{hyb} and B_{hyb} alleles in hybrid have equal expression (i.e. $f_{\mathrm{H}}^{\mathrm{A}} = f_{\mathrm{H}}^{\mathrm{B}}$). In such a case, the slope of the correlation between RAE (i.e. $\log_2(f_{\mathrm{H}}^{\mathrm{A}}/f_{\mathrm{H}}^{\mathrm{B}})$ in our model) and interparental expression divergence (i.e. $\log_2(f_{\mathrm{A}}/f_{\mathrm{B}})$ in our model) equals 0 (see red dots and line in Figure 1B).

2) By contrast, the gene is supposedly ***under cis- regulation*** when RAE matches parental expression divergence (i.e. $\log_2(f_{\mathrm{H}}^{\mathrm{A}}/f_{\mathrm{H}}^{\mathrm{B}}) \sim \log_2(f_{\mathrm{A}}/f_{\mathrm{B}})$ in our model notation). In such a case, the slope of the correlation between RAE and interparental expression divergence equals 1 (see blue dots and line in Figure 1B). Each allele thus tends to conserve the expression level from parental values and does not respond to orthologous regulatory elements.

Combined cis- and trans- effects are assumed in cases when RAE deviates from both patterns. Specifically:

3) The gene is categorized as ***under compensating cis- / trans- interaction*** if the cis- and trans-effects act in the opposite direction. In such a case the slope of the correlation between RAE and interparental expression divergence is greater than 1 (see yellow dots and line in Figure 1B).

4) Alternatively, ***enhancing cis- /trans- interaction*** is assumed if the cis- and trans-effects act in the same direction. In such a case the slope of the correlation between RAE and interparental expression divergence is greater than 0, but smaller than 1 (see green dots and line in the Figure 1B).

Supplementary Note 2

As relevant biological values for our model parameters we adopted the values of dissociation constants and total concentrations of TFs and of specific and non-specific binding sites from Veitia et al. [6]. In particular, assuming a nuclear volume $5 \cdot 10^{-13}$L, 150 binding specific binding sites of each type $\mathrm{S_A}$ and $\mathrm{S_B}$ in the AB-hybrid (i.e., 300 in total per nucleus), 10^6 non-specific binding sites, 557 TFs per nucleus, and the following values of the dissociation constants

$$K_{\mathrm{AA}} = K_{\mathrm{AB}} = K_{\mathrm{BB}} = K_{\mathrm{BA}} = 10^{-9}\ \mathrm{L/mol}, \qquad K_{\mathrm{AN}} = K_{\mathrm{BN}} = 10^{-5}\ \mathrm{L/mol},$$

we obtain

$$c(\mathrm{TF_A}) = c(\mathrm{TF_B}) = 9.25 \cdot 10^{-10}\ \mathrm{mol/L}, \qquad c(\mathrm{S_A}) = c(\mathrm{S_B}) = 4.48 \cdot 10^{-10}\ \mathrm{mol/L},$$

$$\text{and} \quad c(\mathrm{N}) = 3.32 \cdot 10^{-6}\ \mathrm{mol/L}.$$

At these values, the corresponding simulation resulted in equal fractional occupancy in both parental species with 50% binding sites occupied. Simultaneously, approximately 50% of TFs were bound (either to specific or non-specific binding sites) and 50% remained free.

Modelling scheme (1)

To achieve a reference scenario for our models, we simulated only one specific binding site per subgenome by reducing the respective values of $c(\mathrm{S_A})$ and $c(\mathrm{S_B})$ by a factor 150 (see modelling scheme (2) below). Also, to simulate the assumption that no TFs occur in a free state, we reduced all dissociation constants by a factor 10 (see modelling scheme (3) below), which resulted in a negligible proportion of free TF molecules. Therefore, we assumed a basic **reference set** of parameter values in scenario (1) so that:

$$K_{\mathrm{AA}} = K_{\mathrm{AB}} = K_{\mathrm{BB}} = K_{\mathrm{BA}} = 10^{-10};\ K_{\mathrm{AN}} = K_{\mathrm{BN}} = 10^{-6},$$

$$c(\mathrm{TF_A}) = c(\mathrm{TF_B}) = 3.75 \cdot 10^{-9}, \qquad c(\mathrm{S_A}) = c(\mathrm{S_B}) = 3.32 \cdot 10^{-12}, \qquad c(\mathrm{N}) = 7.332 \cdot 10^{-5}.$$

Such values resulted in the equal fractional occupancies of the specific binding sites of both parental species $f_{\mathrm{A}} = f_{\mathrm{B}} = 0.5$ and almost 99% of bound TFs.

In order to investigate how the expression divergence between parental species impacts on hybrids, we subsequently assumed that the ratio of fractional occupancies may vary between species A and B in the range 0.5 – 2 (i.e. $\log_2(f_{\mathrm{A}}/f_{\mathrm{B}})$ ranges between -1 and 1). For this purpose, we assumed the species modulate its gene expression by changing the concentrations of their TF molecules and hence we adjusted the parameter values of $c(\mathrm{TF_A})$ and $c(\mathrm{TF_B})$ accordingly. This allowed us to modify the ratio $f_{\mathrm{A}}/f_{\mathrm{B}}$ achieving values of 1/2, 2/3, 1, 3/2 and 2 as can be seen on x-axes of the Figures 3-7 (e.g., for the reference runs we set the couple $(c(\mathrm{TF_A}), c(\mathrm{TF_B}))$ to (1.86, 7.41), (2.55, 5.8), (3.75, 3.75), (5.8, 2.55) and (7.41, 1.86), with all values scaled by the factor 10^{-9}).

Each numerical calculation of fractional occupancies of specific binding sites of parental species A and B as well as of various types of hybrids has on its input two vectors, the vector $\mathbf{C}$ of total concentrations and the vector $\mathbf{K}$ of dissociation constants. Vector $\mathbf{C}$ contains 5 total concentrations: two values of TFs concentrations, two values of specific binding sites' concentrations originating from their parental species for each type of alleles and one value of the concentration of non-specific binding sites. These

concentrations are set in dependence on the type of parental alleles entering the nucleus in a given run (type of parents and/or hybrids) and on the specific choices of the modelling schemes (1)-(3), (5) and (6) described below. Vector $\mathbf{K}$ contains 6 dissociation constants, their values are set by the choice of parameter ρ in modelling scheme (4) and by the choice of adaptation in modelling scheme (6).

When adjusting the concentration vector $\mathbf{C}$ (see below for details), we assumed that the cell volume of the parental species A is $\boldsymbol{V_A} = 1$ and the cell volume of the parental species B is expressed as its multiplicities so that $\boldsymbol{V_B} = \boldsymbol{\nu_{BA}} \boldsymbol{V_A}$. We used ν_{BA} values 1, 1/2, and 2 to simulate the situation that both species have equal cell size or differ by a factor of two, see also modeling scheme (5) below. Another coefficient, $\boldsymbol{\omega_{BN}}$, was used to reflect how many times higher is the concentration of non-specific binding sites of parental species B compared to species A, see modelling scheme (6) below. Modelling of particular types of individuals, either parental species or diploid or polyploid hybrids, may then be achieved by simple setting of following values of each parameter. Vector $\mathbf{C}$ for both parental species is set as follows:

AA-parent: $\mathbf{C} = \big(c(\mathrm{TF_A}); c(\mathrm{TF_A}); c(\mathrm{S_A}); c(\mathrm{S_A}); c(\mathrm{N})\big)$,

BB-parent: $\mathbf{C} = (c(\mathrm{TF_B})/\nu_{\mathrm{BA}}; c(\mathrm{TF_B})/\nu_{\mathrm{BA}}; c(\mathrm{S_B})/\nu_{\mathrm{BA}}; c(\mathrm{S_B})/\nu_{\mathrm{BA}};\ \omega_{\mathrm{BN}}\, c(\mathrm{N})/\nu_{\mathrm{BA}})$.

For hybrids an additional parameter W_{H} appears reflecting our assumption that the cell volume of each hybrid type H is the average of the cell volumes of its parents: Volume of the cell of the **AB-hybrid** is $V_{\mathrm{AB}} = (V_{\mathrm{A}} + V_{\mathrm{B}})/2 = (V_{\mathrm{A}} + \nu_{\mathrm{BA}} V_{\mathrm{A}})/2 = (1 + \nu_{\mathrm{BA}}) V_{\mathrm{A}}/2$. Therefore, volume of the cell of the AB-hybrid compared to the volume V_{A} of the parental species A is $(V_{\mathrm{A}} + V_{\mathrm{B}})/2/V_{\mathrm{A}} = W_{\mathrm{AB}}/2$ with $W_{\mathrm{AB}} = 1 + \nu_{\mathrm{BA}}$, hence vector $\mathbf{C}$ has a form

$\mathbf{C} = (2\, c(\mathrm{TF_A})/W_{\mathrm{AB}}; 2\, c(\mathrm{TF_B})/W_{\mathrm{AB}}; 2\, c(\mathrm{S_A})/W_{\mathrm{AB}}; 2\, c(\mathrm{S_B})/W_{\mathrm{AB}};\ (1 + \omega_{\mathrm{BN}}) c(\mathrm{N})/W_{\mathrm{AB}})$.

Similarly, volume of the **AAB-hybrid** compared to the volume V_{A} is $(2V_{\mathrm{A}} + V_{\mathrm{B}})/2/V_{\mathrm{A}} = W_{\mathrm{AAB}}/2$ with $W_{\mathrm{AAB}} = 2 + \nu_{\mathrm{BA}}$, hence vector $\mathbf{C}$ has a form

$\mathbf{C} = (4\, c(\mathrm{TF_A})/W_{\mathrm{AAB}}; 2\, c(\mathrm{TF_B})/W_{\mathrm{AAB}}; 4\, c(\mathrm{S_A})/W_{\mathrm{AAB}}; 2\, c(\mathrm{S_B})/W_{\mathrm{AAB}};\ (2 + \omega_{\mathrm{BN}}) c(\mathrm{N})/W_{\mathrm{AAB}})$,

while volume of the **ABB-hybrid** compared to the volume V_{A} is $(V_{\mathrm{A}} + 2V_{\mathrm{B}})/2/V_{\mathrm{A}} = W_{\mathrm{ABB}}/2$ with $W_{\mathrm{ABB}} = 1 + 2\nu_{\mathrm{BA}}$ and the vector $\mathbf{C}$ has a form

$\mathbf{C} = (2\, c(\mathrm{TF_A})/W_{\mathrm{ABB}}; 4\, c(\mathrm{TF_B})/W_{\mathrm{ABB}}; 2\, c(\mathrm{S_A})/W_{\mathrm{ABB}}; 4\, c(\mathrm{S_B})/W_{\mathrm{ABB}};\ (1 + 2\omega_{\mathrm{BN}}) c(\mathrm{N})/W_{\mathrm{ABB}})$.

With these settings, we could then adjust the model to tackle various biological scenarios as described in section 3b of the main text:

Modelling scheme (2):

To investigate the effect of **single vs. multiple specific binding sites**, we compared the reference run as above, with a model where number of specific binding sites per diploid nucleus was assumed 300 as in Veitia et al. [6]. Therefore, we multiplied the concentrations of specific binding sites by 150, i.e. $c(S_{\mathrm{A}}) = c(S_{\mathrm{B}}) = 0.448 \cdot 10^{-9}$.

Modelling scheme (3):

In order to investigate how the **presence of free TFs** impacts on the model behavior, we compared the reference run (modelling scheme (1), where <1% of TFs were free) with a model, where we increased the values of all 6 dissociation constants 10 times. The latter setting resulted in a state of approximately 50% of TFs being free.

Modelling scheme (4):

In this setting, we investigated the **effects of limited cross-regulation**. To do so, the dissociation constants, i.e., the elements of the vector $\mathbf{K} = (K_{\mathrm{AA}}, K_{\mathrm{AB}}, K_{\mathrm{BB}}, K_{\mathrm{BA}}, K_{\mathrm{AN}}, K_{\mathrm{BN}})$ for each simulation run were set with the help of the coefficient ρ and (24)-(26) or (29)-(31) depending on the modelling scheme (6), see below for details.

We set $\rho = 1$ for the full cross-regulation, and $\rho = 10^4$ for no cross-regulation (where the affinity of TFs to heterospecific binding sites is the same as to any non-specific binding sites).

To investigate the effects of somewhat limited cross-regulation, we performed the runs also with $\rho = 2, 5, 10, 100$, of which results only for $\rho = 2$ are shown in figures 3-7.

Modelling scheme (5):

In this setting, we investigated the **effects of unequal cell volume sizes** between species and in hybrids. We therefore set $v_{\mathrm{BA}} = 1$ if both parental species have the same cell volume and we set $v_{\mathrm{BA}} = 1/2$ or $v_{\mathrm{BA}} = 2$ if parental species B have half or 2-times greater, respectively, cells comparing to species A. Hybrids' cell volumes were calculated as averaged from these values as described above.

Modelling scheme (6) (coefficient ρ being from modelling scheme (4))**:**

In this setting, we investigated the **effects unequal contents of non-specific binding sites** among parental species, introduced previously by Bottani et al. [7]. We set $\omega_{\mathrm{BN}} = 1$ or $\omega_{\mathrm{BN}} = 10$ in case parental species B having the same or 10-times greater amount of non-specific binding sites than species A, respectively. Concentrations of non-specific sites in hybrids were calculated from these values as described above.

In the case $\omega_{\mathrm{BN}} = 10$, we simultaneously investigated the effects of assumptions about how the **adaptation to increased euchromatic ratio occurs in the parental species B**. We set the value of the dissociation constant K_{AA} either to the reference value 10^{-10} or to the increased value 10^{-9} depending on the modelling scheme (3). Depending on the type of adaptation we then set the "cross" dissociation constants K_{AB}, K_{BA}, and also K_{BB} as follows.

In a **scenario 6.a** we assumed the species B has adapted to the increased concentration of its non-specific binding sites *via* **increasing the affinity of its promotor sites**, which allowed him to achieve the same fractional occupancy of specific binding sites as the species A. To simulate this scenario, we set the dissociation constants such that the strength of cross-regulation given by the coefficient ρ (see modelling scheme (4)) satisfies

$$\rho = \frac{K_{\mathrm{BA}}}{K_{\mathrm{AA}}} = \frac{K_{\mathrm{AB}}}{K_{\mathrm{BB}}}\,. \tag{22}$$

Let us emphasize that the *second letter* in the indices at K in (22) is the same in each ratio, we read this notation as "relative to the promoter positions", the affinity of specific binding sites for orthologous TFs is ρ-fold higher than for heterologous TFs, and this is true for both types of parental alleles in the nucleus. Let us note that the larger is the affinity, the smaller is the respective dissociation constant.

Simultaneously, the relation of bindings between parental species is controlled by the values of dissociation constants such that

$$\frac{K_{\mathrm{BB}}}{K_{\mathrm{AA}}} = \frac{K_{\mathrm{AB}}}{K_{\mathrm{BA}}} = \gamma\,, \tag{23}$$

where we set $\gamma = 1$ or $\gamma = 0.1011$ for all simulation runs where both parental species have the same concentration of non-specific binding sites or such concentrations is 10-times higher in parental species B, respectively, depending on the choice in modelling scheme (4).

In fact, for each particular simulation run we set the triplet K_{AA}, γ and ρ and the other dissociation constants controlling the affinity to either specific promoters are then given by

$$K_{\mathrm{BA}} = \rho\, K_{\mathrm{AA}},\ K_{\mathrm{BB}} = \gamma\, K_{\mathrm{AA}},\ K_{\mathrm{AB}} = \rho\, K_{\mathrm{BB}}. \tag{24)-(26}$$

In a **scenario 6.b** we assumed the adaptation of the species B to increased euchromatic ratio has been achieved via **modification of its TF molecule affinities.** This allowed the species B to achieve the same fractional occupancy of its specific binding sites as the species A. We thus assume the relations

$$\frac{K_{\mathrm{AB}}}{K_{\mathrm{AA}}} = \frac{K_{\mathrm{BA}}}{K_{\mathrm{BB}}} = \rho, \quad \frac{K_{\mathrm{BB}}}{K_{\mathrm{AA}}} = \frac{K_{\mathrm{BA}}}{K_{\mathrm{AB}}} = \gamma, \tag{27)-(28}$$

where $\gamma = 1$ or $\gamma = 0.31$ for all simulation runs where both parental species have the same concentration of non-specific binding sites or such concentrations is 10-times higher in parental species B, respectively. Let us emphasize that now the *first letter* in the indices at K in the first two ratios in (27) is the same since the cross-regulation is controlled from the point of view of TFs, TFs affinity for orthologous specific binding sites is ρ-fold higher than for heterologous ones.

After setting K_{AA}, γ and ρ, we now have

$$K_{\mathrm{AB}} = \rho\, K_{\mathrm{AA}},\ K_{\mathrm{BB}} = \gamma\, K_{\mathrm{AA}},\ K_{\mathrm{BA}} = \rho\, K_{\mathrm{BB}}\,. \tag{29)-(31}$$

Let us remark that the relations in (27)–(31) differ from those in (22) – (26).

The remaining constants K_{AN} and K_{BN} are set independently, both equally in the order of 10^4-times larger than K_{AA}.

Numerical procedure for solving the nonlinear system of equations (7)-(11)

The system of nonlinear equations (7)-(11) in the main text for a vector $\mathbf{F} = ([\mathrm{TF_A}], [\mathrm{TF_B}], [\mathrm{S_A}], [\mathrm{S_B}], [\mathrm{N}])$ of five unknown concentrations of free components of the system participating in the binding process was solved numerically with help of Matlab/octave function `fsolve()` with the initial vector $(-10^{-11}, 10^{-9}, 2 \cdot 10^{-10}, 2 \cdot 10^{-10}, 10^{-3})$ and a preset tolerance 10^{-16}. The negative first component was chosen artificially since with this negative value the script performed better convergence to the equilibrium and robust behavior for various types of parameter values. The output of the `fsolve()` function was used as an initial vector for the second run of the `fsolve()` function to obtain a more precise solution of the system as one can see in the source octave script file in Supplementary Code 1 file.

Using the vector $\mathbf{F}$ of calculated concentrations, the desired fractional occupancies were calculated. For parental species A, f_{A} was calculated using formula (18) in the main text, for parental species B, f_{B} using (19). For diploid hybrid AB and triploid hybrids AAB and ABB, the fractional occupancies of A-type specific sites were calculated using (20), for B-type specific sites using (21).

Script (source code) processed by octave software

We present in Supplementary Code 1 file a source code used to obtain a numerical solution of the model system (7)-(11) for a particular choice of parameter values. It can be processed by a commercial software Matlab or by a free clone Octave. Under Linux we simply run

```
octave -f solveOccup.m
```

in the command prompt.

References

1. Yoo, M.-J., Szadkowski, E. & Wendel, J. F. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171 (2013).
2. Mattioli, K. *et al.* Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biology* **21**, 210 (2020).
3. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).

4. 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).
5. Shi, X. *et al.* Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat Commun* **3**, 950 (2012).
6. Veitia, R. A., Bottani, S. & Birchler, J. A. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet.* **29**, 385–393 (2013).
7. Bottani, S., Zabet, N. R., Wendel, J. F. & Veitia, R. A. Gene Expression Dominance in Allopolyploids: Hypotheses and Models. *Trends in Plant Science* **23**, 393–402 (2018).