PNAS

The duplication of genomes and genetic networks and its potential for evolutionary adaptation and survival during environmental turmoil

Mehrshad Ebadi[a,b], Quinten Bafort[a,b], Eshchar Mizrachi[c], Pieter Audenaert[d], Pieter Simoens[d], Marc Van Montagu[a,b,*], Dries Bonte[e,*], and Yves Van de Peer[a,b,c,f,*]

[a]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium; [b]Center for Plant Systems Biology, VIB, 9052 Gent, Belgium; [c]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa;[d]Department of Information Technology–IDLab, Ghent University-IMEC, 9052 Gent, Belgium; [e]Department of Biology, Terrestrial Ecology Unit, Ghent University, 9000 Ghent, Belgium; [f]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, China

*__Email:__ marc.vanmontagu@ugent.be (M.V.M.), dries.bonte@ugent.be (D.B.), and yves.vandepeer@psb-vib.ugent.be (Y.V.d.P)

Author Contributions: All authors conceived and designed the research; M.E. performed the experiments; M.E., Q.B., D.B., and Y.V.d.P. designed the experiments and analyzed the data; M.E., D.B., and Y.V.d.P. wrote the article.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences - Evolutionary Biology

Keywords: Whole Genome Duplication – Polyploidy – Autopolyploidy - Gene Regulatory Networks – Adaptation – Survival – Environmental Turmoil – Cataclysmic Events

Abstract

The importance of whole genome duplication (WGD) for evolution is controversial. Whereas some view WGD mainly as detrimental and an evolutionary dead end, there is growing evidence that polyploidization can help overcome environmental change, stressful conditions, or periods of extinction. However, despite much research, the mechanistic underpinnings of why and how polyploids might be able to outcompete or outlive non-polyploids at times of environmental upheaval remain elusive, especially for autopolyploids, in which heterosis effects are limited. On the longer term, WGD might increase both mutational and environmental robustness due to redundancy and increased genetic variation, but on the short – or even immediate – term, selective advantages of WGDs are harder to explain. Here, by duplicating artificially generated Gene Regulatory Networks (GRNs), we show that duplicated GRNs – and thus duplicated genomes – show higher signal output variation than non-duplicated GRNs. This increased variation leads to niche expansion and can provide polyploid populations with substantial advantages to survive environmental turmoil. In contrast, under stable environments, GRNs might be maladaptive to changes, a phenomenon that is exacerbated in duplicated GRNs. We believe that these results provide new insights into how genome duplication and (auto)polyploidy might help organisms to adapt quickly to novel conditions and to survive ecological uproar or even cataclysmic events.

Significance statement

Polyploid organisms abound, but long-term polyploid establishment is much rarer and likely not random. Hence, polyploidy is considered either an evolutionary dead end or a force that can help organisms survive environmental changes and stress. How and why polyploids, especially autopolyploids, might outcompete nonpolyploids during times of environmental upheaval is unclear. On a longer timescale, whole genome duplications may increase genetic robustness and variation, but their benefits on the short-term are harder to explain. We show that duplicating genomes and their encoded gene regulatory networks increase signal output variation, leading to niche expansion and increased potential for surviving environmental turmoil. These findings highlight how polyploidy might help organisms adapt to changing conditions and survive disruption but might be maladaptive under stable conditions.

Introduction

Whole-genome duplication (WGD) leading to polyploidy is a common phenomenon that has been studied for over 100 years, especially in flowering plants (1). Because of the well-known detrimental effects arising from genome doubling, most WGD events are not successful. Genomic instability, mitotic and meiotic abnormalities, and minority cytotype exclusion are all expected to quickly remove new polyploids from the population (2-4). Nevertheless, there are numerous polyploid organisms around us. Furthermore, even those organisms that are currently considered 'functional' diploids usually bear signatures of a polyploid ancestry (5, 6). Several of these ancestral polyploidy events can be traced back to the origin and diversification of major phylogenetic lineages, including vertebrates, fishes, and flowering plants; and within flowering plants, core eudicots, monocots, orchids, grasses, composites, and legumes (6-8).

This phylogenetic signal of polyploidy success suggests an important role for WGD in promoting phenotypic diversity, with a subsequent facilitating role in speciation (9-11). Speciation typically occurs under restricted conditions where certain genotypes can exploit novel ecological opportunities under the presence of mating barriers with others (12). More importantly, polyploidisation is often associated with the expression of new, often exaggerated, phenotypes that have the potential to promote niche expansion and a subsequent radiation in novel environments. Doubling the amount of DNA does for instance necessitate larger cell nuclei and cell size and has already major consequences on organismal developmental and physiological responses (13, 14). Size independent phenotypic changes have been documented on stress physiology and other traits that provide advantages under extreme environments (15, 16). It has been suggested that such potential niche expansion advantages in novel environments might be responsible for phylogenetic records showing a rise of polyploids at certain epoch boundaries, such as for instance the K-Pg boundary, a geological period characterized by major episodes of global climatic change and mass extinction (6, 17-22), or around recent glaciation maxima (23). Studies in yeast have shown that polyploidy can accelerate evolutionary adaptation to challenging environments, because WGD induced regulatory redundancy followed by divergence, allowing a wider range of phenotypic responses to environmental stresses (24, 25).

The increasing numbers of genes that diversify in function due to a relaxed functional constraint on one of both copies (i.e., sub- or neofunctionalization (26)), is likely not the sole explanation of WGD's evolutionary success under stress. Doubling of gene regulatory networks may equally increase the frequency of beneficial mutations (24), and therefore

enlarge the genetic and phenotypic variation for selection to act on. In this respect, WGD can be seen as a complex super mutation of/within the genome. The increased genetic variation and the buffering effect of their duplicated genes has led to an increased recognition of the adaptive potential of polyploidy (6, 27, 28).

Recent work by some of us based on a computational framework aimed at mimicking biological evolution (29, 30) suggested that so-called digital organisms (DOs) with an unduplicated genome performed better - as in, adapted faster - than DOs with a duplicated genome in stable environments, while the opposite was true for unstable environments (31). Somewhat similar observations were made with populations of so-called 'virtual cells' (32, 33). These insights were generated by the implementation of WGD as a series of random mutations of large – adaptive or maladaptive – effect. Gene regulatory networks (GRNs) shape the mechanistic pathway between genotype and phenotype. We build on the observation that duplicated GRNs seem to have magnified impact (31) and hypothesize that the phenotypic variance generated by such networks exceeds the one of the ancestral simpler networks. The eventual propagation of information through (artificial) networks, and thus the eventual distribution of output signals is uncertain. Signal propagation across a network can be considered as a sum of different node values. The eventual variance of the distribution of output signals in such systems with a double number of nodes will then be the sum of the variances of the distribution of output signals across all nodes and their (doubled) covariance. Hence, increasing the number of nodes, and given covariances not being strongly negative, variance of the distribution of output signals of the population of doubled networks should always increase. To what extent the duplicated structure of the network results in a different signal propagation relative to the non-duplicated version, or to networks with the same number of nodes but with random structure, remains understudied. Few studies thus far considered the duplication of entire GRNs, both *in silico* or *in vivo* and previous research on the effect of network duplication focused almost exclusively on the rewiring of the network after (gene and genome) duplication or on processes buffering the 'immediate' effects of duplication (34-42).

The rewiring of networks, (re)diploidization and fractionation (gene loss and genomic rearrangements) have important consequences for adaptation at the longer term as they generate new functions and phenotypes. Immediate consequences of WGD on evolvability are also expected from the doubling of both genes and their connections within the gene regulatory network. Because this obvious route for polyploid evolutionary success has not been explored, we use extensive simulations to test how the duplication of artificial ancestral (further also referred to as 'simple') gene regulatory networks (aGRNs) of different sizes and shapes impacts

the standing phenotypic variation for selection to act on. We generated aGRNs that represent scale-free genetic networks with general output functions that translate the aGRN to a gene product that is considered as the relevant phenotypic trait of interest under environmental change. We demonstrate that WGD increases phenotypic variation more than can be expected from doubling the number of genes alone. Importantly, this increased variation results from a non-random expansion of the phenotypic space by the proliferation of trait values along the same direction as the ancestral state. WGD therefore 'magnifies' or 'exaggerates' the phenotypic profile of the simple networks. Finally, by explicitly linking phenotype to fitness, we show these amplified phenotypes to be only adaptive during sudden environmental changes or periods of rapid extinction. WGD does impose a direct rescuing mechanism by enlarging the phenotypic space for selection to act on during episodes of strong environmental turmoil.

Results

Previous simulation studies showed that digital organisms (DOs) with one genome copy generally adapted faster than DOs with a duplicated genome in relatively stable environments, but not in unstable environments (31). Furthermore, if DOs with duplicated genomes did adapt to stable(r) environments, they did so with a restricted number of mutations, compared to DOs with one genome copy. By contrast, if DOs with single copy genomes adapted to more drastically changed environments, they needed more mutations to adapt than the DOs with duplicated genomes. From this observation, i.e., fewer mutations being allowed and fewer mutations with a higher impact in duplicated genomes (and their encoded GRNs) (31), we assumed that changes in duplicated GRNs, either through mutations or sensed input cues changing node 'values', have an enhanced impact. Therefore, here, we tested different ways to evaluate the dynamics of signals sent through the networks and how they translate in output generated by the nonduplicated and duplicated networks. Different signals sent through the network mimic different environmental cues, such as, for instance, differences in temperature, where greater differences in values represent greater environmental turmoil (see Methods).

The increased phenotypic variance of network duplication. Examples of 'simple' and 'duplicated' artificial Gene Regulatory Networks (aGRNs) can be found in Figures 1 and 2 (see Methods). For information on how these are constructed and used in the current study, we refer to Methods and *SI Appendix*.

Figure 3*A* shows a Phenotypic Trajectory Analysis (see Methods) of a population of simple (10 nodes) and duplicated networks in a single environment. It provides an example of the increase in the phenotypic variation of the duplicated population compared to the simple (ancestral) ones. Although the phenotypic effect of WGD depends on the network's topology (i.e., the genotype), the average phenotypic value of the duplicated networks is on average more extreme than that of the simple networks. It is noteworthy to mention that the phenotypic variance of duplicated networks is on average significantly larger than that of simple networks with the same number of nodes, but not having the typical 'duplicated' network structure (Fig. 3*B*, comparing, for instance, the mean variances of the output nodes from the 20-node duplicated networks with those of simple 40-node networks (red arrows) (see also *SI Appendix*, Fig. S3.2). Similarly, the average phenotypic variance σ as measured by multiplying variance of both (mean) output node values of populations (1000) of simple networks of 10 nodes and their duplicated networks is more than four times higher for duplicated (σ=0.176) compared to single networks (σ=0.041; Fig. 3*C*). For each pair of simulated simple-duplicated networks, the phenotypic vector length increases by about 30% (average vector length for single networks: 0.77±0.31; for duplicated networks: 1.00±0.30). Furthermore, this increase in phenotypic value (trait) is in the same direction as the phenotype of the single network. The relative angle between both vector angles is 0°±5° (Fig. 3*D*). For instance, if we consider gene expression as a trait, when gene expression is at a certain 'high level' in a simple network, gene expression will generally be further increased in the duplicated network. The same is true for repression of gene expression: in the duplicated network, gene repression will be higher/stronger. Of course, there are exceptions to the rule, indicated by vectors that point in contradictory directions, such as vectors in the lower left quadrant that point upwards rather than downwards (Fig. 3*A*). This pattern holds true for networks of all sizes and initialization conditions (see *SI Appendix* S2 and S3), but the dispersion of the relative angles slightly increases in larger networks. Genome doubling thus affects the phenotypic trait in the same 'direction' as in the nonduplicated 'ancestral' network, in a multiplicative manner, but the directionality, and thus predictability of the trait change decreases when networks increase in size. The duplication of the particular structure of a genetic network, rather than the sole increase in nodes, therefore underlies the observed pattern of phenotypic (or niche) expansion. Thus, WGD seems to systematically 'exaggerate' the obtained phenotypic value of the single network rather than driving it into a random direction.

Selection along environmental gradients. The relative fitness between the simple and duplicated networks, as for instance expressed in differential survival or reproduction, is higher for phenotypes generated by simple networks in the reference environment, and when deviations between the new and reference environment are small. In contrast, when environmental change is large(r), the fitness of the duplicated networks exceeds that of the simple networks (Fig. 4, *A* and *C*). By comparing differential fitness between single and duplicated networks, $(\boldsymbol{w}_{duplicated} - \boldsymbol{w}_{single})/\boldsymbol{w}_{single}$, it becomes obvious that even small mismatches between the environment and phenotype impose a strong selection against the duplicated genomes (Fig. 4, *B* and *D*). With increasing environmental change, the fitness of the duplicated networks exceeds that of the simple networks, indicating that they will be favored compared to the non-duplicated networks. Fitness differences decrease with an increased number of nodes in the single network. When environmental change is too large, fitness differences equalize at zero since none of the networks can persist. This pattern is not qualitatively affected by the number of nodes in the single network.

Discussion

The longer-term consequences of WGDs have been discussed at large. Whole genome duplications increase both mutational and environmental robustness due to redundancy and increased genetic variation (6, 14-16, 27, 28). Many studies have reported on the co-option of extra duplicates specifically retained following WGD in different biological processes or pathways, increasing biological complexity or creating biological novelty (43-46). However, losing and retaining (a selection of) genes, the rewiring of gene interactions, and/or the functional divergence of genes takes time and selective advantages of WGDs on the short – or even immediate – term, often remain elusive. We have previously wondered about the 'conundrum' between the many examples of recurrent polyploidy and the existence of many polyploids of recent origin, which seem to contrast with the evidence of relatively few polyploidy events that have been established on the long-term, certainly within the same evolutionary lineage (6, 47, 48). The long-term fixation of polyploidy does not seem to occur randomly in space and time. One notable example is the biased distribution of 'survived' WGD events across independent plant lineages at the Cretaceous–Paleogene or K-Pg boundary, about 66 million years ago (Mya) (22). Other 'waves' of WGDs may correlate with periods of global climatic change during the Paleocene–Eocene, ca. 56–54 mya (18), or the last glaciations (23). The possible correlation between the 'establishment' of WGDs at times of environmental

upheaval is interesting, but, although some interesting hypotheses have been put forward (49, 50), remains to be explained. The fact that polyploids can survive drastically changing conditions or cataclysmic impacts while their diploid progenitors cannot, suggests a short-term, perhaps even immediate, evolutionary advantage for polyploids.

Some of the immediate consequences of polyploidy have been well described (13, 14, 27). One of the most consistent effects of WGD is an increase in cell size, but physiological effects have also been often observed. For instance, first-generation autotetraploids of *Arabidopsis thaliana* instantaneously enhanced their salt tolerance when compared to their diploid ancestors (51). Neo-autotetraploid *Arabidopsis* lines were shown to experience a tradeoff, demonstrating lower fitness compared to diploid progenitors under non-saline conditions, but higher fitness in response to saline challenge. The authors proposed that in conditions of salinity stress the autopolyploid lineages would benefit from a fitness advantage that could contribute to their establishment and persistence. In turn, autotetraploid *Arabidopsis* had been shown to be also more drought tolerant (52). Tetraploid rice (*Oryza sativa*) and citrange (*Citrus sinensis* L. Osb. × *Poncirus trifoliata* L. Raf.) too, have an increased tolerance to salt and drought stress because of WGD, which affects the expression of genes involved in stress and phytohormone response pathways (53, 54). Similarly, tetraploid rootstock-grafted watermelon (*Citrullus lanatus*) plants are more tolerant to salt stress than are diploid plants (55). Although such physiological and cellular responses to stress have thus been frequently documented for polyploids (16), the exact molecular processes underlying these responses remain obscure (56). Both the ‘gigas’ effect shown by polyploids, as well as observations in shifts in photosynthetic rates or stress tolerance, are in line with our findings when considering polyploidy at the genomic and gene regulatory network level. The established extreme phenotypes in nature are merely the result from selection on the expanded phenotypic variation following WGD, incidentally improving fitness under novel environmental conditions (57), rather than the outcome from any directional and deterministic trait change in response to environmental change (which may increase polyploidization rates by itself (58)). As shown in Figure 3, simply considering the particular structure of duplicated networks, these networks show greater variation in trait values, solely likely being able to explain observations such as increased drought and salt tolerance. In an elegant simulation experiment in yeast, van Hoek and Hogeweg (59) showed that WGD can lead to increased fitness under conditions that require elevated fluxes for certain pathways by increasing the absolute dosage of all the genes in a pathway.

Direct comparison of our results with experimental data on autotetraploid transcriptional regulation is not straightforward. First, experimental data consist of transcript/protein or metabolite numbers relative to transcriptome size, cell number, or dry weight, whereas our output represents up- or downregulation compared to the population mean. Second, as with all models, our model is a crude simplification of reality, in which full dosage compensation is assumed, while dosage compensation in real GRNs is not well understood. There is some evidence that real biological networks have built-in structural mechanisms to deal with dosage shifts (60), but here again the variation between cells and genotypes will be considerable. Finally, experimental studies on the effects of genome duplication on transcription are rare, especially those focusing on the immediate effects of autopolyploidy. WGD increases the transcriptome size, but the degree of change depends on the genotype, and dosage changes of individual genes are variable (61). Gene expression differences normalized with transcriptome size and cell number are limited and the exact quantity depends on the genotype. Even different *Arabidopsis* ecotypes (thus with a slightly different genetic makeup) have been shown to have significantly different transcriptome responses for many genes in newly synthesized tetraploids (60). Thus, this study supports the notion that the response to polyploidy is (highly) variable and depends on the genomic composition, and indirectly corroborate our findings that even small changes in simple GRN networks can lead to quite different responses in their duplicated counterparts (Fig. 3). It is also interesting to note that gene expression alterations in the autotetraploids used (62) were developmentally stage specific, implicating that certain GRNs were active or inactive during different times or conditions, as expected. Gene expression alterations have been associated with trait changes that might be adaptive and therefore polyploidy might confer an immediate advantage, depending on the environmental conditions and the GRNs active (and useful) at a certain moment in time.

In stable, non-changing environments, polyploidy will often be disadvantageous, as shown by our simulations, but also observed in vivo. For instance, in *Heuchera cylindrica,* an herbaceous perennial plant, increased nutrient requirements following polyploidy constrain the ability of new polyploids to establish in the nutrient-poor habitats the diploid progenitors thrive in (63). Similar observations were made for the autopolyploid complex *Dianthus broteri*, where, although higher ploidies have developed specific photochemical processes to survive in extremely warm conditions, the reduced performance of higher cytotypes render them less competitive in the ‘normal’ (non-stressed) environment (64). Differential fitness is a first and foremost criterion underlying adaptive dynamics theory (65). Our simulation results suggest that duplicated networks – or their hosts - will be able to coexist to eventually replace their

simple progenitors only under substantial environmental change (Fig. 4), for instance, when they end up in contrasting environments, or when the environment is quickly changing. Developmental stochastic noise increases persistence in moderately fluctuating environments (66, 67), but exaggerated phenotypic changes are essential to persist when changes of the adaptive landscape occur from major disturbances. During such events, any mutation must, by necessity, shift the value towards that new fitness peak if they are to increase fitness (Fig. 5). Or, in other words, when the original adaptive peak is sinking, overshooting is necessary to reach the new rising adaptive peak under discontinuous and/or fast environmental change (Fig. 5).

We need to notice that, in contrast to our approach here, polyploids are continuously but in low frequencies produced by their non-duplicated ancestors. Polyploids emerge from meiotic failures that lead to unreduced gametes, which are documented to occur in low frequencies, i.e., 0.1 to 2% in vascular plants (68). This implies that every 1/1000 to 2/100 offspring (seeds) will experience this potential niche expansion. This number probably increases during times of environmental stress (16, 58). Given high fecundity in most (WGD) plants, our assumption of a one-to-one doubling event is thus not too far from reality. This implies that any potential for establishment will depend on their fitness advantage compared to their ancestors and the level of standing genetic variation of the ancestor population. Since we show polyploids to have 'exaggerated' traits of their ancestors, fitness advantages are to be expected when rapid and drastic environmental change is already in line with earlier ambient selection. For instance, when ancestor populations evolved under continuous warming, extreme heat waves will promote polyploid invasion. If such a period of warming would be followed by extreme cold, fitness advantages would disappear because the extreme phenotype would then exaggerate evolved maladaptations. On the other hand, we do observe - albeit much rarer - cases where the orientation of vectors in trait-space (Fig. 3, *A* and *D*), describing the relative contributions of the output traits to divergence between simple and duplicated networks, are almost opposite. In such cases, even when the fitness landscape changes more drastically, causing niche shifts rather than niche expansion, polyploids might be the 'hopeful monsters' being able to adapt, where their diploid progenitors go extinct (Fig. 5). As the overall observed directionality in trait expression after doubling decreases with increasing network size (see *SI Appendix*), we can expect WGD to provide more fuel for selection in organisms with more complex genomes under severe, but unpredictable, environmental changes. This way, whole genome duplication or polyploidy might even explain large 'jumps' in evolution, or so-called saltational evolution (69). We here deliberately use a narrative of a single tangible

trait (phenotype) responding to a unidirectional environmental change but would like to emphasize that our model is equally valid for more complex multifactorial environmental changes that provoke selection on a restricted set of genes. Evidently, when multiple genes interact in more complex networks, trade-offs and pleiotropic effects might eventually provoke maladaptive dynamics under such conditions and lead to failure of establishment because optimal phenotype-environment matching cannot be reached (70).

We thus show that, at least in theory, immediate consequences of polyploidy can be significant, but it remains to be further tested whether they can indeed explain the preferential survival of polyploids over diploids during periods of sudden environmental change or times of extinction, as previously suggested (6, 48, 71). Reconstructing what occurred tens of millions of years ago, such as during the K/Pg extinction event, remains extremely challenging. Nevertheless, genomes of extant organisms might hold some clues. For instance, Yu et al. (20), showed that, based on the analysis of 25 plant genomes, genes functioning at low temperature and in darkness and been duplicated through WGDs at around the K/Pg boundary, have been subsequently selected for retention following duplication. Immediately after the Chicxulub asteroid impact, global cooling and darkness have been shown to be the two main stresses (72). One can imagine that increased expression of genes functioning in shade avoidance and/or cold-responsive pathways, thereby enhancing the perception of light signals and/or increasing cold stress tolerance, might have increased the chance of survival directly after the cataclysmic events responsible for the K/Pg extinction. Continued selection on such highly expressed genes might then explain their retention on the longer-term.

Evolution experiments with real biological organisms might be another means to get further insights into the mechanistic underpinnings explaining why duplicated GRNs might confer a selective advantage for polyploids during stressful times. By ‘replaying the duplication tape of life’, fitness of nonpolyploid and polyploid species can be evaluated under normal and stressful conditions (24, 73-76). Transcriptomes can be sequenced, and phenotypic and physiological responses measured and linked to the duplication of genomes and gene regulatory networks. Finally, a more explicit eco-evolutionary modeling approach building on the work presented here but also considering the history of selection, multidimensionality, magnitude, frequency, and direction of environmental change is needed. This too remains the topic of future work.

Methods

Defining and initializing aGRNs. In the current study, we consider artificial gene regulatory networks (aGRNs), mimicking gene regulatory networks in the traditional sense i.e., a set of genes or proteins that interact with each other to define and control a specific function (42, 77-80). For instance, such networks can transduce signals from environmental cues into a proper phenotypic behaviour that allows an organism to respond to environmental changes. In our aGRNS, we discriminate between 'regulatory' genes or proteins (like transcription factors, TFs), regulating the activity of other genes or proteins, and so-called 'output' genes or proteins, which produce an 'output', such as a structural protein or a metabolite. We also consider so-called 'input' genes or proteins, which can 'sense' the environment, and which can receive an input value. All these different genes or proteins form the nodes of the network, while edges between nodes represent their interactions. Furthermore, the following rules apply: 1) networks have a fixed number of nodes and are built by a preferential attachment algorithm and thus have properties that are close to scale-free networks, and 2) all edges are directed and have weights to mimic the strength of regulation (interaction). For instance, a weight can be considered as the strength with which a regulator binds to its target, or alternatively, as the strength with which it induces - or represses - expression of its target. It should be mentioned that all 'simple' or 'single' (non-duplicated, ancestral) networks have two output nodes (while the duplicated network has four output nodes). As a result, for better interpretation of the outcomes, plots are two-dimensional (see below).

Although there is still debate as to what extent biological networks are truly 'scale-free' (81), there are reasons to believe that many biological networks at least have certain features similar to scale-free networks, such as a high diversity of node degrees and absence of nodes in the network that could be used to characterize the rest of the nodes (82). Therefore, here, we consider directed, weighted, scale-free networks as our initial networks. To generate these directed scale-free networks, we used the 'Preferential Attachment' algorithm (83-85). Using this algorithm, nodes have a higher chance to connect with nodes with a higher degree (more connections) compared to other nodes ('rich get richer') (85). Another significant characteristic of real biological GRNs is the high number of feed-forward loops (FFL; A regulates B, B regulates C, A regulates C)(86, 87). To enrich our aGRNs with FFLs, we used the algorithm of Herrera and Zufiria (88). By using this algorithm, the clustering coefficient of the network increases which in turn causes an increase in the number of triadic motifs in the network. Then, by controlling and changing the direction of edges, we can easily raise the number of FFL motifs in the network. It should also be noted that, for computational reasons, generated networks containing feedback loops (never-ending loops; A regulates B, B regulates C, and C

regulates A, or self-regulation like A regulating A) are discarded. Besides, although such motifs do occur in real biological networks, they are rare (89).

As stated above, in our aGRNs, three different kinds of nodes are distinguished (Fig. 1). Nodes with zero in-degree and non-zero out-degree are referred to as input nodes. The number of input nodes is variable and depends on the preferential attachment algorithm, but usually lies between 20% and 40%. Nodes with non-zero in-degree and non-zero out-degree represent regulators, affecting other nodes (genes), such as transcription factors. Finally, nodes with non-zero in-degree and zero out-degree are output nodes defining the 'phenotype' or 'behaviour'. To allow analyses of the phenotypic effects in an easily conceivable two-dimensional space (see further), the number of output nodes is artificially set to 2. In practice, this means that, for a network with n nodes, the network is built on n-2 nodes, and after the n-2 network has been built, the last two nodes are added. These two nodes cannot connect to others, but others can connect to them, again by applying the preferential attachment algorithm. Every edge has a weight corresponding to the 'strength' of the regulatory interaction between genes. The weight of the edges is determined randomly by a standard Normal Distribution, generating values between -1 and 1, with positive values indicating stimulation, and negative values repression. After initialization of the network, the weights of the edges are fixed. However, to model changes in the network after receiving different values of input nodes, i.e., mimicking environmental cues, we have defined an 'activation level' for each node. Initially, activation levels of all nodes are set to 0, but during simulation, the input signal will determine the activation level of the input node(s), which will then be propagated through the network changing the activation level of each downstream node in function of all incoming edge weights and the activation level of all previous nodes (90). Concretely, this way, when the value/expression of one node/gene is increased (or decreased), this would lead to increased (or decreased) dosage of a regulator, in turn being responsible for the increased (or decreased) production of its target, and so on.

Network duplication. To mimic WGDs, we simply duplicate all nodes of the network. However, this means that, if a regulator A regulates nodes B and C, its duplicate A' regulates the duplicated targets B' and C', but also the original targets B and C. In turn, the original regulator A also regulates all four targets, B, B', C, and C' (Fig. 2). The edge weights between corresponding nodes in the non-duplicated (A-B) and duplicated (A-B, A'-B', A'-B, A-B') network remain unaltered. It should be noted that such operation mimics only part of the effects of a polyploidisation event, i.e., the effects of genome doubling (autopolyploidy) and not those

of genome merging (allopolyploidy). As a result, throughout the paper, we will only consider autopolyploidy, where the 'own' genome gets duplicated, rather than allopolyploidy, where the duplicated genome is obtained from the merging of genomes of different species. Although there is still discussion on the ratio of autopolyploids versus allopolyploids in the polyploid realm, there is reason to believe that autopolyploids are much more frequent than previously thought (91). We are aware of the fact that, in autopolyploids, unlike in allopolyploids, the duplicated genes might be seen more as different alleles of the same gene, rather than as different genes, but we feel this will not have a major effect on our conclusions, because even when considered only different alleles, it will affect certain traits (e.g., due to dosage effects) (13, 14, 16), and when there is no recombination, they can be considered separate genes.

Signal propagation in the network. One of the main purposes of our simulations is to see how signals, such as environmental cues, propagate over simple (non-duplicated) versus duplicated networks, the hypothesis here being that, because of the specific structure and a denser wiring of duplicated networks (Fig. 2), greater parts of the network – and thus more genes - are affected, with consequently, greater variation in output values. As far as we know, this has not been studied in a biological context, and certainly not in the context of duplicated networks and polyploidy or genome duplication. Signal propagation functions in the network (see further) will determine the output values, and thus the phenotype. We evaluated output changes by using constrained propagation using the hyperbolic tangent function (tanh) with the max value of "+1" and min value of "-1" (eq. 1). This function is typically used to determine the activation level of nodes in neural networks (90, 92, 93). Similar 'sigmoid' functions have also been used previously to study signal processing in complex regulatory gene networks (40, 94):

$$A_a = tanh(\sum_{i \in Ng} A_i \times W_{i \to a}) \qquad \text{(Eq. 1)}$$

where A_a is the activation level of node a, $W_{i \to a}$ is the weight of the edge from node i to node a, and Ng is the list of nodes that are connected to node A.

Depending on the different input signals, different outputs will be reached. This constrained implementation mimics biological networks in that it considers minimal and maximum values for, for instance, an increase in gene expression (increase in gene expression is not unlimited). We additionally provide a sensitivity analysis for an unconstrained linear propagation algorithm in *SI Appendix*, Text S1.

Environmental changes, such as for instance changes in temperature, are mimicked by changing the signal values of input nodes (input values are drawn from a uniform distribution between -1 and +1) and following their propagation over the network. Values from the two output nodes, *o1*, *o2*, are interpreted as a phenotype in a two-dimensional (2D) trait space (hence the two output nodes). Since we have four output nodes for duplicated networks, like *o1, o2, and o1' and o2',* each output value is calculated as the average of the corresponding output nodes, e.g. *(o1+o1')/2.* We thus consider a conservative but realistic full dosage compensation of the gene expression after WGD (27). This representation allows us to quantify phenotypic changes by means of Phenotypic Trajectory Analysis, PTA (95, 96). In brief, this approach allows us to understand whether evolutionary divergence between pairs of populations, here the ancestral (single/simple) and duplicated networks, is parallel, convergent, divergent, or random (97). To this end, vectors $\vec{o}[o1, o2]$ are drawn from the 'phenotype/trait value' (e.g., positions of the two output nodes in a 2D space, see Fig. *3A*) of an ancestral simple network to the 'phenotype/trait value' of its duplicated counterpart. Changes in the vector length demonstrate how much the eventual trait value (the phenotype produced by the duplicated genome) is changing compared to the initial value of the simple aGRN. The distribution of these differential vector lengths therefore identifies the strength of the phenotypic (and putative niche) shifts, and thus the strength of the divergence due to genome doubling. The orientation of a vector in trait-space describes the relative contributions of the output traits to divergence between that pair of populations. Changes in angular dispersion between the ancestral and duplicated phenotype indicate whether phenotypic changes among all independent network doubling events occur in parallel for all doubling events (absolute angles in the trait space similar, hence showing directional or parallel evolution), in the same direction of the initial phenotypic position (relative angles between the ancestral and doubles genotypes are zero, showing niche expansion), or completely random (both absolute and relative angles randomly distributed across trait space). We calculated overall phenotypic variation at the population-level (hence a population of 10K single versus doubled genomes generated by the same initialization) by calculating the variance of the mean of the two output nodes, and by multiplying the variance of the two output values. Artificial gene regulatory networks consisting of 10, 20, 40, 60, 80, and 100 nodes were generated and exposed to 10,000 environmental conditions by drawing the input values (activation levels for the input nodes) from a uniform distribution between -1, and +1, as stated previously. We report these variances for 400 populations of 10K networks per category (e.g., number of nodes in the single network; Fig. 3*B*). We specifically test how doubling of simple scale-free networks of size *n* affects

phenotypic variation relative to the variation generated by randomly selected scale-free networks with the same number of nodes ($2n$) but the same parameter distributions. Unless explicitly mentioned, we present data for ancestral (simple) networks consisting of 10 nodes and node/edge values initiated from the uniform distribution. We tested the robustness of our analysis by sensitivity analyses for the full range of node numbers (*SI Appendix*, Text S2) and network initialisation from Gaussian N(0,1) and mixed uniform U[-1,1]-Gaussian N(0,1) distributions (see *SI Appendix*, Text S3).

Linking phenotypes to the environment: fitness. Besides studying theoretical phenotype-WGD mapping, we moved one step further and tested the hypothesis that duplicated genomes provide fitness benefits under larger environmental changes. To this end, we simplified the two-dimensional phenotype vector $\vec{o}[o1, o2]$ towards its one-dimensional average value [$\bar{o} = (o1 + o2)/2$] as it was shown to generate qualitatively similar insights (Pearson correlation r between 0.45-0.52). We start from a population of simple networks that are well adapted to their environment and assume that the individual phenotypes are all centred around the fitness optimum. We simulated a genome doubling effect of all simple genetic networks and quantified the mean fitness $\bar{\boldsymbol{w}}$ of both the simple and duplicated populations in the reference environment and under environmental changes of different magnitudes.

For each size of simple networks (resp. 10, 20, 40, 60, 80, and 100 nodes), we constructed 10,000 simple and their duplicated networks as described above. All these 10K networks have different topologies and different edge weights. For all these simple and duplicated networks, we here provide insights from simulations with an input value of 0.01, creating a large compilation of networks with different output values centered around 0 (see above). This specific input value represents our 'reference environment' and guarantees that the average phenotype of the population is close to the fitness optimum (0.01), when we assume that fitness $\boldsymbol{w}$ is inversely proportional to the difference between input and output value. To assess fitness in the reference environment and how it is affected by the underlying fitness function, the performance of each network was calculated using both a negative linear and a Gaussian fitness function.

For the linear function, $\boldsymbol{w} = 1 - |A_i - A_o|$ and $\boldsymbol{w} = 0$ when $|A_i - A_o| > 1$ (Eq. 2),

while for the Gaussian function, $\boldsymbol{w} = \frac{1}{\sqrt{2\pi}} e^{\frac{-|A_i - A_0|^2}{2}}$ (Eq. 3).

Networks with a phenotype (output A_o) similar to the reference environment (input A_i) will have the maximal fitness, and this value decreases to zero under large deviation from the

reference environment. The mean fitness $\bar{\boldsymbol{w}}$ of the simple networks under this reference environment was set to the maximum of 1, to which all other measured mean fitness values were contrasted with.

Next, the populations of these simple and duplicated aGRNs were subjected to deviating environmental conditions as input A_i, our environmental change gradient, and $\boldsymbol{w}$ was again calculated according to the output phenotype A_o as in Eq. 2 and Eq. 3. To this end, the input value A_i of one randomly chosen input node is changed gradually with ΔA_i. If there is more than one input node, which is always the case for duplicated networks, the input of the other input nodes is kept at zero. Given the fitness values $\boldsymbol{w}$ ranging between 0 and 1 under all conditions, it could be considered as a survival rate: networks with an output equal to the input have a survival of 100%, whereas those differing a lot will approach a survival rate of zero, pending the used fitness function (Eq. 2-3). The mean fitness over the 10K simple or duplicated networks ($\bar{\boldsymbol{w}}$) then represents the average population-level survival rate. Given the choice of the reference environment, $\bar{\boldsymbol{w}}$ in the environment with input 0.01 is maximal for the simple networks and used as a baseline for the performance of doubled networks in this reference environment and for all simple and doubled networks in environments with a different input value, hence environmental change. We represent both the fitness relative to this baseline for networks of 10 nodes, and the differential fitness, $\frac{\bar{w}_{double}-\bar{w}_{simple}}{\bar{w}_{simple}}$ for simulations of networks with different number of nodes.

Data availability. Documentation and software to generate artificial scale-free gene regulatory networks and their duplicated versions simulating the result of whole genome duplication (WGD) can be found at: https://github.com/Mehrshad-Ebadi/SC-as-aGRNs. Examples of networks and their duplicated versions of different sizes can be found there as well.

Acknowledgements

We thank Yao Yao (Biocomputing and Developmental Systems Group, University of Limerick, Limerick, Ireland) and Felipe Kauai (VIB-Ghent University, Ghent, Belgium) for helpful discussions. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522) and from Ghent University (Methusalem funding, BOF.MET.2021.0005.01) (to Y.V.d.P.) and Ghent University (UGent BOF/STA/202009/039, BioGraph BOF.24Y.2019.0010.01) (to P.A.). The

authors want to thank all reviewers for their constructive comments and suggestions that greatly helped improve the paper.

References

1. P. S. Soltis, X. Liu, D. B. Marchant, C. J. Visger, D. E. Soltis, Polyploidy and novelty: Gottlieb's legacy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369** (2014).
2. L. Comai, The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836-846 (2005).
3. C. Morgan, H. Zhang, C. E. Henry, F. C. H. Franklin, K. Bomblies, Derived alleles of two axis proteins affect meiotic traits in autotetraploid Arabidopsis arenosa. *Proc. Natl. Acad. Sci. U S A* **17**, 8980–8988 (2020).
4. D. A. Levin, Minority cytotype exclusion in local plant populations. *Taxon* **24**, 35–43 (1975).
5. J. F. Wendel, The wondrous cycles of polyploidy in plants. *Am. J. Bot.* **102**, 1753-1756 (2015).
6. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411-424 (2017).
7. J. S. Taylor, I. Braasch, T. Frickey, A. Meyer, Y. Van de Peer, Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**, 382-390 (2003).
8. P. Dehal, J. L. Boore, Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
9. P. S. Soltis, D. E. Soltis, The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* **60**, 561-588 (2009).
10. J. B. Landis, D. E. Soltis, P. S. Soltis, Comparative transcriptomic analysis of the evolution and development of flower size in Saltugilia (Polemoniaceae). *BMC Genomics* **18**, 475 (2017).
11. J. H. Leebens-Mack *et al.*, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679-685 (2019).
12. D. Schluter, Ecology and the origin of species. *Trends Ecol. & Evol.* **16**, 372-380 (2001).
13. K. Bomblies, When everything changes at once: finding a new normal after genome duplication. *Proc. Biol. Sci.* **287**, 20202154 (2020).
14. P. Baduel, S. Bray, M. Vallejo-Marin, F. Kolář, L. Yant, The "polyploid hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* **6**, 117 (2018).
15. M. te Beest *et al.*, The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* **109**, 19-45 (2012).
16. Y. Van de Peer, T. L. Ashman, P. S. Soltis, D. E. Soltis, Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11-26 (2021).
17. J. A. Fawcett, S. Maere, Y. Van de Peer, Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc. Natl. Acad. Sci. U S A* **106**, 5737-5742 (2009).
18. L. Cai *et al.*, Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* **221**, 565-576 (2019).
19. E. J. M. Koenen *et al.*, The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. *Syst. Biol.* 10.1093/sysbio/syaa041 (2020).
20. S. Wu, B. Han, Y. Jiao, Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* **13**, 59-71 (2020).
21. J. Chang *et al.*, The genome of the king protea, Protea cynaroides. *Plant J.* **113**, 262-276 (2023).

22. K. Vanneste, G. Baele, S. Maere, Y. Van de Peer, Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* **24**, 1334-1347 (2014).
23. P. Y. Novikova, N. Hohmann, Y. Van de Peer, Polyploid Arabidopsis species originated around recent glaciation maxima. *Curr. Opin. Plant Biol.* **42**, 8-15 (2018).
24. A. M. Selmecki *et al.*, Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349-352 (2015).
25. A. L. Scott, P. A. Richmond, R. D. Dowell, A. M. Selmecki, The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Mol. Biol. Evol.* **34**, 2690-2703 (2017).
26. A. Force *et al.*, Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545 (1999).
27. J. J. Doyle, J. E. Coate, Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* **180**, 1-52 (2019).
28. Y. Van de Peer, S. Maere, A. Meyer, The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725-732 (2009).
29. Y. Yao, K. Marchal, Y. Van de Peer, Improving the adaptability of simulated evolutionary swarm robots in dynamically changing environments. *PLoS One* **9**, e90695 (2014).
30. Y. Yao, V. Storme, K. Marchal, Y. Van de Peer, Emergent adaptive behaviour of GRN-controlled simulated robots in a changing environment. *PeerJ* **4**, e2812 (2016).
31. Y. Yao, L. Carretero-Paulet, Y. Van de Peer, Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy. *PLoS One* **14**, e0220257 (2019).
32. T. D. Cuypers, P. Hogeweg, A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS Comput. Biol.* **10**, e1003547 (2014).
33. T. D. Cuypers, J. P. Rutten, P. Hogeweg, Evolution of evolvability and phenotypic plasticity in virtual cells. *BMC Evol. Biol.* **17**, 60 (2017).
34. G. C. Conant, K. H. Wolfe, Functional partitioning of yeast co-expression networks after genome duplication. *PLOS Biol.* **4**, 545-554 (2006).
35. J. Berg, M. Lässig, A. Wagner, Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol. Biol.* **4**, 51 (2004).
36. G. C. Conant, Rapid reorganization of the transcriptional regulatory network after genome duplication in yeast. *Proc. Biol. Sci.* **277**, 869-876 (2010).
37. W. Peng, R. Song, M. Acar, Noise reduction facilitated by dosage compensation in gene networks. *Nat. Comm.* **7**, 12959 (2016).
38. M. Acar, B. F. Pando, F. H. Arnold, M. B. Elowitz, A. van Oudenaarden, A general mechanism for network-dosage compensation in gene circuits. *Science* **329**, 1656-1660 (2010).
39. Y. S. Posadas-García, C. Espinosa-Soto, Early effects of gene duplication on the robustness and phenotypic variability of gene regulatory networks. *BMC Bioinformatics* **23**, 509 (2022).
40. S. Ciliberti, O. C. Martin, A. Wagner, Innovation and robustness in complex regulatory gene networks. *Proc. Natl. Acad. Sci. U S A* **104**, 13591-13596 (2007).
41. R. De Smet, Y. Van de Peer, Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* **15**, 168-176 (2012).
42. T. Schlitt, A. Brazma, Current approaches to gene regulatory network modelling. *BMC Bioinformatics* **8**, S9 (2007).
43. T. Unver *et al.*, Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci U S A* **114**, E9413-E9422 (2017).

44. S. Sato *et al.*, The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635-641 (2012).
45. J. Cai *et al.*, The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47**, 65-72 (2015).
46. L. Zhang *et al.*, The water lily genome and the early evolution of flowering plants. *Nature* **577**, 79-84 (2020).
47. G. K. Wong *et al.*, Sequencing and analyzing the transcriptomes of a thousand species across the tree of life for green plants. *Annu. Rev. Plant Biol.* **71**, 1.1–1.25 (2020).
48. L. Carretero-Paulet, Y. Van de Peer, The evolutionary conundrum of whole genome duplication. *Am. J. Bot.* **107**, 1101-1105 (2020).
49. D. A. Levin, D. E. Soltis, Factors promoting polyploid persistence and diversification and limiting diploid speciation during the K–Pg interlude. *Curr. Opin. Plant Biol.* **42**, 1-7 (2018).
50. M. Freeling, Picking up the Ball at the K/Pg Boundary: The distribution of ancient polyploidies in the plant phylogenetic tree as a spandrel of asexuality with occasional sex. *Plant Cell* **29**, 202-206 (2017).
51. D. Y. Chao *et al.*, Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. *Science* **341**, 658-659 (2013).
52. J. C. del Pozo, E. Ramirez-Parra, Deciphering the molecular bases for drought tolerance in Arabidopsis autotetraploids. *Plant Cell Environ.* **37**, 2722–2737 (2014).
53. P. M. Yang, Q. C. Huang, G. Y. Qin, S. P. Zhao, J. G. Zhou, Different drought-stress responses in photosynthesis and reactive oxygen metabolism between autotetraploid and diploid rice. *Photosynthetica* **52**, 193-202 (2014).
54. M. Ruiz *et al.*, Tetraploidy enhances boron-excess tolerance in carrizo citrange (Citrus sinensis L. Osb. x Poncirus trifoliata L. Raf.). *Front. Plant Sci.* **7**, 701 (2016).
55. H. Zhu *et al.*, Genome duplication improves the resistance of watermelon root to salt stress. *Plant Physiol. Biochem.* **133**, 11-21 (2018).
56. D. Fox, D. E. Soltis, P. S. Soltis, T.-L. Ashman, Y. Van de Peer, Polyploidy: a biological force from cells to ecosystems. *Trends Cell Biol.* **30**, 688-694 (2020).
57. J. Ramsey, D. W. Schemske, Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* **33**, 589-639 (2002).
58. N. De Storme, G. P. Copenhaver, D. Geelen, Production of diploid male gametes in Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. *Plant Physiol.* **160**, 1808-1826 (2012).
59. M. J. A. Van Hoek, P. Hogeweg, Metabolic adaptation after whole genome duplication. *Mol. Biol. Evol.* **26**, 2441-2453 (2009).
60. R. Song, P. Liu, M. Acar, Network-dosage compensation topologies as recurrent network motifs in natural gene networks. *BMC Systems Biol.* **8**, 69 (2014).
61. M. J. Song, B. I. Potter, J. J. Doyle, J. E. Coate, Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana*. *Plant Cell* **32**, 1434-1448 (2020).
62. Z. Yu *et al.*, Impact of natural genetic variation on the transcriptome of autotetraploid Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U S A* **107**, 17809-17814 (2010).
63. T. J. Anneberg, K. A. Segraves, Neopolyploidy causes increased nutrient requirements and a shift in plant growth strategy in *Heuchera cylindrica*. *Ecology* 10.1002/ecy.4054, e4054 (2023).
64. J. López-Jurado, F. Balao, E. Mateos-Naranjo, Polyploidy-mediated divergent light-harvesting and photoprotection strategies under temperature stress in a Mediterranean carnation complex. *Environ. Exp. Bot.* **171**, 103956 (2020).

65. F. Dercole, S. Rinaldi, *Analysis of Evolutionary Processes: The Adaptive Dynamics Approach and Its Applications* (Princeton University Press, 2008).
66. M. Acar, J. T. Mettetal, A. van Oudenaarden, Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471-475 (2008).
67. A. Eldar, M. B. Elowitz, Functional roles for noise in genetic circuits. *Nature* **467**, 167-173 (2010).
68. J. M. Kreiner, P. Kron, B. C. Husband, Evolutionary dynamics of unreduced gametes. *Trends Genet.* **33**, 583-593 (2017).
69. G. Theissen, Saltational evolution: hopeful monsters are here to stay. *Theory Biosci.* **128**, 43-51 (2009).
70. S. P. Brady *et al.*, Causes of maladaptation. *Evolutionary Applications* **12**, 1229-1242 (2019).
71. R. Lohaus, Y. Van de Peer, Of dups and dinos: evolution at the K/Pg boundary. *Curr. Opin. Plant Biol.* **30**, 62-69 (2016).
72. P. Schulte *et al.*, The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science* **327**, 1214-1218 (2010).
73. Q. Bafort *et al.*, Studying whole-genome duplication using experimental evolution of *Chlamydomonas*. in Polyploidy: Methods and Protocols, Y. Van de Peer, Ed. (Springer US, New York, NY, 2023), 10.1007/978-1-0716-2561-3_18, pp. 351-372.
74. T. Wu *et al.*, Studying whole-genome duplication using experimental evolution of *Spirodela polyrhiza*. in Polyploidy: Methods and Protocols, Y. Van de Peer, Ed. (Springer US, New York, NY, 2023), 10.1007/978-1-0716-2561-3_19, pp. 373-390.
75. Q. Bafort, T. Wu, A. Natran, O. De Clerck, Y. Van de Peer, The immediate effects of polyploidization of *Spirodela polyrhiza* change in a strain-specific way along environmental gradients. *Evol. Lett.* **7**, 37-47 (2023).
76. T. J. Anneberg, E. M. O'Neill, T. L. Ashman, M. M. Turcotte, Polyploidy impacts population growth and competition with diploids: multigenerational experiments reveal key life-history trade-offs. *New Phytol.* **238**, 1294-1304 (2023).
77. L. Mendoza, E. R. Alvarez-Buylla, Dynamics of the genetic regulatory network for Arabidopsis thaliana flower morphogenesis. *J. Theor. Biol.* **193**, 307-319 (1998).
78. S. Kidokoro, K. Shinozaki, K. Yamaguchi-Shinozaki, Transcriptional regulatory network of plant cold-stress responses. *Trends Plant Sci.* **27**, 922-935 (2022).
79. T. R. Sorrells, A. D. Johnson, Making sense of transcription networks. *Cell* **161**, 714-723 (2015).
80. T. Schlitt, A. Brazma, Modelling in molecular biology: describing transcription regulatory networks at different scales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 483-494 (2006).
81. A. D. Broido, A. Clauset, Scale-free networks are rare. *Nat. Commun.* **10**, 1017 (2019).
82. R. Albert, Scale-free networks in cell biology. *J Cell Sci* **118**, 4947-4957 (2005).
83. B. J. Prettejohn, M. J. Berryman, M. D. McDonnell, Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Front. Comput. Neurosci.* **5**, 11 (2011).
84. A. L. Barabasi, Scale-free networks: a decade and beyond. *Science* **325**, 412-413 (2009).
85. A. L. Barabasi, E. Bonabeau, Scale-free networks. *Sci. Am.* **288**, 60-69 (2003).
86. M. B. Cooper, M. Loose, J. F. Brookfield, Evolutionary modelling of feed forward loops in gene regulatory networks. *Biosystems* **91**, 231-244 (2008).
87. F. Mottes, C. Villa, M. Osella, M. Caselle, The impact of whole genome duplications on the human gene regulatory networks. *PLoS Comput. Biol.* **17**, e1009638 (2021).
88. C. Herrera, P. J. Zufiria (2011) Generating scale-free networks with adjustable clustering coefficient via random walks. in *2011 IEEE Network Science Workshop*, pp 167-172.

89. N. H. Tran, K. P. Choi, L. Zhang, Counting motifs in the human interactome. *Nat. Commun.* **4**, 2241 (2013).
90. A. Moreira, C. Rennó-Costa (2021) Evolutionary Strategies Applied to Artificial Gene Regulatory Networks. doi: https://doi.org/10.1101/2021.09.28.462218 (bioRxiv).
91. M. S. Barker, N. Arrigo, A. E. Baniaga, Z. Li, D. A. Levin, On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **210**, 391-398 (2016).
92. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436-444 (2015).
93. C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: comparison of trends in practice and research for deep learning. https://doi.org/10.48550/arXiv.1811.03378 (2018).
94. A. Wagner, Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc. Natl. Acad. Sci. U S A* **91**, 4387-4391 (1994).
95. S. P. De Lisle, D. I. Bolnick, A multivariate view of parallel evolution. *Evolution* **74**, 1466-1481 (2020).
96. D. C. Adams, M. L. Collyer, A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evolution* **63**, 1143-1154 (2009).
97. D. I. Bolnick, R. D. H. Barrett, K. B. Oke, D. J. Rennison, Y. E. Stuart, (Non)Parallel Evolution. *Ann. Rev. Ecol. Evol. Syst.* **49**, 303-330 (2018).

FIGURE LEGENDS

Fig. 1. Two examples of an aGRN of 10 nodes generated by the preferential attachment algorithm. All nodes represent regulatory genes or proteins, except nodes 8 and 9 in both networks, which are output nodes. Nodes 5 and 6 can act as input nodes since all edges are outgoing. Weight values are also indicated. Positive weight values represent induction, while negative weight values indicate repression (as for example in gene expression). The topology of a specific aGRN is unique and can be considered the genotype, while the output nodes or node values define the phenotype. See text for details

Fig. 2. Example of a simple or ancestral (left) and duplicated (right) aGRN.

Fig. 3. Phenotypic Trajectory Analysis (see Methods) comparing a population of single versus its duplicated networks. (*A*) The value of one output node is plotted against the value of the second output node for simple (blue dots) and duplicated (orange) networks for a simple GRN of 10 nodes. Thinning has been applied and from the 1,000,000 values only a fraction is shown, to facilitate interpretation. (*B*) The variance for simple and duplicated networks for networks of 10, 20, 40, and 80 nodes (400 networks consisting of 10K single/double networks per size category). Variance of the output is increasing with node additions, but duplicated genomes always have higher variance compared to their unduplicated counterparts. Red arrows denote the difference in variance between duplicated networks and random networks with an equal number of nodes but not having the typical duplicated topology (structure doubling versus node doubling) (*C*) Cumulative density function of the phenotypic variance σ as measured by multiplying variance of both (mean) output node values in the 10K simulated simple GRNs of 10 nodes, and their duplicated counterparts. (*D*) Angular dispersion of the relative angles between the single and doubled networks for 10K simulations of simple GRNs of 10 nodes, and their duplicated counterparts.

Fig. 4. Mean fitness $\bar{w}$ of simple and duplicated networks relative to the fitness of the population of simple networks in the reference environment, as a function of different input values, assuming a Gaussian fitness function (eq. 3) (*A*) and a linear fitness function (eq. 2) (*C*). Differential fitness as a function of different input variables that represent environmental change with fitness modelled by a Gaussian (*B*) and linear function for networks of different

size (*D*). Note that we only report relative $\bar{w}$ for input values leading to non-zero values for the single network.

Fig. 5. 3D representation of fitness landscapes in which hills, corresponding to local adaptive peaks, are surrounded by valleys or depressions, corresponding to regions of the phenotype space where no survival is possible. Polyploidy may allow a wider and faster exploration of phenotypic space, ultimately conferring a potential adaptive advantage under challenging environmental conditions. Blue-green dots are individuals that can survive, red dots denote organisms that cannot survive. In a stable environment (top left panel), non-polyploid organisms are expected to have reached their local adaptive peaks. WGD results in an expansion of the phenotypic space covered by the population, although some polyploid genotypes might survive, most polyploids cannot survive in this environment (bottom-left panel). Adaptive landscapes are readily distorted by environmental challenges, such as cataclysmic or extinction events (right panels), resulting in shifts in the relative locations of their adaptive peaks. Under these conditions, although most diploids are expected to perish (top right panel), some polyploid organisms (which could be referred to as 'hopeful monsters'), featured by wider accessible phenotype space (see text for details), have better chances to fall near the peak of a newly formed adaptive hill and thus to acquire the necessary evolutionary innovations to colonize novel niches (bottom right panel).