Gene gain and loss are crucial factors that shape the evolutionary success of diverse organisms. In the past two decades, more attention has been paid to the significance of gene gain through gene duplication or de novo genes. However, gene loss through natural loss-of-function (LoF) mutations, which is prevalent in the genomes of diverse organisms, has been largely ignored. With the development of sequencing techniques, many genomes have been sequenced across diverse species and can be used to study the evolutionary patterns of gene loss. In this review, we summarize recent advances in research on various aspects of LoF mutations, including their identification, evolutionary dynamics in natural populations, and functional effects. In particular, we discuss how LoF mutations can provide insights into the minimum gene set (or the essential gene set) of an organism. Furthermore, we emphasize their potential impact on adaptation. At the genome level, although most LoF mutations are neutral or deleterious, at least some of them are under positive selection and may contribute to biodiversity and adaptation. Overall, we highlight the importance of natural LoF mutations as a robust framework for understanding biological questions in general.

Key words: adaptive evolution, biodiversity, essential genes, loss-of-function, natural variation

Xu Y.-C. and Guo Y.-L. (2020). Less Is More, Natural Loss-of-Function Mutation Is a Strategy for Adaptation. Plant Comm. 1, 100103.
we emphasize the importance of LoF mutations for adaptive evolution in natural populations.

**GENETIC VARIATIONS THAT CAUSE LoF MUTATIONS**

There are four major genetic variations that can lead to LoF mutations (Figure 1A). First, a nonsense SNP may lead to a premature stop codon, producing a truncated protein sequence. For example, in Arabidopsis, FRIGIDA (FRI) alleles with premature stop codons explain a large fraction of flowering-time variation (Le Corre et al., 2002; Lempe et al., 2005; Shindo et al., 2005; Werner et al., 2005; Atwell et al., 2010). Second, a SNP that occurs at a canonical splice site may affect splicing. Specifically, a SNP that occurs in a splice donor site causes intron retention in the mRNA, whereas a SNP that occurs in a splice acceptor site removes an exon from the original mRNA. Splice site variations may eventually lead to frameshifts or premature stop codons (Balasubramanian et al., 2011). Third, insertion or deletion variants with non-integral multiples of 3 located in the gene coding region can lead to frameshifts by disrupting the full-length transcript (MacArthur et al., 2012; Lim et al., 2013). Fourth, the loss of an initiation codon can lead to LoF mutations. The loss of transcription start codon (ATG) variations prevents gene transcription if there is no alternative start codon near the gene. For example, in humans, the loss of the start codon in the FRMD7 gene leads to idiopathic infantile nystagmus disease (Choi et al., 2015).

Given that most genes in a genome can be transcribed into multiple transcripts, only mutations that affect all transcripts are regarded as LoF mutations in many studies (Figure 1B) (Balasubramanian et al., 2011; MacArthur et al., 2012; Kaiser et al., 2015).

**LoF MUTATIONS ARE ABUNDANT IN DIVERSE SPECIES**

With the rapid development of sequencing technologies, LoF mutations have been studied in diverse species at the population level. Almost every genome contains many LoF mutations in either heterozygous or homozygous states. Most LoF mutations in a genome tend to be present at low allele frequencies and have a heterozygous status (Lek et al., 2016; Minikel et al., 2020). In selfing or consanguineous outcrossing species, there is a higher ratio of homozygous LoF mutations (Saleheen et al., 2017; Minikel et al., 2020). In humans, an analysis of 7597 genomes identified 17 764 stop-gain variants and 13 915 frameshift variants within 11 369 protein-coding genes (Rausell et al., 2014). Based on 1432 whole exome sequences from five isolated European populations, 173 homozygous LoF mutations were identified within 167 genes (Kaiser et al., 2015). Analysis of whole genomes sequencing from 2636 Icelanders and chip-genotype data from 101 584 additional individuals identified 6795 autosomal LoF variants within 4924 genes (Sulem et al., 2015). Our recent study of 1071 A. thaliana genomes from worldwide accessions revealed 60 819 LoF variants within 12 907 genes (Xu et al., 2019).

The evolutionary pattern of LoF variants in natural populations can be influenced by various factors. Redundant genes within larger gene families are more likely to gain LoF mutations due to the presence of paralogs that buffer their functional effects (MacArthur et al., 2012; Xu et al., 2019). Furthermore, our recent study revealed that the level of nucleotide diversity and the density of transposable elements are positively correlated with the presence of LoF mutations (Xu et al., 2019). Given that
et al. (2012) found that transcription is decreased in only 25% of genes eventually degraded in this manner. For example, MacArthur et al. (2009) predicted that transcripts regulated by the NMD pathway are degraded due to an unstable protein-folding state (Williams et al., 2003). By contrast, at the protein level, a truncated protein may be functionally co-opted by the cell to fulfill the functions of the wild-type protein or to act as dominant-negative if the LoF truncation is located further than 50 base pairs upstream of the last exon, the truncated protein may be degraded through the NMD pathway (Figure 1B) (Balasubramanian et al., 2011). In humans, based on 1151 individual genomes from 56 worldwide populations, 55% of genes are predicted to undergo transcript degradation through the NMD pathway (Yngvadottir et al., 2009). Nevertheless, only a small fraction of transcripts predicted to be regulated by the NMD pathway are eventually degraded in this manner. For example, MacArthur et al. (2012) found that transcription is decreased in only 25% of the genes predicted to be regulated by the NMD pathway. By contrast, at the protein level, a truncated protein may be degraded due to an unstable protein-folding state (Williams et al., 2003).

FUNCTIONAL EFFECTS OF LoF MUTATIONS

Compared with functional protein-coding genes (Figure 2A), LoF mutations are natural gene knockouts that can provide insight into gene function in diverse organisms (Cheetham et al., 2020). For example, natural variants at the FRI locus from more than 1000 Arabidopsis genomes have helped to reveal the function of the FRI central domain (Zhang and Jimenez-Gomez, 2020). In A. thaliana, natural brix LoF alleles confer root adaptation to acidic soil, and a natural knockout allele of ARMA-DILLO REPEAT-CONTAINING KINESIN1 causes root hair branching (Gujas et al., 2012; Rishmawi et al., 2014). In addition, different LoF mutations of duplicated genes in diverse A. thaliana populations can lead to hybrid incompatibility (Blevins et al., 2017). In rice, a natural LoF mutation in GSE5, which encodes a plasma membrane-associated protein, contributes to grain size diversity (Duan et al., 2017).

However, not all LoF mutations lead to complete functional knockouts (Figure 2B). First, the effect of an LoF mutation on a gene depends on its location. Many LoF mutations at the 5' or 3' ends of affected genes do not completely destroy their functions (Figure 2C) (de Valles-Ibanez et al., 2016; MacArthur et al., 2012), given that an alternative start codon at the 5' end may rescue the transcript, and an LoF mutation at the 3' end will probably not remove the core functional domain. Second, the truncated proteins may act as dominant-negative mutations. For example, the nonsense allele of sex-determining region Y-box transcription factor 9 (MiniSOX9) acts in a dominant-negative manner to buffer the effect of the wild-type protein, contributing to biological diversity and drug resistance (Billmyre et al., 2017). In budding yeast, loss-of-start and nonsense mutations in AMN1 (antagonist of mitotic exit network) cause a transition from multicellularity to unicellularity (Kuzdzal-Fick et al., 2019). In humans, much more attention has been paid to the effects of LoF mutations that are correlated with Mendelian diseases. The precise annotation of LoF mutations has played an important role in the diagnosis of rare diseases (Cummings et al., 2020). LoF mutations have been found in defensive genes of many individuals and may lead to diseases (Saleheen et al., 2017). For example, in 9% of the European population, two LoF variants occurred independently in the filaggrin (FLG) gene and were associated with an increased risk of atopic dermatitis (Palmer et al., 2006). Consanguineous unions can confer more homozygous LoF variants, making them good systems for complex disease study (Saleheen et al., 2017). Based on a population with a high rate of consanguinity, association studies between 49 138 LoF mutations and more than 200 biochemical and disease traits revealed many LoF mutations in diverse disease genes (Saleheen et al., 2017).

In the human fungal pathogen Cryptococcus deuterogattii, an LoF mutation in the mismatch repair component MSH2 can affect phenotypic diversity and drug resistance (Billmyre et al., 2017). In budding yeast, loss-of-start and nonsense mutations in AMN1 (antagonist of mitotic exit network) cause a transition from multicellularity to unicellularity (Kuzdzal-Fick et al., 2019). In humans, much more attention has been paid to the effects of LoF mutations that are correlated with Mendelian diseases. The precise annotation of LoF mutations has played an important role in the diagnosis of rare diseases (Cummings et al., 2020). LoF mutations have been found in defensive genes of many individuals and may lead to diseases (Saleheen et al., 2017). For example, in 9% of the European population, two LoF variants occurred independently in the filaggrin (FLG) gene and were associated with an increased risk of atopic dermatitis (Palmer et al., 2006). Consanguineous unions can confer more homozygous LoF variants, making them good systems for complex disease study (Saleheen et al., 2017). Based on a population with a high rate of consanguinity, association studies between 49 138 LoF mutations and more than 200 biochemical and disease traits revealed many LoF mutations in diverse disease genes (Saleheen et al., 2017).
type allele (Abdel-Samad et al., 2011). Third, a functional pseudogene may exist. In Drosophila sechellia, the transcriptional readthrough of the premature stop codon in the chemosensory variant ionotropic glutamate receptor repertoire (Ir75a) can encode a functional receptor (Figure 2D) (Prieto-Godino et al., 2016). Fourth, LoF can be an effective evolutionary mechanism for generating new functional genes with physicochemical properties similar to those of the genes from which they originated, as recently demonstrated (Figure 2E) (Bartonek et al., 2020).

**ESSENTIAL GENES IN NATURAL POPULATIONS**

Essential genes are those genes that are necessary for survival (Meinke et al., 2008). An individual cannot survive if LoF mutations have occurred in essential genes; otherwise, these genes would be considered non-essential (Figure 3A). Essential genes in one species may be functionally redundant in another related species if they have undergone duplication (Figure 3B). The number of essential genes, namely the minimal gene set required for survival, remains a fundamental biological question (Koonin, 2003; Meinke et al., 2008). Beyond its theoretical importance, the study of essential genes in diverse organisms is crucial for synthetic biology, which requires essential gene lists to synthesize new cell lines or new organisms. Despite knowing the number of essential gene families in a lineage such as the green plants, which are estimated to contain approximately 2745–2928 core gene families (Van Bel et al., 2015), our recent study revealed that at least 34% of protein-coding genes do not have any LoF variants based on 1071 genomes from worldwide accessions. These genes are probably essential for the survival of A. thaliana in its natural environment (Xu et al., 2019). In contrast to the 358 genes identified as essential in a genetic study of mutants (Meinke et al., 2008) or the 2675 essential genes predicted by machine-learning methods (Lloyd et al., 2015), our essential gene number is much larger but understandable. Because growth conditions in laboratories are much better than in natural habitats, plants probably tolerate more LoF mutations when grown in laboratories. Therefore, the definition of essential genes is context dependent, reflecting the niche that the organisms inhabit.

**LoF MUTATIONS ARE CRUCIAL FOR ADAPTATION AND DIVERSIFICATION**

LoF mutations in the genome may be neutral, deleterious, or advantageous. Neutral or less-deleterious LoF mutations can be tolerated and may even accumulate during range expansion (Figure 4). However, compared with SNPs at the whole-genome level, LoF mutations are biased toward low allele frequencies, indicating that they are mostly deleterious and under purifying selection (MacArthur et al., 2012; The 1000 Genomes Project Consortium, 2010; Xu et al., 2019).
Nevertheless, the “less is more” hypothesis proposes that gene loss may be beneficial to organisms (Olson, 1999). Adaptive LoF mutations have been observed frequently in bacteria and yeast. An analysis of bacterial fitness in more than 100 different conditions revealed that LoF mutations can provide fitness benefits (Hottes et al., 2013). In the yeast S. cerevisiae, aquaporin genes are critical for freeze-thaw tolerance (Tanghe et al., 2004). However, sensitive strains have lost the function of aquaporin genes independently at least six times to adapt to high-sugar substrates (Will et al., 2010). Adaptive LoF mutations have also been found in plants. In A. thaliana, about 1% of LoF mutations are under positive selection, and the LoF allele of the KUK gene is correlated with longer roots (Xu et al., 2019). Overall, gene loss may be an adaptive process; adaptive LoF mutations may be quickly fixed in specific scenarios such as range expansion to new niches (Figure 4).

LoF mutations play important roles in the evolution and diversification of diverse organisms. In plants, a premature stop codon in GL4 caused smaller grain size and loss of seed shattering during African rice domestication (Wu et al., 2017). Furthermore, LoF mutations have been found to act as evolutionary hotspots for changing plant-pollinator communication and speciation. The floral scent is an important chemical signal between plants and pollinators (Klahre et al., 2011). In the genus Petunia (Solanaceae), a premature stop codon in the gene encoding cinnamate-CoA ligase 1 (CNL1), which produces phenylalanine-derived volatiles, eliminates the scent of Petunia exserta. As a result, P. exserta cannot attract hawkmoths and shifts its pollinator from hawkmoths to hummingbirds (Segatto et al., 2014; Amrad et al., 2016). A similar phenomenon was observed in Capsella (Brassicaceae), which independently lost CNL1 and scent twice, contributing to the transition from the outcrossing species Capsella grandiflora to the selfing species Capsella rubella (Sas et al., 2016). Similarly, flower color plays an important role in pollinator attraction (Bradshaw and Schemske, 2003). In Petunia axillaris, an LoF mutation in ANTHOCYANIN2 (AN2) occurred independently at least five times, changing the flower color from violet-red to white compared with Petunia integrifolia and influencing the shift in pollinator attraction from bees to hawkmoths (Hoballah et al., 2007).

In the Drosophila relative Scaptomyza flava, LoF mutations in odorant receptor genes resulted in a transition to herbivory from its yeast-feeding relatives (Goldman-Huertas et al., 2015). In mammals, LoF mutations in proto-Xist and its four flanking protein genes are associated with the emergence of X-chromosome inactivation and played an important role in the divergence of eutherians and marsupials (Duret et al., 2006). LoF mutations have also been found to be beneficial in human evolution. A CASP12 LoF allele is known to promote resistance to severe sepsis (Saleh et al., 2004), and rare LoF mutations in SLC30A8 can protect against type 2 diabetes (Flannick et al., 2014).

**FUTURE DIRECTIONS**

LoF mutations are prevalent in natural populations of diverse species. Species-wide studies could be performed to understand the genome-wide distribution patterns, functional effects, and evolutionary importance of LoF mutations. Several studies of LoF mutations at the genome level have been performed in natural populations (MacArthur et al., 2012; Xu et al., 2019). In particular, our recent study revealed that the level of nucleotide diversity, the density of transposable elements, and gene family size are positively correlated with the presence of LoF mutations (Xu et al., 2019). However, LoF studies are mostly limited to humans, chimpanzees, and Arabidopsis. More species should be studied to understand the evolutionary patterns and importance of LoF mutations. Many crucial questions about LoF mutations need to be addressed in the future.

First, the identification of LoF mutations is usually reference biased, based on comparison with one reference genome. For example, compared with the LoF mutations caused by stop-gain mutations, stop-loss mutations identified in other species may also be important. Moreover, the functional effects of LoF alleles need to be experimentally verified. As LoF mutations can have both neutral and adaptive effects, understanding the adaptive landscape in natural populations is crucial.

| Species                     | Method                        | Gene number (percentage) | Reference                  |
|-----------------------------|-------------------------------|--------------------------|----------------------------|
| *Bacillus subtilis*         | Mutants                       | 192 (4.7% a)             | Kobayashi et al. (2003)    |
| *Candida albicans*          | Mutants                       | 567 (9% b)               | Roemer et al. (2003)       |
| *Haemophilus influenza*     | Mutants                       | 259 (38% b)              | Akerley et al. (2002)      |
| *Pseudomonas aeruginosa*    | Mutants                       | 321 (6.6% b)             | Poulsen et al. (2019)      |
| *Staphylococcus aureus*     | Antisense RNA                 | 150 (5.6% b)             | Ji et al. (2001)           |
| *Escherichia coli*          | Mutants                       | 303 (7.1% b)             | Baba et al. (2006)         |
| *Saccharomyces cerevisiae*  | Mutants                       | 1105 (18.7% b)           | Giaever et al. (2002)      |
| *Arabidopsis thaliana*      | Mutants                       | 358 (1.3% b)             | Meinke et al. (2008)       |
| *Arabidopsis thaliana*      | Machine learning prediction   | 2675 (9.8% b)            | Lloyd et al. (2015)        |
| *Arabidopsis thaliana*      | LoF at population level       | 9249 (34.6% b)           | Xu et al. (2019)           |
| *Mice*                      | Mutants                       | 38 (42% b)               | White et al. (2013)        |
| *Human*                     | LoF at population level       | 194 (3% b)               | Narasimhan et al. (2016)   |
| *Human*                     | LoF at population level       | 3230 (17.7% b)           | Lek et al. (2016)          |

Table 1. Essential Gene Numbers in Diverse Species.

aIndicates the percentage of all genes in the genome.
bIndicates the percentage of studied genes.
LoF mutations can be neutral, deleterious, or advantageous. Deleterious LoF mutations are usually present at low allele frequencies in natural populations due to purifying selection. Neutral or beneficial LoF mutations can be tolerated or may accumulate during range expansion, and beneficial LoF mutations may be fixed by positive selection.

Figure 4. Adaptive Evolution of LoF Mutations.
LoF mutations can be neutral, deleterious, or advantageous. Deleterious LoF mutations are usually present at low allele frequencies in natural populations due to purifying selection. Neutral or beneficial LoF mutations can be tolerated or may accumulate during range expansion, and beneficial LoF mutations may be fixed by positive selection.

Accessions may be premature stop codons in the reference genome instead. Pan-genome studies in diverse species would compensate for this reference bias (Zhao et al., 2018; Golicz et al., 2020; Liu et al., 2020).

Second, the evolutionary effects of natural LoF mutations on other genes in the same pathway may be complicated. When genes become non-functional, other genes in the same pathway may accumulate LoF mutations as well (Zufall and Rausher, 2004). Such an evolutionary tendency could explain Dollo's law, which posits the irreversibility of character elimination (Gould, 1970). More cases in diverse species are needed to understand the evolutionary consequences of LoF mutations.

Third, the functional effects of LoF mutations are interesting to study. Genome-wide study of LoF variants can provide information on gene lethality based on the frequency of LoF alleles for a given gene in natural environments (population gene lethality) (Albalat and Canestro, 2016). Our recent study revealed that 34% of the genes in A. thaliana are probably essential in the natural environment (Xu et al., 2019). However, essential genes validated by functional analysis may have LoF mutations as well. In our recent study of A. thaliana, 11.7% of the essential genes validated by functional analysis had LoF mutations (Xu et al., 2019). In yeast S. cerevisiae, a study of 1106 essential gene knockouts found that 88 (9%) of them could survive through adaptive evolution, and these were defined as evolvable essential genes (Liu et al., 2015). In addition, an essential gene mutation study of E. coli and S. cerevisiae revealed that revertant mutants could occur in essential gene mutations when cells were grown under stress conditions, causing the mutant to regain its gene function to some extent (Kheir Gouda et al., 2019; Rodrigues and Shakhnovich, 2019). The study of revertant LoF mutations provides insight into the drug resistance of microorganisms, which is useful in synthetic biology in medicine. How individuals with natural mutations in essential genes can survive in natural populations is an intriguing question for in-depth study.

Fourth, most studies of LoF mutations have focused on variations in coding regions; however, a mutation in the promoter or UTR could also affect gene expression (Yang et al., 2018; Niu et al., 2019; Xu et al., 2020). Synthetic biology can offer interesting insights into the beneficial effects of regulator loss. For example, in Chinese hamster ovary cells, repressor loss in the promoter region of PuroR leads to high gene expression and drug resistance (Farquhar et al., 2019).

Finally, LoF mutations that are abundant in natural populations can provide valuable genetic resources for the functional study of genes and are particularly useful for crop breeding. New genes (gene duplication and de novo genes) or new functional alleles can make fruitful contributions to crop breeding. For instance, the COLD1 allele from the wild rice Oryza rufipogon confers chilling tolerance (Ma et al., 2015), and the 27-kDa γ-zein gene duplication contributes to maize protein quality through endosperm modification (Liu et al., 2016). However, the agricultural importance of LoF mutations is largely unknown. In plants, several studies have reported that LoF mutations can also increase crop yield. For example, LoF of GW2, a gene that encodes a RING-type E3 ubiquitin ligase, can increase grain width and weight (Song et al., 2007). Similarly, an LoF mutation of MEI2-LIKE PROTEIN4 (OML4) leads to large and heavy grains in rice (Oryza sativa) (Lyu et al., 2020). Nevertheless, there have been no studies of LoF mutations at the genome level in crops to date.

Overall, there are abundant natural LoF mutations in the genomes of diverse organisms, and this important genetic variation is correlated with biodiversity and adaptation. In particular, LoF mutations in natural populations provide an invaluable resource and a robust framework for gaining theoretical biological insight while simultaneously improving crop breeding in the context of climate change.

Funding
This work was supported by the National Natural Science Foundation of China (31925004), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27010305), and the Innovative Academy of Seed Design, Chinese Academy of Sciences.

Authors Contributions
Y.L.G. conceived the study. Y.C.X. and Y.L.G. wrote the manuscript.
ACKNOWLEDGMENTS

We thank members of the Guo lab for suggestions and comments about this work, and especially the anonymous reviewers for their help improving the manuscript. We apologize for our inability to cite many relevant publications due to space limitations. No conflict of interest declared.

REFERENCES

Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N., and Cummings, B.B., Karczewski, K.J., Kosmicki, J.A., Seaby, E.G., Watts, N.A., Singer-Berk, M., Mudge, J.M., Karjalainen, J., Satterstrom, F.K., O’Donnell-Luria, A.H., et al. (2020). Transcript expression-aware annotation improves rare variant interpretation. Nature 581:452–458.

de Valles-Ibanez, G., Hernandez-Rodriguez, J., Prado-Martinez, J., Luisi, P., Marques-Bonet, T., and Casals, F. (2016). Genetic load of loss-of-function polymorphic variants in great apes. Genome Biol. Evol. 8:871–877.

Duan, P., Xu, J., Zeng, D., Zhang, B., Geng, M., Zhang, G., Huang, K., Huang, L., Xu, R., Ge, S., et al. (2017). Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. Mol. Plant 10:685–694.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science 312:1653–1655.

Farquhar, K.S., Charlebois, D.A., Szenk, M., Cohen, J., Nevozhay, D., and Balazsi, G. (2019). Role of network-mediated stochasticity in mammalian drug resistance. Nat. Commun. 10:2766.

Fink, G.R. (1987). Pseudogenes in yeast. Cell 49:5–6.

Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burtt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R., et al. (2014). Loss-of-function mutations in SLC9A8 protect against type 2 diabetes. Nat. Genet. 46:357–363.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391.

Goldman-Huertas, B., Mitchell, R.F., Lapoint, R.T., Faucher, C.P., Hildebrand, J.G., and Whitman, N.K. (2015). Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. Proc. Natl. Acad. Sci. U S A 112:3026–3031.

Golitz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., and Edwards, D. (2020). Pan genomics comes of age: from bacteria to plant and animal applications. Trends Genet. 36:132–145.

Gould, S.J. (1970). Dollo on Dollo’s law: irreversibility and the status of evolutionary laws. J. Hist. Biol. 3:189–212.

Gujas, B., Alonso-Blanco, C., and Hardtke, C.S. (2012). Natural Arabidopsis bxr loss-of-function alleles confer root adaptation to acidic soil. Curr. Biol. 22:1962–1968.

Guo, Y.L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J. 73:941–951.

Guo, Y.L., Todesco, M., Hagmann, J., Das, S., and Weigel, D. (2012). Independent FLC mutations as causes of flowering-time variation in Arabidopsis thaliana and Capsella rubella. Genetics 192:729–739.

Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P., and Gerstein, M. (2003). Identification of pseudogenes in the Drosophila melanogaster genome. Nucleic Acids Res. 31:1033–1037.

Hoballah, M.E., Gubitza, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell’Olivo, A., Arnold, M., and Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in Petunia. Plant Cell 19:779–790.

Hottes, A.K., Freddolino, P.L., Khare, A., Donnell, Z.N., Liu, J.C., and Tavazoie, S. (2013). Bacterial adaptation through loss of function. PLoS Genet. 9:e1003617.

Inokuchi, H., Kodaira, M., Yamamoto, K., and Ozeki, H. (1979). Variability of 3′-terminal region of transfer ma2gln gene in E. coli - unequal crossover with presumptive pseudogenes. Jpn. J. Genet. 54:437.

Jacc, C., Miller, J.R., and Brownlee, G.G. (1977). A pseudogene in 5S DNA of Xenopus laevis. Cell 12:109–120.

de Valles-Ibanez, G., Hernandez-Rodriguez, J., Prado-Martinez, J., Luisi, P., Marques-Bonet, T., and Casals, F. (2016). Genetic load of loss-of-function polymorphic variants in great apes. Genome Biol. Evol. 8:871–877.

Duan, P., Xu, J., Zeng, D., Zhang, B., Geng, M., Zhang, G., Huang, K., Huang, L., Xu, R., Ge, S., et al. (2017). Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. Mol. Plant 10:685–694.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science 312:1653–1655.

Farquhar, K.S., Charlebois, D.A., Szenk, M., Cohen, J., Nevozhay, D., and Balazsi, G. (2019). Role of network-mediated stochasticity in mammalian drug resistance. Nat. Commun. 10:2766.

Fink, G.R. (1987). Pseudogenes in yeast. Cell 49:5–6.

Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burtt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R., et al. (2014). Loss-of-function mutations in SLC9A8 protect against type 2 diabetes. Nat. Genet. 46:357–363.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391.

Goldman-Huertas, B., Mitchell, R.F., Lapoint, R.T., Faucher, C.P., Hildebrand, J.G., and Whitman, N.K. (2015). Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. Proc. Natl. Acad. Sci. U S A 112:3026–3031.

Golitz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., and Edwards, D. (2020). Pan genomics comes of age: from bacteria to plant and animal applications. Trends Genet. 36:132–145.

Gould, S.J. (1970). Dollo on Dollo’s law: irreversibility and the status of evolutionary laws. J. Hist. Biol. 3:189–212.

Gujas, B., Alonso-Blanco, C., and Hardtke, C.S. (2012). Natural Arabidopsis bxr loss-of-function alleles confer root adaptation to acidic soil. Curr. Biol. 22:1962–1968.

Guo, Y.L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J. 73:941–951.

Guo, Y.L., Todesco, M., Hagmann, J., Das, S., and Weigel, D. (2012). Independent FLC mutations as causes of flowering-time variation in Arabidopsis thaliana and Capsella rubella. Genetics 192:729–739.

Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P., and Gerstein, M. (2003). Identification of pseudogenes in the Drosophila melanogaster genome. Nucleic Acids Res. 31:1033–1037.

Hoballah, M.E., Gubitza, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell’Olivo, A., Arnold, M., and Kuhlemeier, C. (2007). Single gene-mediated shift in pollinator attraction in Petunia. Plant Cell 19:779–790.

Hottes, A.K., Freddolino, P.L., Khare, A., Donnell, Z.N., Liu, J.C., and Tavazoie, S. (2013). Bacterial adaptation through loss of function. PLoS Genet. 9:e1003617.

Inokuchi, H., Kodaira, M., Yamamoto, K., and Ozeki, H. (1979). Variability of 3′-terminal region of transfer ma2gln gene in E. coli - unequal crossover with presumptive pseudogenes. Jpn. J. Genet. 54:437.

Jacc, C., Miller, J.R., and Brownlee, G.G. (1977). A pseudogene in 5S DNA of Xenopus laevis. Cell 12:109–120.

Plant Communications 1, 100103, November 9 2020 © 2020 The Author(s).
Plant Communications

Ji, Y.D., Zhang, B., Van Horn, S.F., Warren, P., Woodnutt, G., Burnham, M.K.R., and Rosenberg, M. (2001). Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 293:2266–2269.

Kaiser, V.B., Svtini, V., Prendergast, J.G., Chau, Y.Y., Campbells, A., Pataric, I., Barroso, I., Joshi, P.K., Hastie, N.D., Miljkovic, A., et al. (2015). Homozygous loss-of-function variants in European cosmopolitan and isolate populations. Hum. Mol. Genet. 24:5464–5474.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581:434–443.

Kheir, G., Yong, M.Y., Yurieva, M., Srinivasan, K.G., Liu, J., Lim, J.S., Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Kaiser, V.B., Svinti, V., Prendergast, J.G., Chau, Y.Y., Campbells, A., Pataric, I., Barroso, I., Joshi, P.K., Hastie, N.D., Miljkovic, A., et al. (2015). Homozygous loss-of-function variants in European cosmopolitan and isolate populations. Hum. Mol. Genet. 24:5464–5474.

Loss-of-Function Mutation Strategy for Adaptation

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A., Zhang, H., Liu, Z., Shi, M., et al. (2020). Pan-genome of wild and cultivated soybeans. Cell https://doi.org/10.1016/j.cell.2020.1005.1023.

Lloyd, J.P., Seddon, A.E., Moghe, G.D., Simenc, M.C., and Shiu, S.-H. (2015). Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. Plant Cell 27:2133–2147.

Lyu, J., Wang, D., Duan, P., Liu, Y., Huang, K., Zeng, D., Zhang, L., Dong, G., Li, Y., Xu, R., et al. (2020). Control of grain size and weight by the GSK2-LARGE1/OML4 pathway in rice. Plant Cell 32:1905–1918.

Ma, Y., Dai, X., Yu, X., Luo, W., Zheng, X., Zeng, D., Pan, Y., Lin, X., Liu, H., Zhang, D., et al. (2015). COLD1 confers chilling tolerance in rice. Cell 160:1209–1221.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science 339:823–826.

Meinke, D., Muralla, R., Sweeney, C., and Dickerman, A. (2008). Identifying essential genes in Arabidopsis thaliana. Trends Plant Sci. 13:483–491.

Meinke, D., W., et al. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat. Rev. Microbiol. 1:127–136.

Kuzdzal-Fick, J.J., Chen, L., and Balazsi, G. (2019). Disadvantages and benefits of evolved unicellularity versus multicellularity in budding yeast. Ecol. Evol. 9:8509–8523.

Le Corre, V., Roux, F., and Rebour, X. (2002). DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol. Biol. Evol. 19:1261–1271.

Lee, M.G., Lewis, S.A., Wilde, C.D., and Cowan, N.J. (1983). Evolutionary history of a multigene family—an expressed human beta-tubulin gene and 3 processed pseudogenes. Cell 30:477–487.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O’Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D. (2005). Diversity of flowering responses in wild Arabidopsis thaliana strains. PLoS Genet. 1:109–118.

Li, W.H., Gojobori, T., and Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239.

Li, Z.W., Chen, X., Wu, Q., Hagmann, J., Han, T.S., Zou, Y.P., Ge, S., and Guo, Y.L. (2016). On the origin of de novo genes in Arabidopsis thaliana populations. Genome Biol. Evol. 8:2190–2202.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Ware, J., Ruderfer, D.M., Fromer, M., et al. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77:235–242.

Liu, G., Yong, M.Y., Yurieva, M., Srinivasan, K.G., Liu, J., Lim, J.S., Poidinger, M., Wright, G.D., Zolezzi, F., Choi, H., et al. (2015). Gene essentiality is a quantitative property linked to cellular evolvability. Cell 163:1388–1399.

Liu, H., Shi, J., Sun, C., Gong, H., Fan, X., Qiu, F., Huang, X., Feng, Q., Zheng, X., Yuan, N., et al. (2016). Gene duplication confers enhanced expression of 27-kDa gamma-zein for endosperm modification in quality protein maize. Proc. Natl. Acad. Sci. U S A 113:4964–4969.
Loss-of-Function Mutation Strategy for Adaptation

of ARMDILLO REPEAT-CONTAINING KINESIN that causes root hair branching by mapping-by-sequencing. Plant Physiol. 166:1280–1287.

Rodrigues, J.V., and Shakhnovich, E.I. (2019). Adaptation to mutational inactivation of an essential gene converges to an accessible suboptimal fitness peak. eLife 8:e50509.

Roemer, T., Jiang, B., Davison, J., Ketela, T., Veillette, K., Breton, A., Tandia, F., Lintea, A., Sillaots, S., Marta, C., et al. (2003). Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery. Mol. Microbiol. 50:167–181.

Saleh, M., Vaillancourt, J.P., Graham, R.K., Huyck, M., Srinivasula, S.M., Alnemri, E.S., Steinberg, M.H., Nolan, V., Baldwin, C.T., Hotchkiss, R.S., et al. (2004). Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. Nature 429:75–79.

Saleheen, D., Natarajan, P., Armean, I.M., Zhao, W., Rasheed, A., Khetarpal, S.A., Zink, F., Hjartarson, E., Sigurdsson, G.T., Jonasdottir, A., Bonnycastle, L.J., Lintula, A., Sillaots, S., Marta, C., et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature 544:235–239.

Sas, C., Muller, F., Kappel, C., Kent, T.V., Wright, S.I., Hilker, M., and Lenhard, M. (2016). Repeated inactivation of the first committed enzyme underlies the loss of benzaldehyde emission after the selfing transition in Capsella. Curr. Biol. 26:3313–3319.

Segatto, A.L., Caze, A.L., Turchetto, C., Khare, U., Kuhlmeier, C., Bonatto, S.L., and Freitas, L.B. (2014). Nuclear and plastid markers reveal the persistence of genetic identity: a new perspective on the evolutionary history of Petunia exserta. Mol. Phylogenet. Evol. 70:504–512.

Shindo, C., Aranzana, M.J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., and Dean, C. (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. Plant Physiol. 138:1163–1173.

Song, X.J., Huang, W., Shi, M., Zhu, M.Z., and Lin, H.X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. 39:623–630.

Sulem, P., Helgason, H., Oddson, A., Stefansson, H., Gudjonsson, S.A., Zink, F., Hjartarson, E., Sigurdsson, G.T., Jonasdottir, A., Jonasdottir, A., et al. (2015). Identification of a large set of rare complete human knockouts. Nat. Genet. 47:448–452.

Tanghe, A., Van Dijck, P., Colavizza, D., and Thevelein, J.M. (2004). Aquaporin-mediated improvement of freeze tolerance of Saccharomyces cerevisiae is restricted to rapid freezing conditions. Appl. Environ. Microb. 70:3377–3382.

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Torrelo, J., Suyama, M., Zdobnov, E., and Bork, P. (2003). A genome-wide survey of human pseudogenes. Genome Res. 13:2559–2567.

Van Bel, M., Proost, S., Wischntizki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., and Vandepoele, K. (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. Plant Physiol. 158:590–600.

Vanin, E.F. (1985). Processed pseudogenes—characteristics and evolution. Annu. Rev. Genet. 19:253–272.

Werner, J.D., Borevitz, J.O., Uhlenhaut, N.H., Ecker, J.R., Chory, J., and Weigel, D. (2005). FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. Genetics 170:1197–1207.

White, J.K., Gerdin, A.K., Karp, N.A., Ryder, E., Buljan, M., Buswell, J.N., Salisbury, J., Clare, S., Ingham, N.J., Podrini, C., et al. (2013). Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell 154:452–464.

Will, J.L., Kim, H.S., Clarke, J., Painter, J.C., Fay, J.C., and Gasch, A.P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. PLoS Genet. 6:e1000893.

Williams, R.S., Chasman, D.I., Hau, D.D., Hui, B., Lau, A.Y., and Glover, J.N. (2003). Detection of protein folding defects caused by BRCA1- BRCT truncation and missense mutations. J. Biol. Chem. 278:53007–53016.

Xie, J.B., Li, Y., Liu, X.M., Zhao, Y.Y., Li, B.L., Ingvarsson, P.K., and Zhang, D.Q. (2019). Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. Plant Cell 31:563–578.

Xu, D., Gokcumen, O., and Khurana, E. (2020). Loss-of-function tolerance of enhancers in the human genome. PLoS Genet. 16:e1008663.

Xu, Y.C., Niu, X.M., Li, X.X., He, W., Chen, J.F., Zou, Y.P., Wu, Q., Zhang, Y.E., Busch, W., and Guo, Y.L. (2019). Adaptation and phenotypic diversification in Arabidopsis through loss-of-function mutations in protein-coding genes. Plant Cell 31:1012–1025.

Yang, L., Takuno, S., Waters, E.R., and Gaut, B.S. (2011). Lowly expressed genes in Arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation. Mol. Biol. Evol. 28:1193–1203.

Yang, L., Wang, H.N., Hou, X.Z., Zhou, Y.P., Han, T.S., Niu, X.M., Zhang, J., Zhao, Z., Todesco, M., Balasubramanian, S., et al. (2018). Parallel evolution of common allelic variants confers flowering diversity in Capsella rubella. Plant Cell 30:1322–1336.

Yngvadottir, B., Xue, Y., Searle, S., Hunt, S., Delgado, M., Morrison, J., Whittaker, P., Deloukas, P., and Tyler-Smith, C. (2009). A genome-wide survey of the prevalence and evolutionary forces acting on human nonsense SNPs. Am. J. Hum. Genet. 84:224–234.

Zhang, L., and Jimenez-Gomez, J.M. (2020). Functional analysis of FRIGIDA using naturally occurring variation in Arabidopsis thaliana. Plant J. 103:154–165.

Zhao, L., Liao, P., Jones, C.D., and Begun, D.J. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. Science 343:769–772.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. 50:278–284.

Zheng, L.L., Zhou, K.R., Liu, S., Zhang, D.Y., Wang, Z.L., Chen, Z.R., Yang, J.H., and Qu, L.H. (2018). dreamBase: DNA modification, RNA regulation and protein binding of expressed pseudogenes in human health and disease. Nucleic Acids Res. 46:D85–D91.

Zou, C., Lehti-Shiu, M.D., Thibaud-Nissen, F., Prakash, T., Buell, C.R., and Shiu, S.H. (2009). Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol. 151:3–15.

Zufall, R.A., and Rausher, M.D. (2004). Genetic changes associated with floral adaptation restrict future evolutionary potential. Nature 428:847–850.