Electronic Supplementary Material (ESI) for Integrative Biology.

Supplementary text 1

Topological stability analysis of gene regulatory networks:

Definitions:

A gene regulatory network can be seen as a directed graph composed by a set of vertices V and a set of edges E. Each vertex $v_i \in V$ corresponds to a transcription factor. Any edge $e_{ij} \in E$ corresponds to an ordered pair of transcription factors $e_{ij} = \left(v_i, v_j\right)$ so that the transcription factor v_i regulates the expression of the gene coding the transcription factor v_j. The number of transcription factors in the network is known as the cardinal of the graph and can be noted as $|V|$.

The time evolution of the number of copies of a particular transcription factor in a cell can be described by the following differential equation.

$$\frac{dn_i}{dt} = r_i\left(\left\{n_j / e_{ji} \in E\right\}\right) - \delta_i n_i \tag{1}$$

The previous equation indicates that the rate of transcription (and indirectly its rate of translation) of a transcription factor, depends on the number of copies of all the transcription factors that are involved in its regulation. If a transcription factor v_j is not involved in the regulation of the expression of the transcription factor v_i, $e_{ji} \notin E$ and r_i is not a function of n_j. The parameter δ_i is the specific degradation rate of the transcription factor v_i.

The system of equations governing the regulatory network can be written using the following vector notation:

$$\frac{d\vec{n}}{dt} = \vec{r}\left(\vec{n}\right) - \Delta\vec{n} \tag{2}$$

The matrix Δ is a diagonal matrix that contains the specific degradation rates δ_i.

A fixed point of the regulatory network corresponds to a vector $\vec{n}_0$ for which all the time derivatives are equal to zero. In the neighborhood of a fixed point the time evolution of the system can be described by the following equation:

$$\vec{n} - \vec{n}_0 = \sum_{j=1}^{|V|} e^{\lambda_j t}\vec{u}_j \tag{3}$$

The coefficients λ_j and the vectors $\vec{u}_j$ are respectively the eigenvalues and the eigenvectors of the differential matrix of the system, which is expressed as follows:

$$D = \frac{\partial\left(\vec{r}\right)}{\partial(\vec{n})} - \Delta \tag{4}$$

The differential matrix of the system is related to the adjacency matrix of the associated graph A in the way that if D_{ij}=0, A_{ij}=0.

We define the support of an eigenvector $\vec{u}_i$ as the set of transcription factors $S_i = \{v_k\}$ for which $\vec{u}_i(k) \neq 0$. The symbol $\vec{u}_i(k)$ represents the kth element of $\vec{u}_i$.

Determination of the supports of the eigenvectors:

Remark 1: From linear algebra it is known that the number of eigenvectors is equal to the number of variables in the system, which in our case is $|V|$. Therefore we can establish a one to one map between transcription factors and eigenvectors.

Remark 2: If there is a path from v_i to v_j and a path from v_j to v_k then there is also a path from v_i to v_k.

Lemma 1: *For any vertex v_i we define the set $P_i = \{v_k\} \subset V$ as the set of vertices for which there exists a path from v_i to v_k. If there is no path from v_i to v_j then D_{jk}=0 $\forall v_k \in P_i$*

Proof: If $D_{jk} \neq 0$ there is an edge directed from v_k to v_j. As $v_k \in P_i$ there is a path from v_i to v_k and as it is stated in remark 2 there is also a path from v_i to v_j, which contradicts the proposition of the lemma.

Remark 3: The set P_i can contain v_i or not. If it contains v_i, this vertex is located in a circuit and all the other vertices in the circuit belong also to P_i.

Remark 4: The subset of V, $R_i = \{v_j \neq v_i / v_j \notin P_i\}$ is the complement of $Q_i = \{v_i\} \cup P_i$, $R_i = Q_i^c$

Theorem 1:. *For each vertex v_i, there are $|Q_i|$ eigenvectors of D with a support $S_i \subseteq Q_i$.*

$|Q_i|$ is the cardinal of the set Q_i.

Proof: An eigenvector of D is defined by the following equation:

$$D\vec{u}_i = \lambda_i \vec{u}_i \tag{5}$$

Component by component we can write:

$$\sum_k D_{jk}\vec{u}_i(k) = \lambda_i \vec{u}_i(j) \tag{6}$$

If $v_j \in R_i$, and $v_k \in Q_i$, according to lemma 1 D_{jk}=0. Therefore there are solutions of equation (6) for which $\vec{u}_i(j) = 0 \forall v_j \in R_i$, which implies that there is at least an eigenvector with a support that does not intersect the set R_i, $R_i \cap S_i = 0$. If the support S_i does not intersect R_i it can be the empty set or it can be contained in the complement of R_i, which is Q_i.

Now it is necessary to show that there are solutions with non-zero values in at least one of the other components of $\vec{u}_i$ and therefore its support is not the empty set. The sub-graph that contains only the vertices $v_k \in Q_i$ is a subsystem with a differential matrix *D'* that contains only the rows and columns corresponding to the vertices $v_k \in Q_i$. This matrix has as many eigenvectors as the number of nodes, $v_k \in Q_i$, each eigenvector contains at least one non-zero element, therefore there are $|Q_i|$ solutions for which $\exists v_k \in Q_i / \vec{u}_i(k) \neq 0$, which implies that there are also $|Q_i|$ eigenvectors of *D* with a support. $S_i \subseteq Q_i$

Theorem 2: *If $v_i \notin P_i$ there exists a single eigenvector $\vec{u}_i$ with a support $S_i \subseteq Q_i$, that also contains v_i. $v_i \in S_i$. This eigenvector has an eigenvalue equal to $-\delta_i$.*

Proof: according to theorem 1 there are $|Q_i|$ eigenvectors of *D* with a support. $S_i \subseteq Q_i$. As $v_i \notin P_i$ then $|Q_i| = 1 + |P_i|$. If $v_i \notin P_i$ there is no path starting in v_i that goes back to itself, therefore according to lemma 1 D_{ik}=0 $\forall v_k \in P_i$. If all the components $\vec{u}_i(j) = 0 \forall v_j \in (R_i \cup \{v_i\})$ then the system is reduced to a subsystem composed by the elements of P_i and its differential matrix has $|P_i|$ rows and columns and also $|P_i|$ eigenvectors. The support of these is eigenvectors is contained in P_i. Ifg there are $|P_i|$ eigenvectors contained in P_i and $|P_i| + 1$ eigenvectors with supports contained in Q_i, given the fact that by deffinition $P_i \subset Q_i$, there must be exactly one single eigenvector with a support $S_i \subset Q_i$ and $S_i \not\subset P_i$, which implies that $v_i \in S_i$ or equivalently $\vec{u}_i(i) \neq 0$.

Once we know that this vector exists and that is unique we can determine its associated eigenvalue.

If $v_i \notin P_i$ there is no path starting in v_i that goes back to itself, therefore according to lemma 1 D_{ik}=0 $\forall v_k \in P_i$. If $S_i \subseteq Q_i \Rightarrow \vec{u}_i(j) = 0 \forall v_j \in R_i$. This means that equation (6) can be simplified to the form: $-\delta_i \vec{u}_i(i) = \lambda_i \vec{u}_i(i)$. As $\vec{u}_i(i) \neq 0$ then $\lambda_i = -\delta_i$. This proves completely the theorem.

Remark 5: If $v_k \in P_i$ then $P_k \subseteq P_i$.

Lemma 2: *If v_i and v_j are in the same cycle, the sets P_i and P_j are identical.*

Proof: If both vertices are in the same cycle, there is a path from v_i to v_j and also a path from v_i to v_j, therefore $v_j \in P_i$ and $v_i \in P_j$. Taking account of remark 5 $P_j \subseteq P_i$ and $P_i \subseteq P_j$, therefore P_i=P_j.

Remark 6: Given lemma 2 we can see that the relation “being together in the same cycle” divides the set of vertices in equivalence classes. These equivalence classes are subsets of *V* that we will call “cyclic sets” and note as C_r and the intersection between two different cyclic sets is empty $C_r \subseteq V$, $C_r \cap C_s = 0 \forall r \neq s$. The vertices that are not in any cycle can be seen as cyclic sets of only one element and they satisfy also the two previously mentioned properties of the cyclic sets.

Lemma 3: *Each vertex in V belongs to one and only one cyclic set.*

Proof: Let's assume that a vertex v_i belongs to two different cyclic sets C_r and C_s. Then there is a path from v_i to any element of C_r and there is a path from any element of C_s to v_i. This means that there is a path from any element of C_s to any element of C_r. In an analogous way we can show that there is also a path from any element of C_r to any element of C_s, therefore the cyclic sets C_r and C_s are identical, which contradicts our initial proposition.

Remark 7: To each cyclic set C_r we can associate a set P_r, which is identical to $P_i \forall v_i \in C_r$. By definition $C_r \subseteq P_r$.

Theorem 3: *For any cyclic set C_r there are $|C_r|$ eigenvectors whose supports S_i are contained in P_r and contain elements of C_r.* $S_i \subseteq P_r$, $S_i \cap C_r \neq 0$.

Proof: From theorem 1 we know that there are $|P_r|$ eigenvectors with supports contained in P_r. From lemma 3 we know that each element of P_r belongs to one and only one cyclic set. Now we are going to proceed by induction.

Let's assume that C_r=P_r, in this case theorem 1 states that there are $|P_r| = |C_r|$ eigenvectors with supports contained in P_r and therefore also in C_r. as these supports are contained in C_r they must contain elements of C_r therefore theorem 3 is valid for this case.

Now let's assume that P_r contains C_r and other cyclic sets C_g that satisfy the condition C_g=P_g $\forall C_g \subset P_r / C_g \neq C_r$. For each C_g we have already proven than there are $|C_g|$ eigenvectors with supports contained in C_g. As the intersection between different cyclic sets is empty the intersection between the support of these eigenvectors and C_r is empty. This means that the $|C_g|$ eigenvectors with supports contained in C_g have supports contained in P_r and do not intersect C_r. There are in total $\sum_g |C_g|$ eigenvectors with these characteristics. As we know that there are $|P_r|$ eigenvectors with supports contained in P_r there must be $|P_r| - \sum_g |C_g| = |C_r|$ eigenvectors contained in P_r that have a non-empty intersection with C_r. This proves theorem 3 for this case.

The theorem can be proven for any case. by repeating the same reasoning.

Conclusions:

By computing the sets P_i corresponding to each vertex in the regulatory network we can obtain sets that contain the support of each eigenvector of the differential matrix $S_i \subseteq Q_i$. For some very particular values of the partial derivatives contained in the differential matrix, there might be some cases in which some of the vertices in Q_i contain zero values ($S_i \subset Q_i$) but in general it can be assumed that S_i=Q_i.

If a vertex is not contained in a cycle (there is no pathway coming back to itself) there is an eigenvector with support $S_i \subseteq Q_i$ (very likely S_i=Q_i) which has a negative eigenvalue and corresponds to a stable mode of the system. If there are no cycles in the network all the eigenvalues are negative and all the steady states are stable.

If there are cycles in the network, for each cyclic set C_r there are $\left|C_r\right|$ eigenvectors with supports that contain elements of C_r (except for some very particular values of the elements of the differential matrix they will contain all the elements of C_r). The eigenvalues associated to these vectors can have positive and negative real parts and also imaginary components (depending on the values of the elements of the differential matrix). This means that these eigenvectors involving the vertices in C_r can become unstable due to changes in the parameters of the system and give rise to reordering phenomena in which due to some slight changes in parameters (mutations or environmental changes), the regulatory network can move from a fixed point to that becomes unstable to a completely different one. These reordering phenomena (known as catastrophes in catastrophe theory) will be typically associated to strong changes in the expression level of the transcription factors belonging to the set P_r. That's why using the sets P_r for gene-set expression analysis might be a good alternative to using gene ontologies etc. This is because using these sets might give us better insights into the nature of the possible reordering phenomena.

Supplementary text 2

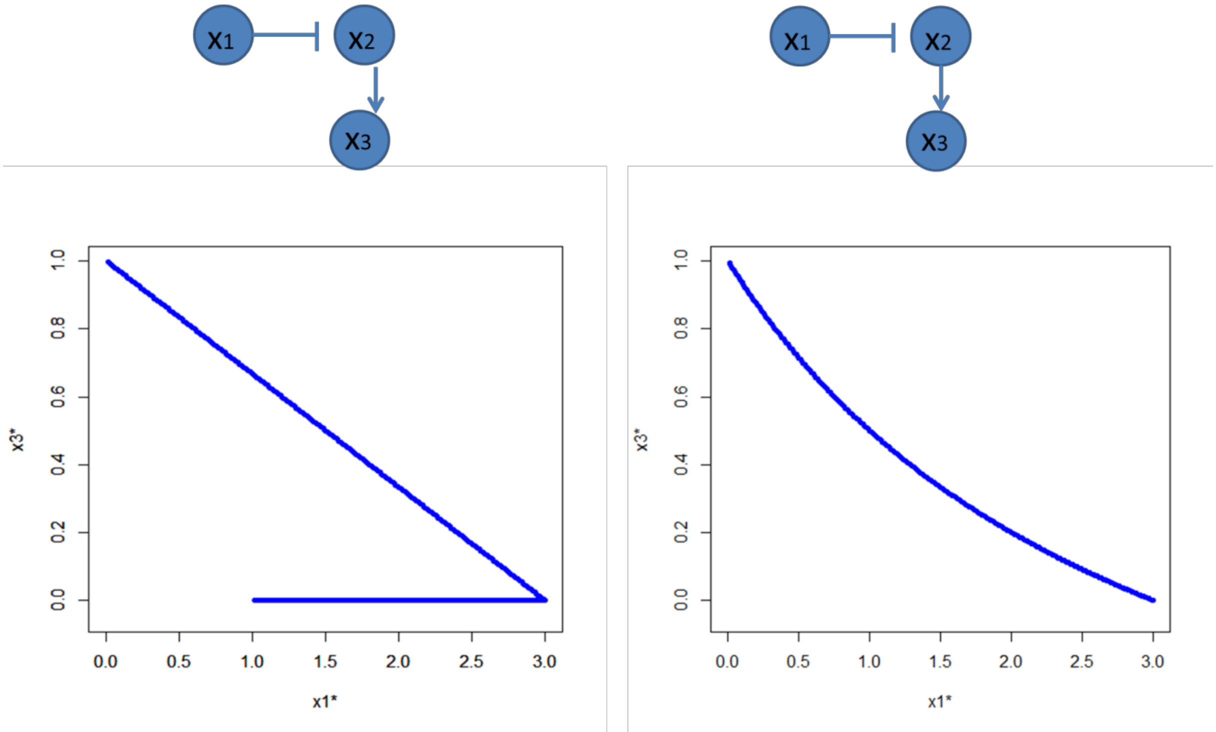

Figure S1. The figure illustrates how the presence of a regulatory loop induces a qualitative different behavior characterized by two alternative steady states, one of which becomes unstable when the parameter x_1^* falls below a certain threshold. This would cause the system to transit from the lower steady state to the upper one giving rise to a rearrangement similar to a non-equilibrium phase transition. The system will remain in the upper steady state even if the parameter that caused the transition comes back to the original value, giving rise to a hysteresis cycle. These irreversible transitions are characteristic of living systems. Some examples are the transitions between different phases of the cell cycle, the commitment of a cell to apoptosis or the differentiation of stem cells.

Example of topological stability:

We consider a system of three transcription factors. Two of them activate the expression of each other and the third one represses the expression of one of them. Let's consider that the activity of the repressor x_1 is governed by some external input and can be seen as a variable parameter of the system.

The associated system of differential equation takes the form:

$$\frac{dx_2}{dt} = k_2^{\max} \frac{x_3}{x_3 + K_3 + x_1 / K_I} - \delta_2 x_2 \qquad [1]$$

$$\frac{dx_3}{dt} = k_3^{\max} \frac{x_2}{x_2 + K_2} - \delta_3 x_3 \qquad [2]$$

The two eigenvalues of this system can be calculated using the following equation:

$$\lambda = \frac{-\left(\delta_2 + \delta_3\right) \pm \sqrt{\left(\delta_2 + \delta_3\right) - \delta_2\delta_3 + \left(\partial x_2 / \partial x_3\right)\left(\partial x_3 / \partial x_2\right)}}{2} \qquad [3]$$

Both eigenvalues will be negative and the system will be stable if:

$$\delta_2\delta_3 > \left(\partial x_2 / \partial x_3\right)\left(\partial x_3 / \partial x_2\right) \qquad [4]$$

In our case this condition is equivalent to:

$$1 > \frac{\kappa_2^* \kappa_3^*}{\left(x_3^* + x_1^* + 1\right)^2 \left(x_2^* + 1\right)^2} \qquad [5]$$

We have written the previous equation using the following adimensional quantities:

$$x_1^* = \frac{x_1}{K_3 K_I} \qquad [6]$$

$$x_2^* = \frac{x_2}{K_2} \qquad [7]$$

$$x_3^* = \frac{x_3}{K_3} \qquad [8]$$

$$\kappa_2^* = \frac{k_2^{\max}}{\delta_2 K_3} \qquad [9]$$

$$\kappa_3^* = \frac{k_3^{\max}}{\delta_3 K_2} \qquad [10]$$

On the other hand the system has two possible steady states. The first one corresponds to $x_2^*=0$ and $x_3^*=0$ and the second one to:

$$x_2^* = \frac{\kappa_2^*\kappa_3^* - \left(1 + x_1^*\right)}{\kappa_3^* + x_1^* + 1} \qquad [11]$$

$$x_3^* = \frac{\kappa_3^* - 1}{\kappa_2^* + 1} - \frac{x_1^*}{\kappa_2^* + 1} \qquad [12]$$

If we give to κ_3^* and κ_2^* values of 2, the first solution is stable for values of x_1^* higher than 1. The second steady state is always stable because equation [5] takes a constant value of 0.25.

If the system is in the first steady state and as a result of an environmental stimulus x_1^* drops below the critical value of 1, the system will reorganize itself by moving to the second steady state and it remains there even if x_1^* recovers its previous value. Similar sudden and irreversible rearrangements are observed in the progress of eukaryotic cells through the cell cycle or in cell differentiation.

If the third transcription factor does not regulate the second one, the system would only have a single steady state for any parameterization and x_1^* and x_3^* are related by a continuous function.

In this case the only steady state is stable and defined by the following equations:

$$x_2^* = \frac{\kappa_2^*}{x_1^* + 1} \qquad [13]$$

$$x_3^* = \frac{\kappa_2^*\kappa_3^*}{\kappa_2^* + x_1^* + 1} \qquad [14]$$

These characteristics are a result of the system’s topology and are independent of the particular parameterization.

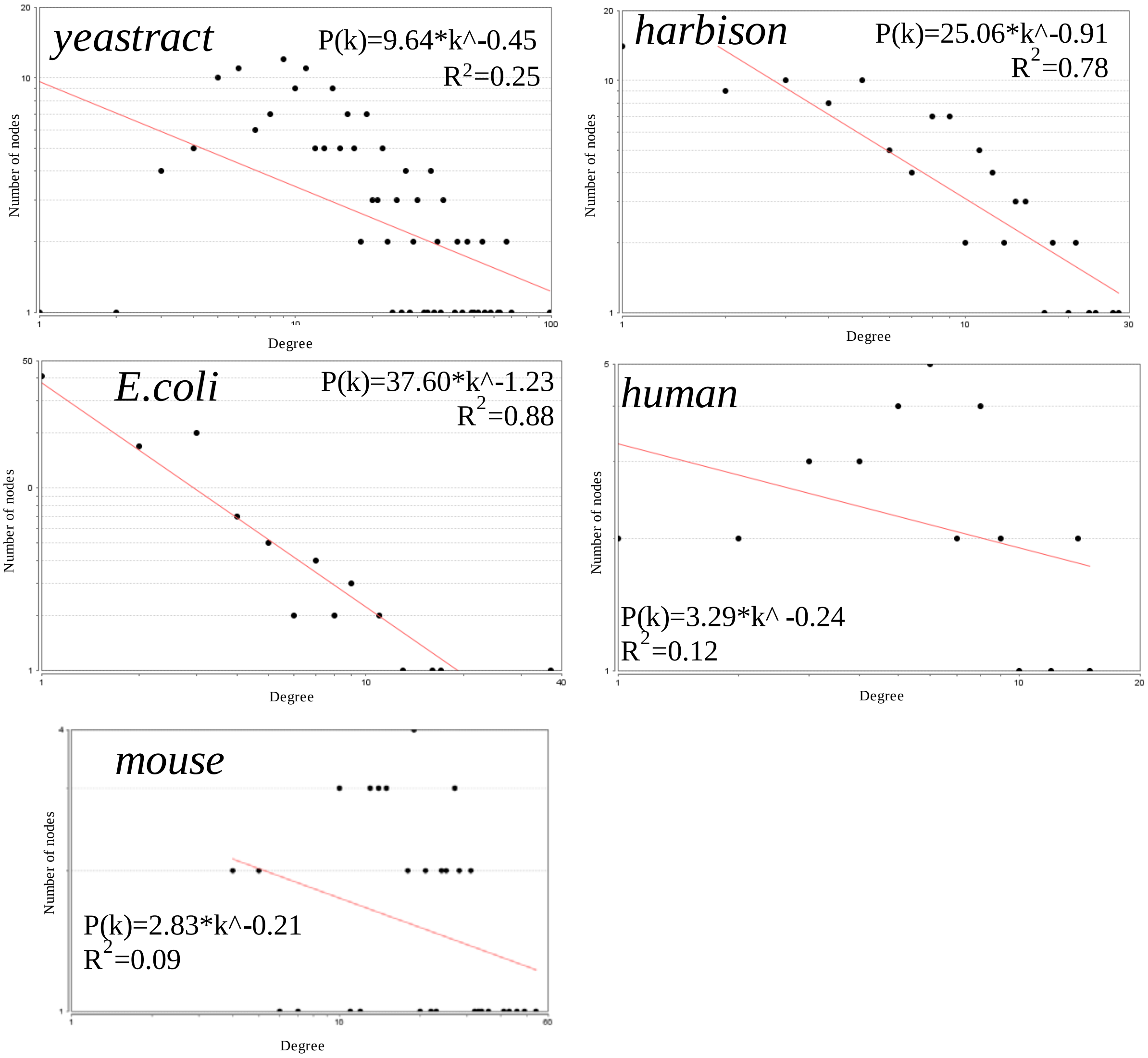

Figure S2 - Degree distribution plots on log-log scale for different networks. The red line and the formula represent the fitted power law distribution and the R2-value represents how good the fit is.

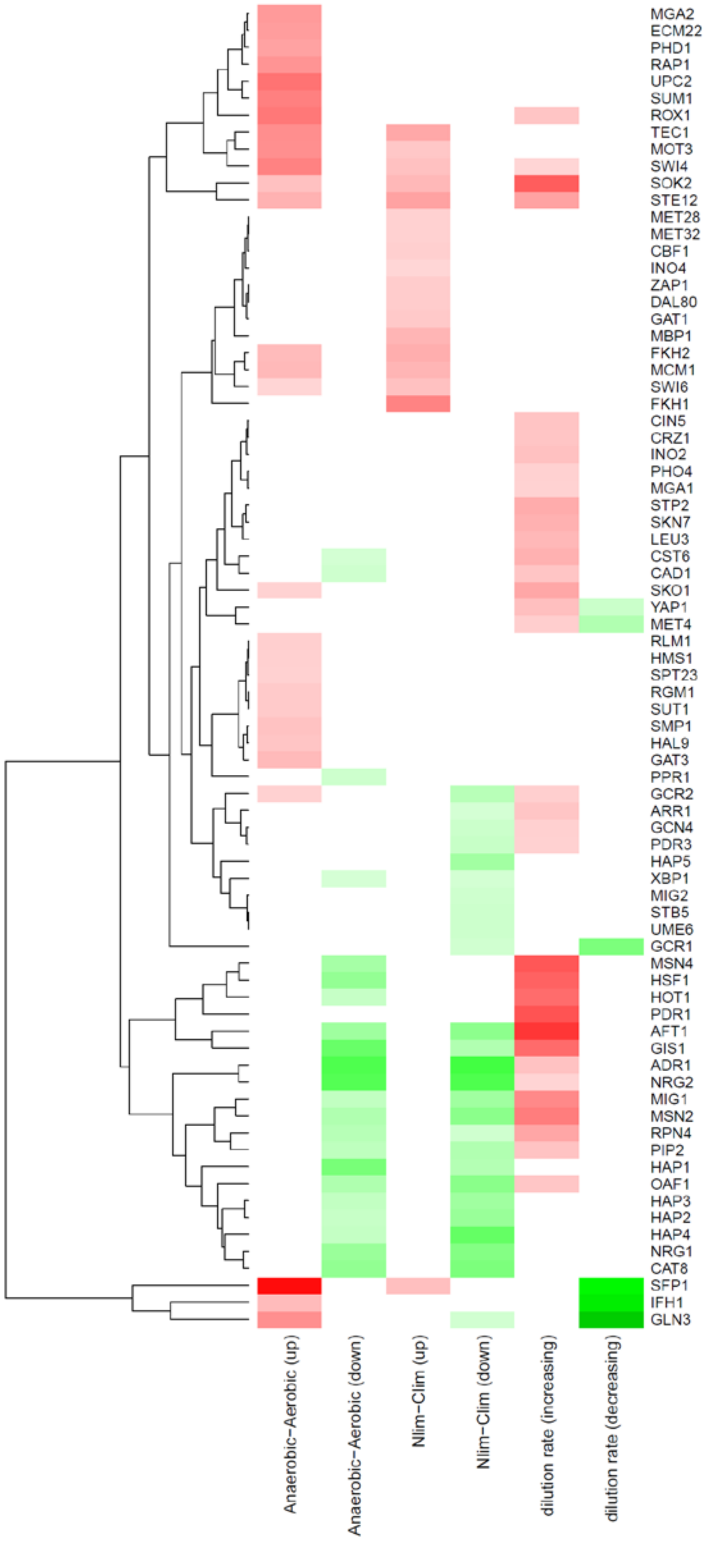

Figure S3 - TFs with up- or down-regulated target genes as a function of the environment. Red color represents that most of the target genes are up-regulated in one condition compared to the other, and green means down-regulation of a majority of the target genes.