Trends in
Plant Science

CellPress

Opinion

The ORFans' tale: new insights in plant biology

Ali Zeeshan Fakhar,[1,2] Jinbao Liu,[1,2] Karolina M. Pajerowska-Mukhtar,[1] and M. Shahid Mukhtar [1,*,@]

Orphan genes (OGs) are protein-coding genes without a significant sequence similarity in closely related species. Despite their functional importance, very little is known about the underlying molecular mechanisms by which OGs participate in diverse biological processes. Here, we discuss the evolutionary mechanisms of OGs' emergence with relevance to species-specific adaptations. We also provide a mechanistic view of the involvement of OGs in multiple processes, including growth, development, reproduction, and carbon-metabolism-mediated immunity. We highlight the interconnection between OGs and the sucrose nonfermenting 1 (SNF1)-related protein kinases (SnRKs)–target of rapamycin (TOR) signaling axis for phytohormone signaling, nutrient metabolism, and stress responses. Finally, we propose a high-throughput pipeline for OGs' interspecies and intraspecies gene transfer through a transgenic approach for future biotechnological advances.

Highlights

Orphan genes (OGs) are taxonomically restricted, without significant sequence similarities to closely related species; they might have evolved from noncoding sequences via rearrangements or duplication and divergence.

OGs from divergent species can interact with conserved regulators – such as sucrose non-fermenting 1 (SNF1)-related protein kinases (SnRKs), nuclear factor Y subunit C4 (NF-YC4) – and display the evolutionarily conserved functions.

OGs have been shown to promote growth and development in the carbon metabolism of plants, possibly through the target of rapamycin (TOR)–SnRK1 signaling axis.

OGs have been shown to compete with pathogenic effectors for their interaction with central regulators, and thereby participate in a growth–defense trade-off.

OGs can retain their function when expressed as transgenes in other species, and can thus be used for biotechnological intervention against various environmental stresses.

Historically, gene evolution has been driven by multiple mechanisms, of which gene duplication and rearrangement are considered to be particularly important. Every temporal series of populations, organisms, or genes harbors OGs, sequences also known as ORFans (Orphan Open Reading Frames) that do not have homologs in other lineages, nor is their evolutionary origin clearly understood [1]. For these reasons, OGs are known as taxonomically restricted genes (TRGs) without identifiable sequence similarities in closely related species [2]. OGs have a narrow phylogenetic distribution, with up to 30% of them found in all gene catalogs of each species analyzed [3]. Systematic studies in primates [4] and plants suggest that OGs evolved through gene duplication and divergence, as well as exaptation from transposable elements (TEs) (detailed later) [5]. Additionally, other plausible pathways of their evolution include chromosomal rearrangements and *de novo* origination from noncoding genomic areas [6]. Identifying OGs is crucial for understanding their evolution and functional roles in various organisms. Comparative genomics using the Basic Local Alignment Search Tool (BLAST) was one of the earliest methods used to detect OGs by comparing genome sequences from different species to identify conserved genes [7]. This method, however, has limitationsas some OGs may have evolved rapidly and are not conserved across species, making them difficult to discover. Consequently, other methods – such as BLASTp, Phylostratigraphy, and ORFan-Finder – have been developed as alternatives to overcome these limitations [8]. While several OGs have been identified in plants to date (Table 1) (T. Racovski, PhD thesis, University of Exeter, 2019) [9–20], this number is merely the tip of the iceberg. The current challenges for the identification of OGs include a lack of chromosome-scale genome assembly, genomic annotation availability, a limited number of high-quality reference genomes, and OG-specific sequencing technology followed by computational pipelines [20]. A case of computational and statistical analyses of the rice genomes led to the discovery of only 1926 out of a total of 18 398 OG candidates as orphan proteins [21]. Despite the emerging evidence that OGs are crucial for plant growth and development, little is known about their functional traits and underlying molecular mechanisms. In this Opinion, we discuss the potential mechanisms of action of OGs in plants, their role in lineage-specific features such as phenotypic diversity, mitigating biotic and abiotic stresses, species-specific adaptive processes, regulation of metabolic pathways, modulation of carbon metabolism, and their role in

[1]Department of Biology, University of Alabama at Birmingham, 1300 University Blvd., Birmingham, AL 35294, USA
[2]These authors contributed equally.

*Correspondence:
smukhtar@uab.edu (M.S. Mukhtar).
@Twitter: @SMukhtarlab (M.S. Mukhtar).

© 2023 Elsevier Ltd. All rights reserved.

Table 1. Orphan genes identified in plants

Orphan gene	Host plant	Function	Refs
PpARDT	*Physcomitrium patens*	Drought tolerance	[9]
Xio1	*Oryza* spp	Resistance against *Xanthomonas oryzae* pv. *oryzae*	[15]
Xa7	*Oryza sativa*	Resistance against *Xanthomonas oryzae* pv. *oryzae*	[18]
BrOGs	*Brassica rapa*	Sugar/nutrient metabolism	[12]
OGs	*Vigna unguiculata*	Drought tolerance/ environmental adaptations	[13]
BSGs	*Brassica rapa*	Growth and development	[11]
Ms2	*Triticum aestivum*	Male sterility	[16]
IAPAR59	*Coffea arabica*	Drought tolerance	[14]
TaFROG	*Triticum aestivum*	Resistance against mycotoxigenic fungus	[17]
LSGs	*Citrus sinensis*	Environmental adaptability	[19]

protein expression. We also propose a pipeline for the identification, characterization, and cross-species transfer of OGs using transgenic technologies and multiomics approaches.

The roles of OGs in plant nutrient status, growth, and development

Due to the lack of sequence similarity and possibly recognizable protein domains, functional motifs, and identifiable folds in the proteins encoded by OGs, the fundamental question is how these OGs operate in various biological processes. What are the molecular mechanisms at work? Do OGs interact with other cellular components and central regulators at the molecular and genomic levels during growth and development/tissue-specific responses? To obtain functional clues of novel genes, widely used approaches include expression [22] and coexpression pattern analysis such as WGCNA (weighted gene correlation network analysis, also known as weighted gene coexpression network analysis) followed by gene ontology (GO) term analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [23], among other tools. For instance, a recent study in melon found that 18 coexpressed OGs are overrepresented in biochemical processes including pentose and glucuronate interconversions, cutin, suberin, and wax biosynthesis, as well as starch and sucrose metabolism [24]. Generally, carbohydrates are involved in multiple growth and developmental processes, especially during male gametogenesis. They maintain the proper nutrition levels in the plant cell but can also act as signaling molecules, playing a crucial role in cell fate determination [25]. The functional analysis of another OG, Qua-Quine Starch ('by any means starch'in Latin; *QQS*) showed its importance in altering the carbon patterns, protein content, and nitrogen partitioning in plants [12,26]. In a large-scale experiment [12], 43 *Brassica rapa*-specific OGs (*BrOGs*) were transformed into arabidopsis (*Arabidopsis thaliana*). When compared with wild-type plants, 19 of the 43 BrOG-overexpression (*BrOGOE*) transgenic lines demonstrated sugar-related phenotypes, while 42 showed varied soluble sugar levels. In addition, *BrOGs* that mediate carbon metabolism in the plant were found to be coexpressed [12]. Additionally, *BrOGOE* showed transcriptional inhibition of the arabidopsis sucrose synthase gene (*AtSUS*), implying a role in SUS-dependent carbon metabolism [11]. Collectively, the data suggest that OGs play a significant role in the regulation of carbon channeling in plants, and possibly also in growth and development, via interactions or coexpression with diverse regulatory networks.

OGs interact with evolutionarily conserved regulators in carbohydrate metabolism

To date, ten to 12 different OGs with distinct phenotypes in plants have been found (Table 1). In addition, *QQS* expression profiling has been accomplished in multiple species, such as tobacco [27], soybean, maize, and rice [28]. However, the underlying molecular mechanisms through

which *QQS* and various other OGs are involved in carbon metabolism remain unknown. The first hint about its possible mechanism of action comes from the interaction of QQS with *Solanum tuberosum* nuclear factor Y subunit C4 (*St*NF-YC4). NF-YC4 is a subunit of the heterotrimeric transcription factor NF-YA/NF-YB/NF-YC, which is highly conserved among eukaryotes, including plants and humans. Overexpression of *NF-YC4* was shown to increase protein contents, mimicking *QQS* phenotypes [29]. Expression profiling, promoter region analysis, and protein–protein interaction analysis collectively suggest that StNF-YC4 may play a role in multiple processes, including metabolism, stress responses, starch storage, and defense against pests and pathogenic invasion [30]. Promoter sequence analysis of *StNF-YC4* revealed the presence of a regulatory motif that resembles the *GATA-1* sequence. Furthermore, this motif is involved in the regulation of a gene encoding glyceraldehyde-3-phosphate dehydrogenase, which limits the regeneration of ribulose 1,5-bisphosphate (RuBP), a central CO_2 acceptor during photosynthesis [27]. Collectively, these data suggest that QQS interacts with NF-YC4 for carbohydrate metabolism in plants and possibly exerts a subsequent effect by altering the Calvin and Krebs cycles (Figure 1). Further investigations are needed to clarify whether these OGs mediate sugar metabolism by transcriptional reprogramming of different regulatory complexes.

The next apparent question is what other central regulators in plants govern carbohydrate metabolism and work in conjunction with OGs. Indeed, *Triticum aestivum* Fusarium Resistance Orphan Gene (*Ta*FROG) protein directly interacts with SnRK1 to regulate energy homeostasis and sugar metabolism in plants [17,31]. SnRK1, an evolutionarily conserved energy master regulator, has been shown to play an essential role in carbohydrate metabolism in plants [32]. Whether other orphan proteins also interact with SnRK1 is a question for future investigation. Interestingly, cellular metabolism in eukaryotes is coordinated through a double-negative feedback loop between SnRK1 and TOR [33], called the SnRK1–TOR axis (Figure 1). TOR is another evolutionarily conserved phosphatidylinositol 3-kinase-related protein kinase that acts at the core of signaling networks for nutrient and hormone sensing, regulating anabolism and catabolism by coordinating growth and various cellular metabolic processes [34]. As shown in Figure 1, the SnRK1–TOR axis is thought to be a central regulatory mechanism for energy metabolism, which is integrated with and mediated by other evolutionarily conserved kinase complexes such as S6 and YAKI kinases [35]. Under nutrient-deficient conditions, SnRK1 acts as a sensor for carbon deficiency and ensures the plant's survival through the repression of TOR [36]. It has yet to be determined how diverse sets of OGs function in the SnRK1–TOR axis and how SnRK1–TOR downstream signaling governs growth and developmental processes.

Molecular mechanisms of OG-mediated carbohydrate metabolism via the SnRK1–TOR axis and downstream signaling

The first line of evidence in this context is the existence of crosstalk between SnRK1–TOR signaling and brassinosteroid (BR)-biosynthetic genes as well as a molecular connection between an OG, GS9 (*Grain Shape Gene on Chromosome 9*), and the BR pathway. GS9 exhibits an antagonistic relationship with *OsOFP14* (*Oryza sativa* Ovate Family Protein 14) [37]. Transcriptional analyses suggest that GS9 potentially participates in spikelet development and modulates the grain shape and appearance in rice. *OsOFP14*, however, also regulates the fruit shape by antagonistically modulating the *GS9* activity [37]. Another OFP, *OsOFP8*, maintains the *OsOFP14*-mediated suppression of *GS9*. Intriguingly, *GS9* overexpression resulted in phenotypes similar to those of *OsOFP8*-RNAi lines [38]. It was also shown that OsOFP8 interacts with OsGSK2 (glycogen synthase kinase 2), a negative regulator of the BR signaling pathway. At the same time, BRs are essential for sugar-/carbon-induced growth response, and in turn, sugar is required to promote the TOR pathway as well as the stabilization of BZR1 (brassinazole-resistant 1). Additionally, the activation of TOR complex (TORC) signaling facilitates the response to the BR pathway by

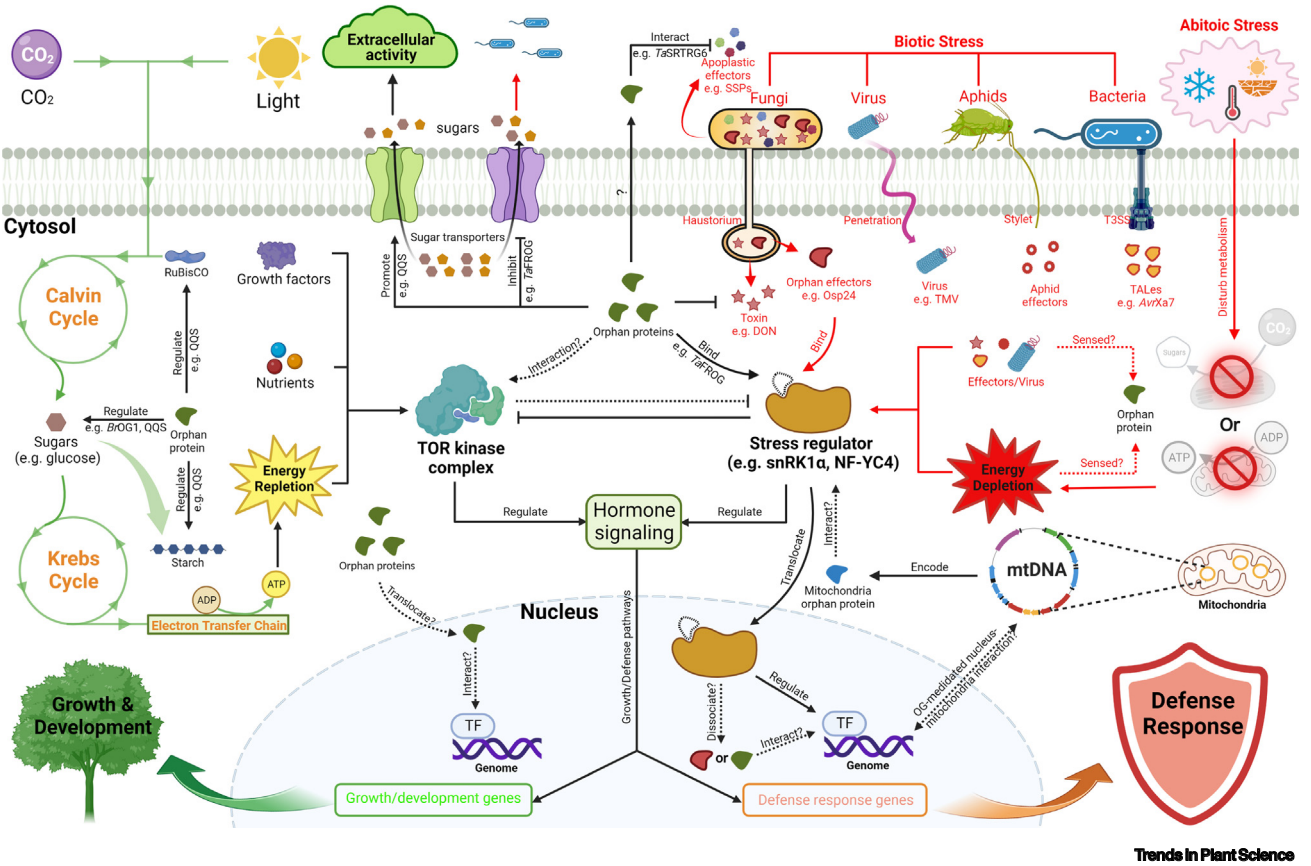

Figure 1. The multifarious roles of orphan genes (OGs) in growth and defense trade-offs. The dynamic balance of plant growth/development and defense response is shown as a green tree and one brown shield, respectively. Upon light activation, carbon from carbon dioxide (CO_2) is fixed into the Calvin cycle, catalyzed by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). Sugars produced by the Calvin cycle are further metabolized by the Krebs cycle and follow the electron transfer chain to produce ATP. The regulatory roles of orphan proteins on RuBisCO and sugar metabolites are illustrated. Energy repletion by ATP production, cooperating with growth factors and nutrients, later fuels the assembly of the target of rapamycin (TOR) kinase complex. A hypothetical interaction between orphan proteins and the TOR kinase complex is displayed. In the presence of biotic or abiotic stresses, host response regulators become activated. Biotic stressors hijack plants in a species-specific invasive way, including a fungal haustorium, virus penetration, aphid stylet, and bacterial type III secretion systems (T3SSs). Furthermore, one existing extracellular interaction between orphan proteins and fungus-released apoplastic effectors is illustrated. Abiotic stresses disrupt plant metabolic processes, such as sugar metabolism and ATP production, and consequently cause energy depletion. The confirmed interactions of stress regulators with exogenous factors are illustrated, such as toxins, orphan protein effectors, aphid effectors, and transcription activator-like effectors (TALEs). The putative direct sensing of these factors by orphan proteins is also indicated. Subsequentially, plant hormones couple with an active TOR kinase complex and stress regulators to regulate gene expression in a growth/development-dependent or defense-dependent manner, governing resource allocations between growth and defense. Additionally, the OG-binding stress regulator is shown to translocate into the nucleus, interacting with transcription factors (TFs) for gene regulation. The dissociation of orphan proteins from the orphan proteins–stress regulator complex and the independent transport of orphan proteins into the nucleus to act on TFs individually are also speculated. Moreover, the regulation of sugar transporters by orphan proteins to promote extracellular activities or prevent resource usurpation by bacteria is illustrated. Finally, mitochondria-originating orphan proteins and OG-mediated nucleus–mitochondria interactions are indicated. Broken lines indicate the putative interactions; black lines indicate regulation by plants; red lines indicate external disruption from biotic and abiotic stressors. Abbreviations: *Br*OGs, *Brassica rapa*-specific Orphan Genes; DON, deoxynivalenol; *QQS, Qua-Quine Starch*; *Ta*FROG, Triticum aestivum Fusarium Resistance Orphan Gene; TMV, tobacco mosaic virus. This figure was created using BioRender (https://biorender.com/).

stabilizing BZR1, potentially by inhibiting its autophagy-mediated degradation [39]. Furthermore, it has also been speculated that OsBZR1 (a positive regulator of the rice BR pathway) could bind the promoter region of the *OsOFP1* gene. Moreover, OsOFP1 physically interacts with DLT (DWARF and Low-Tillering) and OsGSK2, suggesting that BR may regulate OsOFP1 at both transcriptional and translational levels to control grain appearance. *OsOFP19*, another OFP family member, is thought to have a function in BR signaling and plant grain appearance [39,40]. Moreover, OsGSK2 also directly interacts with OsGRF4/GL2, both of which are involved in BR

signaling [41]. Taken together, it can be speculated that GS9 establishes a link between BZR1 accumulation and consequent activation of transcription factors of BR and SnRK1–TOR axis to maintain grain morphology.

Functional studies revealed that, in addition to SnRK1, another SNF Related Kinase, SnRK2, is directly involved in the TOR pathway through the phosphorylation of RAPTOR1 (Regulatory-Associated Protein of TOR 1) and inhibition of TORC signaling [42]. Given that SnRK2 interacts with the TOR signaling pathway, it may also be engaged in OG-mediated functions. In addition to the SnRKs, polyamines have also been implicated in growth and development through TOR signaling [43]. Spermidine (Spd), a prominent polyamine present in animals and plants, has recently been demonstrated to play a crucial role in the growth and development of maize and arabidopsis seedlings via TOR signaling [44]. Polyamines are considered a new kind of biostimulants and play an important role during plant growth and development. Interestingly, a pair of OGs, *OsPHT3* (*Putrescine Hydroxycinnamoyl Transferase 3*), and *OsPHT4* (*Putrescine Hydroxycinnamoyl Transferase 4*) are members of the monocot-specific functional hydroxycinnamoyl putrescine (HP) gene cluster and have been linked to rice growth and development [45]. Based on the existing knowledge, we hypothesize that these OGs may be involved in TOR signaling and growth-related regulatory activities.

Orphan proteins utilize SnRK–TOR signaling to regulate growth-defense trade-offs and pathogen resistance in plants

The growth–defense trade-off is a crucial aspect of organismal homeostasis across kingdoms [46]. Plants, like all other organisms, must allocate resources towards both growth and defense to ensure survival and reproduction [47]. In an agricultural context, there is a well-known risk of losing valuable genetic traits for biotic defenses if crops are bred primarily for growth-related traits. Understanding the molecular mechanisms underlying the growth–defense trade-offs could be important for future crop-breeding strategies aimed at developing high-yielding crops that can also resist biotic stressors. For instance, the OG *TaFROG* encodes an SnRK1-interacting protein and confers disease resistance against a trichothecene mycotoxin deoxynivalenol (DON) [17]. DON is a mycotoxin produced by fungal pathogens *Fusarium graminearum* and *Fusarium culmorum* [48,49]. Interestingly, *F. graminearum* encodes an orphan protein, Osp24, which acts as an SnRK1-dependent susceptibility factor in rice and promotes Fusarium head blight (FHB) disease. During infection, Osp24 interacts with SnRK1 and facilitates its degradation through the proteasome pathway to support fungal pathogenesis. Intriguingly, the TaFROG protein competes with Osp24 for binding with SnRK1, consequently protecting it from degradation (Figure 1) and thereby conferring disease resistance against FHB [10].

While the SnRK–TOR axis signaling components emerged mostly as regulators of growth–defense trade-offs, SnRKs and TOR act as positive and negative regulators of plant disease resistance, respectively. In the case of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*)-mediated infection, SnRK1A regulates plant immunity possibly through the host immune receptor XA21 (*Xanthomonas* resistance 21), whereas TOR suppresses pattern-triggered immunity (PTI). For instance, the OG *Xio1* (*Xoo-induced orphan 1*) promotes resistance against *Xoo* in an immune receptor XA21-dependent manner [15]. However, as a result of growth-to-defense trade-offs, *Xio*-OE lines showed significant developmental retardation. Taken together, it is plausible that specific OGs regulate such trade-off mechanisms via the evolutionarily conserved SnRK–TOR signaling pathway. Transcriptome analysis of OGs suggested that they may induce a similar pattern of immune responses in multiple species [50]. Functional studies also revealed that immunity-related genes exhibit pathogen-specific expression patterns across multiple species, implying a rapid coevolution scenario for OGs and their cognate pathogens [17,51,52]. Future research will focus on whether

these OGs coevolved with their specific pathogen-responsive traits, and if they share any conserved evolutionary history with their pathogen virulence factors (termed effectors). Evolutionary studies in rice revealed that *Xio1* has exclusively evolved in the AA genome. Despite the AA and BB genomes being distinguished using evidence derived from cytogenetics, interspecific crossing, and genomic DNA hybridization, the BB genome, which is phylogenetically closest to the AA genome, does not contain any evidence of this OG [15,53]. Collectively, these findings raise the possibility that *Xio1* is a newly generated OG and that it was created in response to strong selective pressure, which is not specific to the ancient source of gene-for-gene XA21/*Xoo* (raxX-sY) interaction. This coevolution hypothesis is supported by two more OGs, *TaSRTRG6* (*Septoria-responsive taxonomically restricted gene 6*) and *TaSRTRG7* in wheat. Both TaSRTRG6 and TaSRTRG7 interact with small fungal effector candidates secreted by *Zymoseptoria tritici* and enhance the resistance against *Septoria tritici* blotch disease. The proposed coevolution of wheat with *Z. tritici* is further supported by the evidence emerging from the fully sequenced wheat genome [54]. All in all, it is conceivable that OGs and their evolutionarily conserved immunogenic responses are related. They could plausibly modulate the plant immune system via sugar metabolism and possibly interact with major central regulators (i.e., SnRK1/TOR) (Figure 1). Future research on OG's functional characterization in relation to plant immunity will help scientists better grasp their intricate relationships. In addition, as we explore the functions of OGs, there are a number of unanswered questions (see Outstanding questions) that require further research.

OGs in 'sweet immunity' and hormonal-mediated plant defenses

Given the importance of OGs in several growth-related and carbohydrate metabolism processes (Figure 1), one question is whether OG-mediated sugar signals are also involved in the process of 'sweet immunity' in plants [55–57]. Sweet immunity entails the production of specific sugar-based compounds that trigger signaling pathways leading to the inhibition of pathogen growth and infection [58]. Indeed, an orphan protein Xa7 (*Xanthomonas* resistance 7), with no known structures in the Pfam database, provides resistance to rice against *Xoo*, a bacterial blight-inducing rice pathogen [18]. A recent study shed light on the coevolution of this particular OG in the context of pathogen effectors and their host targets in the rice–*Xoo* pathosystem [52]. In particular, to establish effector-triggered susceptibility (ETS), *Xoo* secretes and delivers transcription activator-like effectors (TALEs) including AvrXa7 (Avirulence Xa7) and PthXo3 (Pathogenicity Xo3). These TALEs recognize effector binding elements (EBEs) in the promoter sequence of host target genes such as *SWEET14* (*Sugars Will Eventually Be Exported Transporter 14*), which encodes a sucrose efflux transporter and is a major susceptibility gene [52]. Upon infection with *Xoo, SWEET14* is activated, leading to the export of apoplastic sugar, which promotes the *Xoo* pathogenesis. Since the EBE of *Xa7* is a close mimic of the EBE of *SWEET14*, the evolution of this OG guards the activation of the *SWEET14* gene. Thus, Xa7 triggers effector-triggered immunity (ETI) in response to *Xoo* carrying AvrXa7 or PthXo3 TALEs [52] and overcomes *SWEET14*-mediated ETS. Among other questions yet to be determined is what the underlying molecular mechanisms of Xa7 in immunogenic signal transduction in plants are. Do other OGs also participate in sweet immunity, since a vast majority of them are implicated in carbohydrate metabolism? How do *Xa7* and other OGs crosstalk with hormonal signal transduction pathways?

Evidence for the participation of OGs in the hormonal signaling pathway comes from the functional studies of *OsDR10* (*Oryza sativa defense-responsive gene 10*), a *de novo*-originated rice-specific OG, that is also involved in the regulation of *Xoo*-mediated bacterial blight [59]. RNAi-mediated silencing of *OsDR10* suggested that this OG blocks the accumulation of endogenous salicylic acid (SA), which is a phytohormone predominantly involved in defense signaling in plants. Furthermore, overexpression of *OsDR10* causes an increase in jasmonic acid (JA) accumulation and a decrease in the expression of various *R* genes, hence promoting

disease susceptibility in rice [59]. Another functional study suggested an antagonistic relationship between SA and JA. Specifically, endogenous SA accumulation prevents the conversion of 13S-hydroperoxylinolenic acid to 12-oxo-phytodienoic acid, inhibiting the JA-dependent defensive response in plants [60]. Future research will need to determine whether *Os*DR10 is a target of pathogen effectors for the manipulation of hormonal signaling pathways and the establishment of ETS. It is also known that effector molecules surpass distinct layers of host defenses (ETI and PTI), possibly by targeting the central hormonal signaling hubs [61,62]. Future investigations can provide insights into how *OsDR10* and other OGs integrate hormonal signaling with carbohydrate metabolism and specifically the SnRK–TOR axis to modulate the immune response. For instance, QQS, a well-characterized arabidopsis orphan protein, has also been shown to interact with NF-YC4 (discussed earlier), and is involved in both nutrient regulatory pathways and immune responses against viruses, bacteria, and aphid infection in arabidopsis and soybean [8].

Phytohormone pathways and potential roles of OGs in abiotic stress signaling

Functional studies suggest that OGs are crucial for the production of secondary metabolites and phytohormones [29]. Secondary metabolites are involved in diverse physiological responses in plants and ensure the plant's survival under various environmental stresses. Consistently, expression analysis in *Citrus sinensis* indicated the importance of nine different OGs during abiotic stress response [19]. In another study pertinent to a drought-tolerant *Coffea arabica* cultivar, IAPAR59, several OGs were found to be differentially expressed and may participate in the abscisic acid (ABA) signaling pathway [14], as well as other stress-responsive signaling transduction pathways [63]. Additionally, the transcriptomic analysis of a moss, *Physcomitrium* (*Physcomitrella*) *patens*-specific OG, *PpARDT* (*ABA-responsive drought tolerance*), indicated its potential role in drought tolerance possibly by regulating the ABA signaling pathway [9]. Functional studies showed that ABA also mediates the SnRK1 activity and triggers transcriptional regulation and energy metabolism in plants [64]. It will be fascinating to learn whether *Ta*FROG plays any role in modulating drought stress in an ABA-dependent manner, given that SnRK1 interacts with *Ta*FROG in wheat, as previously discussed. Furthermore, ABA-dependent RAPTOR phosphorylation is thought to be a common way for AMP-activated protein kinase (AMPK) and SnRK1 to decrease TOR activity under energy-depleted situations. The AMPK pathway, conversely, functions as a master regulator of cellular energy homeostasis [65]. Although TOR and SnRKs work antagonistically, it can be speculated that the SnRK–TOR axis mediated by OGs may have a potential role in drought stress response (Figure 1). Moreover, functional studies showed that the BR signaling pathway is coupled to the ABA signaling pathway in a BIN2 (brassinosteroid-insensitive 2) kinase-dependent manner. Inhibiting the BIN2 kinase limits the BR pathway and the interlinked ABA pathway during stress response [66]. So, it may be speculated that BR pathway-specific OGs also may have a potential role in abiotic stress signaling.

Evolution of OGs

Gene structural investigations have revealed that OGs have much shorter protein lengths than nonorphan genes (non-OGs), indicating minimal numbers of exons [67]. Although OGs in plants typically do not exhibit sequence similarities, they possess functional characteristics similar to non-OGs, implying that they share similar evolutionary origins. It remains to be determined what connection there is between adaptive evolution and the emergence of novel agronomic features. For instance, *Brassica rapa* and arabidopsis diverged from one another only about 12.4–13.4 million years ago, yet as a result of selective breeding, some *B. rapa* cultivars display extreme morphological traits [68]. One prominent example of this are the leafy heads of Chinese cabbage, which are phenotypically contrasting with arabidopsis [69]. Since QQS can interact with NF-YC4 homologs across species, including human NF-YC4, it is reasonable to infer that OGs

might have evolved along with their interacting partners or central regulators/factors for their proposed functions, including the modulation of morphological traits.

Apart from the recruitment and *de novo* origination of TEs, mitochondrial fostering also plays a major role in the birth of new OGs (Figure 1) [70]. In a nutshell, the 'mitochondrial fostering' theory posits that a high rate of genomic rearrangement in the mitochondrial genome leads to the emergence of novel sequences that are subsequently inserted into a nuclear chromosome, resulting in the possible emergence of OGs. This process may involve the transfer of mitochondrial DNA, which could either contain gene-coding information or obtain an open reading frame through transposition or other genomic mechanisms already discussed [70]. Given the considerable amount of mitochondrial DNA transferred to the nucleus and the unique evolutionary traits of the mitochondrial genome, this theory offers a compelling explanation for the evolution of OGs. Moreover, in the rice genome, the recruitment of TEs is thought to result in the origination of new OGs [71]. Similarly to TEs, it can also be postulated that, through evolution, the OGs' coding sequences may have been evenly shuffled within the genome over time to interact with certain transcription factors, receptors, and other regulatory networks such as NF-YC and TOR for control of gene expression and transcriptional reprogramming. As discussed above, TaFROG-

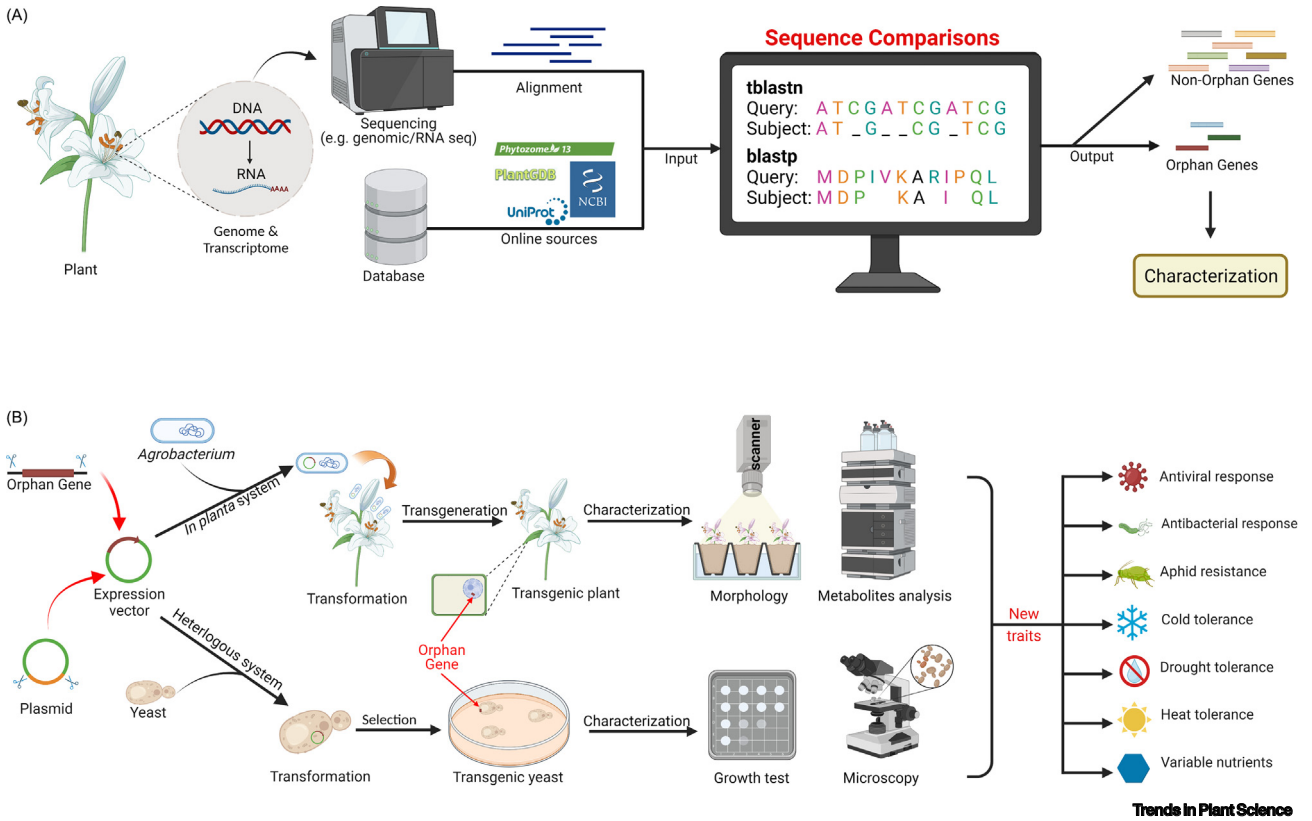

Figure 2. Identification and characterization of plant orphan genes (OGs). (A) Schematics of OGs identification. Plant genomic or transcriptomic information is obtained via sequencing and alignment. Comparisons of sequencing results in the online database further categorize OGs and non-OGs. Candidate OGs are used for characterization. (B) General workflow of OGs characterization via *in planta* and heterogeneous systems. OGs are recombined into expression vectors designated for *in planta* or heterogeneous systems. In an *in planta* overexpression system, *Agrobacterium*-mediated OG transformation into the plant genome is shown. Transgenic plants possessing stable OG expression are designated for downstream analysis, such as morphological study and metabolite analysis. In a heterogeneous system, recombinant vectors are introduced into yeast strains that can be isolated on a selective medium. Techniques including growth tests and microscopy are utilized to characterize OG-overexpressing yeast with improved traits against environmental stresses (e.g., heat, cold, etc.>). This figure was created using BioRender (https://biorender.com/).

SnRK1 interaction modulates energy metabolism, whereas TaFROG-NAC-like protein interaction enhances resistance to Fusarium head blight disease [17].

Role of tissue-specific expression of OGs during developmental processes in plants

Gene cluster analysis reveals that OGs primarily show tissue-specific expression for growth and developmental processes. Expression analyses showed TE-assisted lineage/tissue-specific formation of hydroxycinnamoyl tyramine (HT) gene cluster in rice. For instance, *Ms1* (*Poaceae-specific male sterility 1*) encodes a phospholipid-binding protein that is considered necessary for male fertility and metagenesis, as well as pollen exine development [72]. Male sterility was observed in wheat, barley, and *Brachypodium* plants lacking *Ms2*, another OG that displays anther-specific expression attributed to an insertion of a retroelement into the promoter [16]. While we propose a pipeline for OG cross-species transfer (Figure 2), the cloning of this OG (along with others listed in Table 1) opens new avenues in the development of transgenic plants with desirable phenotypic traits.

Moreover, OGs were found to be much more highly expressed in the flower compared to other tissues of the family Cucurbitaceae [24]. Overexpression of a rice-specific OG *GN2* (*grains number 2*) resulted in lower grain numbers and reduced plant height as compared to control plants [73], implying that the expression of the OGs may be precisely regulated, both spatially and temporally, within the genome/gene cluster. Furthermore, it might be advantageous for modulating different regulatory pathways which are then related to specific phenotypes.

Besides plants, OGs are also potentially involved in major biological processes in other domains of life. For instance, functional analysis in the diamondback moth *Plutella xylostella* revealed that OGs *Tssor-3* and *Tssor-4* showed specific expression patterns in male gonads, indicating their significance in male fitness during evolution [74]. The reproductive system-specific expression of an OG might indicate a genetic sweep out of emerging gene fragments during gene duplication and divergence in their evolutionary course. Additional studies conducted in basal metazoans have further emphasized the functions of these OGs. Evidence from Hydra suggests that genes with restricted taxonomy help to create novelties that are relevant only to the phylum Cnidaria [75]. Here, an evolutionary paradox arises due to the great degree of conservation of signal transduction pathways, and regulatory evolution provides a likely solution to this puzzle. Since regulatory genes are extensively distributed throughout the animal kingdom, the differential utilization of naturally occurring elements may help explain morphological differences between species [51]. Given their significance, OGs are acknowledged as the source of new proteins and are ideal research subjects because of their enormous potential to shed light on the processes that give rise to the structural domains of proteins.

Concluding remarks

The origin of unique OGs remains an enigma. However, detailed computational analyses combined with machine-learning approaches may shed light on the evolution of OGs. In addition to *de novo* origination, duplication, and divergence hypotheses, epigenetic changes may play a key role in the formation of OGs. In the early stages of their formation, the epigenetic landscape of *de novo* OGs may vary significantly among and within populations. The regulatory regions of OGs exhibit epigenetic signatures that are characteristic of enhancers, in contrast to non-OGs that feature classical promoters. Furthermore, some nonexpressed OGs display repressive histone modifications, leading to decreased transcription [76]. It can also be hypothesized that epigenetic remodeling may facilitate the creation of novel OGs. Furthermore, by studying the evolution of OGs, researchers can gain insights into their origins and how they diverge within and between species over extended periods. Such analyses may reveal the mechanisms underlying the

Outstanding questions

What nuclear and cytoplasmic factors are involved in the emergence, turnover, and fixation of OGs? Does epigenetic remodeling have any significant role in the origination of OGs?

How have OGs acquired such a precise gene structure and transcriptional machinery to exert proper temporal and spatial expression? Do OGs act as transcriptional factors directly or act as mediators/regulatory factors?

Do OGs lead to speciation or vice versa? Do OGs get maintained within species via selective sweeps (cross-species transfer using transgenic approaches)?

Since *Ta*FROG and Osp24 have been reported to interact with same central regulatory players, how do plants and their pathogens evolve OGs with similar functions? Do such OGs evolve independently or through horizontal gene transfer?

How can we combine traditional and newly developed methods (e.g., SMOTE and ENN-based XGBoost, Synthetic Minority Oversampling Technique, Edited Nearest Neighbors, and eXtreme Gradient Boosting, the ensemble learning algorithm for strong predictability of data) for the identification and screening of novel OGs?

How do OGs integrate into the main cellular signaling pathway to mediate growth–defense trade-offs? As such, OGs are involved in the BR pathway for multiple regulatory responses; do these OGs have any significant interacting role with the TOR signaling pathway during the temporal and spatial course of development?

Do OGs (e.g., *QQS, BrOGs*) specifically control sugar transporters to reduce the loss of resources and exert defense? If so, does this mechanism contribute to the homeostasis of the defense–growth trade-offs?

Given that *Ta*SRTRG6 and *Ta*SRTRG7 are predicted secreted proteins that physically interact with small secreted proteins of *Zymoseptoria tritici*, possibly in the apoplast, do other plant-secreted orphan proteins interact with apoplastic effectors released by fungi/oomycetes?

Box 1. High-throughput identification and characterization of OGs

The fundamental principles governing OGsis and functional recruitment gleaned from a model plant species such as arabidopsis will expand our understanding of the involvement of orphans in other species. Based on the known functional studies, it is reasonable to hypothesize that an OG from an agronomically important crop may be ectopically transferable to arabidopsis. This may provide a biotechnological platform for the functional and molecular characterization of particular OGs. However, the central question is how we can streamline the functional characterization of OGs that can account for up to 30% of a genome. Towards this, we propose a high-throughput pipeline that entails a large-scale study for the identification and functional characterization of OGs. In step 1, publicly available genomic and transcriptomic data can be utilized for cross-species comparisons to identify OGs (see Figure 2A in main text). Similar to a recent large-scale experiment, which successfully transformed 43 *Brassica*-specific OGs (*Br*OGs) into arabidopsis followed by functional characterization [12], step 2 involves an Agrobacterium system for cross-species transfers of OGs (see Figure 2B in main text). Robotics and artificial intelligence-based phenomics software may be used to screen the compendium of large-scale transgenics for a variety of morphological, physiological, and immunological features, including plant growth, development, and tolerance to various biotic and abiotic stimuli. Using the CRISPR/Cas system, the selected OGs from the model system can further be investigated in the native crop for detailed functional and molecular characterization. Owing to the conserved roles of OGs in diverse organisms, as an alternative to arabidopsis, a yeast-based heterologous system [77–79] can be employed for the overexpression of OGs (see Figure 2B in main text). In this automated and high-throughput platform, yeast cellular morphology and growth trajectories may be utilized as output to characterize the function of a particular OG in response to various environmental challenges, including cold, heat, and chemical-based cellular stresses.

How are OGs involved their pleiotropic functions? The known OGs converge on the TOR–SnRK axis and other evolutionarily conserved functions. From a network biology's perspective, do they operate as hubs?

formation and evolution of OGs and elucidate their functional roles in different organisms. Although OGs in plants generally show no sequence similarities, they share similar functional characteristics with non-OGs, implying that they may function together. OGs play an essential role in sugar metabolism and carbon channeling, biotic and abiotic stress responses, and multiple lineage-specific adaptive processes. Studies on OGs are receiving a lot of interest, although the research field is currently limited to a small number of species, and it is still unknown how broadly OGs act. Even though the OG-mediated alterations of carbon metabolism produce concurrent biotic stress, the downstream morphological and physiological responses vary among different plant species. Additionally, OGs are widely distributed throughout the kingdoms of life, but a fundamental barrier to determining their function is the absence of functional motifs and folds of interdependent components or effects that are still discernible. In this study, we highlighted the possible involvement of the SnRK–TOR axis and BR signaling in the function of OGs (Figure 1), but additional research will be required to confirm any potential interactions relevant to OGs' functionality (Figure 2). Towards this, a high-throughput method for OG detection and characterization needs to be developed (Figure 2 and Box 1). Various gene-editing technologies such as the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein (CRISPR/Cas) system can be helpful for gene profiling and mutant screening, allowing for the functional characterization of OGs in a specific species. Alternatively, the functions of OGs can be analyzed in great detail using a yeast heterologous system. Finally, as encoders of novel proteins, OGs have the potential to disrupt evolution or act as a catalyst for speciation.

Acknowledgments

This work was supported by NSF award IOS-2038872 to M.S.M. and to K.M.P-M. The title 'ORFan's Tale' was inspired by a novel by Pam Jenoff. We apologize for not citing the original references due to a lack of space.

Declaration of interests

No interests are declared.

References

1. Tautz, D. and Domazet-Lošo, T. (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12, 692–702
2. Singh, U. and Wurtele, E.S. (2020) Genetic novelty: how new genes are born. *eLife* 9, e55136
3. Khalturin, K. *et al.* (2009) More than just orphans: are taxonomically restricted genes important in evolution? *Trends Genet.* 25, 404–413
4. Toll-Riera, M. *et al.* (2009) Origin of primate orphan genes: a comparative genomics approach. *Mol. Biol. Evol.* 26, 603–612
5. Donoghue, M.T. *et al.* (2011) Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana*. *BMC Evol. Biol.* 11, 47
6. Van Oss, S.B. and Carvunis, A.-R. (2019) *De novo* gene birth. *PLoS Genet.* 15, e1008160

7. Albà, M.M. and Castresana, J. (2007) On homology searches by protein Blast and the characterization of the age of genes. *BMC Evol. Biol.* 7, 53
8. Qi, M. *et al.* (2019) QQS orphan gene and its interactor NF-YC4 reduce susceptibility to pathogens and pests. *Plant Biotechnol. J.* 17, 252–263
9. Dong, X.-M. *et al.* (2022) Orphan gene PpARDT positively involved in drought tolerance potentially by enhancing ABA response in *Physcomitrium* (*Physcomitrella*) *patens*. *Plant Sci.* 319, 111222
10. Jiang, C. *et al.* (2020) An orphan protein of *Fusarium graminearum* modulates host immunity by mediating proteasomal degradation of TaSnRK1α. *Nat. Commun.* 11, 4382
11. Jiang, M. *et al.* (2018) Mining of brassica-specific genes (BSGs) and their induction in different developmental stages and under *Plasmodiophora brassicae* stress in *Brassica rapa*. *Int. J. Mol. Sci.* 19, 2064
12. Jiang, M. *et al.* (2020) *Brassica rapa* orphan genes largely affect soluble sugar metabolism. *Hortic. Res.* 7, 181
13. Li, G. *et al.* (2019) Orphan genes are involved in drought adaptations and ecoclimatic-oriented selections in domesticated cowpea. *J. Exp. Bot.* 70, 3101–3110
14. Mofatto, L.S. *et al.* (2016) Identification of candidate genes for drought tolerance in coffee by high-throughput sequencing in the shoot apex of different *Coffea arabica* cultivars. *BMC Plant Biol.* 16, 94
15. Moon, H. *et al.* (2022) Oryza-specific orphan protein triggers enhanced resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Front. Plant Sci.* 13, 859375
16. Ni, F. *et al.* (2017) Wheat Ms2 encodes for an orphan protein that confers male sterility in grass species. *Nat. Commun.* 8, 15121
17. Perochon, A. *et al.* (2015) TaFROG encodes a Pooideae orphan protein that interacts with SnRK1 and enhances resistance to the mycotoxigenic fungus *Fusarium graminearum*. *Plant Physiol.* 169, 2895–2906
18. Wang, C. *et al.* (2021) Xa7, a small orphan gene harboring promoter trap for AvrXa7, leads to the durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Rice* 14, 48
19. Xu, Y. *et al.* (2015) Identification, characterization and expression analysis of lineage-specific genes within sweet orange (*Citrus sinensis*). *BMC Genomics* 16, 995
20. Zhang, W. *et al.* (2019) Origination and evolution of orphan genes and *de novo* genes in the genome of *Caenorhabditis elegans*. *Sci. China Life Sci.* 62, 579–593
21. Guo, W.J. *et al.* (2007) Significant comparative characteristics between orphan and nonorphan genes in the rice (*Oryza sativa* L.) genome. *Comp. Funct. Genom.* 2007, 21676
22. Mishra, B. *et al.* (2021) Network biology to uncover functional and structural properties of the plant immune system. *Curr. Opin. Plant Biol.* 62, 102057
23. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30
24. Ma, D. *et al.* (2022) Identification, characterization and function of orphan genes among the current Cucurbitaceae genomes. *Front. Plant Sci.* 13, 872137
25. Clément, C. and Audran, J. (1995) Anther wall layers control pollen sugar nutrition in *Lilium*. *Protoplasma* 187, 172–181
26. Arendsee, Z.W. *et al.* (2014) Coming of age: orphan genes in plants. *Trends Plant Sci.* 19, 698–708
27. Tanvir, R. *et al.* (2022) AtQQS orphan gene and NtNF-YC4 boost protein accumulation and pest resistance in tobacco (*Nicotiana tabacum*). *Plant Sci.* 317, 111198
28. Li, L. and Wurtele, E.S. (2015) The QQS orphan gene of *Arabidopsis* modulates carbon and nitrogen allocation in soybean. *Plant Biotechnol. J.* 13, 177–187
29. Li, L. *et al.* (2015) QQS orphan gene regulates carbon and nitrogen partitioning across species via NF–YC interactions. *Proc. Natl. Acad. Sci.* 112, 14734–14739
30. Tanvir, R. *et al.* (2022) Orphan genes in crop improvement: enhancing potato tuber protein without impacting yield. *Plants* 11, 3076
31. Hulsmans, S. *et al.* (2016) The SnRK1 energy sensor in plant biotic interactions. *Trends Plant Sci.* 21, 648–661
32. Polge, C. and Thomas, M. (2007) SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? *Trends Plant Sci.* 12, 20–28
33. Jamsheer, K.M. *et al.* (2022) A negative feedback loop of TOR signaling balances growth and stress-response trade-offs in plants. *Cell Rep.* 39, 110631
34. Raught, B. *et al.* (2001) The target of rapamycin (TOR) proteins. *Proc. Natl. Acad. Sci.* 98, 7037–7044
35. Jamsheer, K. *et al.* (2019) Evolution of TOR–SnRK dynamics in green plants and its integration with phytohormone signaling networks. *J. Exp. Bot.* 70, 2239–2259
36. Baena-González, E. and Hanson, J. (2017) Shaping plant development through the SnRK1–TOR metabolic regulators. *Curr. Opin. Plant Biol.* 35, 152–157
37. Zhao, D.-S. *et al.* (2018) GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat. Commun.* 9, 1240
38. Yang, C. *et al.* (2016) OVATE family protein 8 positively mediates brassinosteroid signaling through interacting with the GSK3-like kinase in rice. *PLoS Genet.* 12, e1006118
39. Zhang, Z. *et al.* (2016) TOR signaling promotes accumulation of BZR1 to balance growth with carbon availability in *Arabidopsis*. *Curr. Biol.* 26, 1854–1860
40. Yang, C. *et al.* (2018) Os OFP 19 modulates plant architecture by integrating the cell division pattern and brassinosteroid signaling. *Plant J.* 93, 489–501
41. Che, R. *et al.* (2015) Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat. Plants* 2, 15195
42. Rosenberger, C.L. and Chen, J. (2018) To grow or not to grow: TOR and SnRK2 coordinate growth and stress response in *Arabidopsis*. *Mol. Cell* 69, 3–4
43. Ren, M. *et al.* (2012) Target of rapamycin signaling regulates metabolism, growth, and life span in *Arabidopsis*. *Plant Cell* 24, 4850–4874
44. Salazar-Díaz, K. *et al.* (2021) TOR senses and regulates spermidine metabolism during seedling establishment and growth in maize and *Arabidopsis*. *Iscience* 24, 103260
45. Fang, H. *et al.* (2022) Function of hydroxycinnamoyl transferases for the biosynthesis of phenolamides in rice resistance to *Magnaporthe oryzae*. *J. Genet. Genom.* 49, 776–786
46. Davis, J.E. *et al.* (2021) Toward a universal theoretical framework to understand robustness and resilience: from cells to systems. *Front. Ecol. Evol.* 8, 579098
47. Pajerowska-Mukhtar, K.M. *et al.* (2012) The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* 22, 103–112
48. Bai, G.-H. *et al.* (2002) Deoxynivalenol-nonproducing *Fusarium graminearum* causes initial infection, but does not cause disease spread in wheat spikes. *Mycopathologia* 153, 91–98
49. Snijders, C. (2004) Resistance in wheat to *Fusarium* infection and trichothecene formation. *Toxicol. Lett.* 153, 37–46
50. Entwistle, S. *et al.* (2019) Orphan genes shared by pathogenic genomes are more associated with bacterial pathogenicity. *Msystems* 4, e00290-00218
51. Johnson, B.R. (2018) Taxonomically restricted genes are fundamental to biology and evolution. *Front. Genet.* 9, 407
52. Luo, D. *et al.* (2021) The Xa7 resistance gene guards the rice susceptibility gene SWEET14 against exploitation by the bacterial blight pathogen. *Plant Commun.* 2, 100164
53. Kim, H. *et al.* (2007) Comparative physical mapping between *Oryza sativa* (AA genome type) and *O. punctata* (BB genome type). *Genetics* 176, 379–390
54. Brennan, C.J. *et al.* (2020) Taxonomically restricted wheat genes interact with small secreted fungal proteins and enhance resistance to *Septoria tritici* blotch disease. *Front. Plant Sci.* 11, 433
55. Jeandet, P. *et al.* (2022) The role of sugars in plant responses to stress and their regulatory function during development. *Int. J. Mol. Sci.* 23, 5161
56. Bolouri Moghaddam, M.R. and Van den Ende, W. (2013) Sweet immunity in the plant circadian regulatory network. *J. Exp. Bot.* 64, 1439–1449
57. Morkunas, I. and Ratajczak, L. (2014) The role of sugar signaling in plant defense responses against fungal pathogens. *Acta Physiol. Plant.* 36, 1607–1619
58. Yamada, K. *et al.* (2016) Regulation of sugar transporter activity for antibacterial defense in *Arabidopsis*. *Science* 354, 1427–1430

59. Xiao, W. *et al.* (2009) A rice gene of *de novo* origin negatively regulates pathogen-induced defense response. *PLoS One* 4, e4603
60. Doares, S.H. *et al.* (1995) Salicylic acid inhibits synthesis of proteinase inhibitors in tomato leaves induced by systemin and jasmonic acid. *Plant Physiol.* 108, 1741–1746
61. Mukhtar, M.S. *et al.* (2016) Pathogen tactics to manipulate plant cell death. *Curr. Biol.* 26, R608–R619
62. Liu, J. *et al.* (2022) A TIReless battle: TIR domains in plant–pathogen interactions. *Trends Plant Sci.* 27, 426–429
63. Muhammad Aslam, M. *et al.* (2022) Mechanisms of abscisic acid-mediated drought stress responses in plants. *Int. J. Mol. Sci.* 23, 1084
64. Rodrigues, A. *et al.* (2013) ABI1 and PP2CA phosphatases are negative regulators of Snf1-related protein kinase1 signaling in *Arabidopsis*. *Plant Cell* 25, 3871–3884
65. Nukarinen, E. *et al.* (2016) Quantitative phosphoproteomics reveals the role of the AMPK plant ortholog SnRK1 as a metabolic master regulator under energy deprivation. *Sci. Rep.* 6, 31697
66. Bulgakov, V.P. and Avramenko, T.V. (2020) Linking brassinosteroid and ABA signaling in the context of stress acclimation. *Int. J. Mol. Sci.* 21, 5108
67. Heames, B. *et al.* (2020) A continuum of evolving *de novo* genes drives protein-coding novelty in *Drosophila*. *J. Mol. Evol.* 88, 382–398
68. Yang, T.-J. *et al.* (2006) Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*. *Plant Cell* 18, 1339–1347
69. Jiang, M. *et al.* (2022) Research advances and prospects of orphan genes in plants. *Front. Plant Sci.* 13, 947129
70. O'Conner, S. and Li, L. (2020) Mitochondrial fostering: the mitochondrial genome may play a role in plant orphan gene evolution. *Front. Plant Sci.* 11, 600117
71. Jin, G.H. *et al.* (2021) Genetic innovations: transposable element recruitment and *de novo* formation lead to the birth of orphan genes in the rice genome. *J. Syst. Evol.* 59, 341–351
72. Wang, Z. *et al.* (2017) Poaceae-specific MS1 encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl. Acad. Sci.* 114, 12614–12619
73. Chen, H. *et al.* (2017) Emergence of a novel chimeric gene underlying grain number in rice. *Genetics* 205, 993–1002
74. Li, T.-P. *et al.* (2021) Functional analysis of the orphan genes Tssor-3 and Tssor-4 in male *Plutella xylostella*. *J. Integr. Agric.* 20, 1880–1888
75. Milde, S. *et al.* (2009) Characterization of taxonomically restricted genes in a phylum-restricted cell type. *Genome Biol.* 10, R8
76. Werner, M.S. *et al.* (2018) Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Res.* 28, 1675–1687
77. Barberini, M.L. and Muschietti, J.P. (2017) Expression of plant receptor kinases in yeast. *Methods Mol. Biol.* 1621, 21–27
78. Locascio, A. *et al.* (2019) *Saccharomyces cerevisiae* as a tool to investigate plant potassium and sodium transporters. *Int. J. Mol. Sci.* 20, 2133
79. Mehlmer, N. *et al.* (2009) Functional complementation of yeast mutants to study plant signalling pathways. *Methods Mol. Biol.* 479, 235–245