Review / Revue

The importance of being essential: *EMBRYO-DEFECTIVE* genes in Arabidopsis

Martine Devic

Genome and Plant Development Laboratory, CNRS-IRD-Université de Perpignan,
UMR5096, 52, rue Paul Alduy, 66860 Perpignan cedex, France

Available online 2 September 2008

Presented by Christian Dumas

Abstract

With the completion of the sequence of the first bacterial genomes, scientists have been able to address the question: How many genes are required for cell viability? In attempting to reply to this question, the concept of the minimal gene set was developed and validated by systematic gene disruption. In a similar manner, whole genome comparisons and systematic Knock–Out have been performed in eukaryotes and have led to the identification to date of the set of essential genes in yeast and *C. elegans*. In the plant kingdom, the sequence of the Arabidopsis genome together with large-scale functional genomics programs now allow us to address the question of essentiality in Arabidopsis. These concerted efforts have resulted in the identification to date of up to 219 genes essential for seed development (*EMBRYO-DEFECTIVE*, *EMB*, genes). With this basic knowledge, we can start a valid comparison of essentiality in Arabidopsis and in other eukaryotes based on functional categories and orthologous relationships. Furthermore, the function of the *EMB* genes in the particular context of eukaryote evolution driven by whole genome duplications and selective gene loss will be discussed. ***To cite this article: M. Devic, C. R. Biologies 331 (2008).***
© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Essentialité des gènes chez les plantes : le cas des gènes embryonnaires létaux d'Arabidopsis. Avec l'achèvement de la séquence des premiers génomes bactériens, les scientifiques furent en mesure de répondre à la question : combien de gènes sont nécessaires pour conditionner la viabilité cellulaire ? En essayant de répondre à cette question, la notion de l'existence d'un jeu minimal de gènes fut élaborée et validée par inactivation systématique de gènes. De la même manière, des études analogues conduites chez les eucaryotes ont abouti à l'identification de l'ensemble des gènes essentiels chez la levure et *C. elegans*. Dans le règne végétal, le séquençage du génome d'Arabidopsis et la mise en place de programmes de génomique fonctionnelle à haut débit permettent aujourd'hui d'aborder la question de l'essentialité des gènes chez cette plante modèle. Ces efforts concertés ont abouti à l'identification à ce jour de 219 gènes qui sont essentiels au développement des graines (gènes *EMB*). Grâce à ces connaissances, nous pouvons initier une analyse critique de l'essentialité des gènes chez Arabidopsis et chez d'autres eucaryotes sur la base de catégories fonctionnelles et de relations d'orthologie entre gènes. En outre, nous discutons dans la présente revue la fonction des gènes *EMB*, dans le contexte particulier de l'évolution des génomes eucaryotes gouvernée par les duplications génomiques et la perte sélective de gènes. ***Pour citer cet article : M. Devic, C. R. Biologies 331 (2008).***
© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Embryogenesis: a complex developmental process

Embryogenesis is a critical step in the life cycle of the plant during which a multicellular organism is formed from a single cell by division, growth and differentiation [1,2]. Most of the essential developmental processes are initiated within the embryo and will be reiterated throughout the life of the plant. This complex process of seed development can be conceptually divided into three main steps. The first step will establish the pattern of the embryo by rapid cell division. The second step is characterized by cell expansion and massive synthesis and accumulation of reserves [3,4]. Ultimately, the seed will desiccate, arrest its development and enter into dormancy [5]. In addition, the development of the three components of the seed, namely the maternal integuments, the triploid endosperm and the diploid embryo, has to be coordinated to produce a viable seed. It is not surprising therefore that the expression of a large number of genes is required for seed development, as revealed by genetic analyses. Following mutagenesis, a substantial proportion of plants do not produce a normal seed set and as a consequence, homozygous plants cannot be obtained. Typically, the presence of 1/4 of aborted seeds in the siliques represents the principal characteristic of most *emb* mutations. The fertilized ovule is the first homozygous diploid stage of a mutation and the disruption of any gene associated with essential cellular functions will lead to seed abortion. Hence, in addition to the expected (and initially desired) seed-specific functions associated with the three phases of embryogenesis, seed mutants also represent the main source for the identification of essential genes in Arabidopsis (*Arabidopsis thaliana*). The group of David Meinke [6] has collectively named the seed mutations as *embryo-defective* (*emb*) and the corresponding genes as *EMB*, whose functions are essential for seed viability. Based on the number of *emb* mutations resulting from mutagenesis in relation to the size of the Arabidopsis genome, the average size of a gene and genetic mapping of these mutations, the number of essential genes has been estimated to be in the range of 500 to 1000 [7].

2. Essentiality and embryogenesis: identification of genes essential for seed development

To obtain information concerning the genes governing this complex developmental process several approaches have been used to identify the defective genes in embryonic mutants. The first *EMB* genes were identified by chromosome walking in mutant lines created by chemical/physical mutagenesis [8,9]. With the production of large-scale insertional mutant collections and the optimization of the methodologies to determine the sites of T-DNA insertion [10–13], the number of *EMB* genes that have been characterized by forward and reverse genetics has dramatically increased in the recent years. Large-scale screens for seed mutants have been performed by several laboratories [12,14–16]. The results of this international effort have been compiled in a common database called SeedGenes database (http://www.seedgenes.org), of which Dr. David Meinke is the curator. The SeedGenes database contains molecular and phenotypic information on essential Arabidopsis genes giving a seed phenotype when disrupted by mutation [17]. The work of my group has focused on the genetic analysis of the *emb* lines from the Arabidopsis T-DNA insertion lines of the INRA-Versailles collection. The *emb* lines identified among 16,000 T1 plants fell into five major classes named according to the morphology of the embryo: preglobular (14%), globular (34%), heart-transition (23%), torpedo-cotyledon (27%) and disorganized embryo morphology (2%) (Fig. 1). The two critical steps in early development (preglobular) are the resumption of cell division following fertilization [18,19] and the formation of the protoderm (embryonic epidermis) by periclinal division [20]. Arrest of the embryo at the globular stage represents the most frequent phenotype of *emb* mutants and the screening of independent mutant collections has also identified this stage as being the most sensitive step of embryo development [12,21]. The globular-to-heart transition is rich in developmental and metabolic events. Here, the embryo will acquire a bilateral symmetry by emergence of the cotyledons, a relative metabolic autonomy and a photosynthetic capacity associated with the differentiation of proplastids into chloroplasts. The first step of differentiation of the proplastids occurs during embryogenesis at the globular-to-heart transition phase [22]. The *emb* phenotypes in the heart and torpedo classes
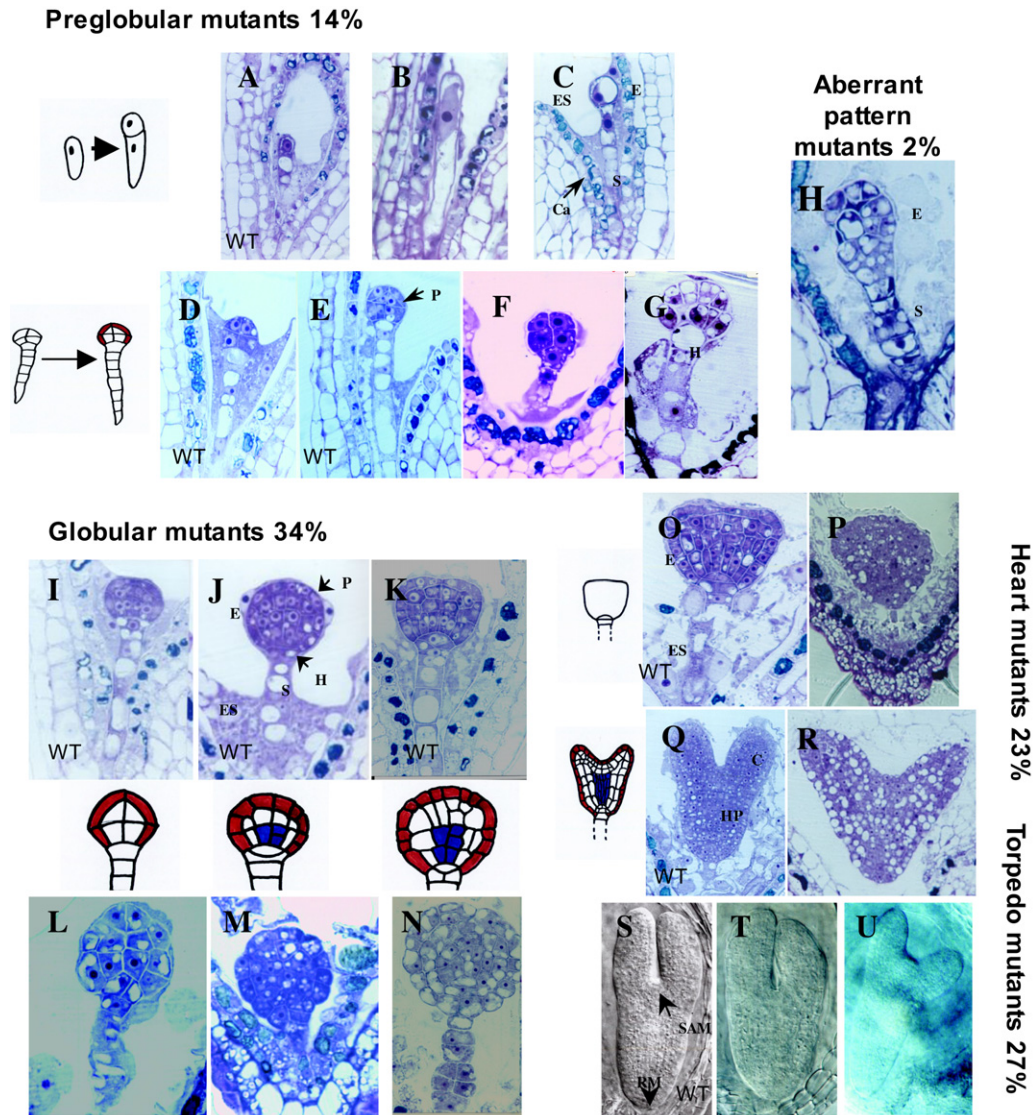
Fig. 1. Phenotypes of the various classes of *embryo-defective* mutants. Schematic representation of the main steps of embryo development occurring during preglobular, globular and heart-transition stages. Longitudinal sections of seeds stained with toluidine blue. Wild type embryos: at early embryogenesis (A) 1 cell, (D) octant or 8-cell and (E) dermatogen or 16-cell stages; at the (I) early-globular, (J) mid-globular and (K) late globular stages; at mid-embryogenesis, (O) triangular and (Q) heart. Mutant embryos: *preglobular* (B) *zeus1* mutant arrested at the zygote stage, (C) *cyclops1* arrested at 1 cell stage, (F) *quatre-quart1* and (G) embryo without protoderm; *globular* mutants (L-N); *heart* mutants (P) arrested at triangular stage and (R) arrested at heart stage; Aberrant pattern mutant (H). Cleared seeds viewed by Nomarski optic (S-T). Wild-type embryos at torpedo (S) stages of development and two examples of *torpedo* mutants (T-U). E, embryo; ES, endosperm; S, suspensor; P, protoderm; H, hypophysis; Ca, catechins; C, cotyledon; HP, hypocotyl; SAM, shoot apical meristem; RM, root meristem. Red color, protoderm; White, ground tissue; Blue, provascular tissue.

may be more related to seed-specific functions involved in seed maturation and desiccation. At the time of writing this review (July 2008), there were 219 confirmed *EMB* genes in the SeedGenes database corresponding to 27.5% of preglobular, 26.1% of globular, 11.4% of heart-transition and 35.1% of torpedo-cotyledon mutants. The enrichment in the molecular characterization

of *EMB* genes in the torpedo-cotyledon class reflects a greater interest of plant biologists for the later stages of seed development, which are associated with potentially important agronomic traits. This enrichment is mainly to the detriment of the globular mutants, which belong to the class characterized by the less informative phenotype. Furthermore, the more recent use of reverse

genetics allowed the rapid characterization of multiple alleles of an *EMB* gene, showing that in some cases the phenotypic class is correlated to the severity of the allele.

3. Can the minimal gene set concept be applied to eukaryotic unicellular and multicelullar organisms?

A previous release of the SeedGenes database allowed the classification of the function of genes required for seed development and a preliminary comparison with essential genes in other organisms [23]. Essential genes are defined as genes being indispensable to support cellular life. They constitute a minimal gene set required for a living cell and can be considered as the foundation of life itself. In addition, proteins encoded by such genes constitute excellent targets for antibacterial agents, for the development of novel drugs against pathogens and of herbicides together with the development of an "antidote" provided to the crop plants but not to the weeds. Therefore, an identification of the essential gene set in plants is of considerable interest not only to answer fundamental questions but also for therapeutic and agronomic uses.

Curiously, the description of a minimal gene set does not necessarily require systematic functional genomics (for a review, [24]). Instead, this detection rather benefited from large-scale comparative genomic studies. Several theoretical studies have endeavored to derive the minimal set of genes that are necessary and sufficient to sustain a functional cell under ideal conditions. A comparison of the first two complete bacterial genomes, *Haemophilus influenza* and *Mycoplasma genitalium*, resulted in a catalogue of the minimal gene set consisting of 256 genes. Similar estimates have been obtained by analyzing viable gene knockouts in *Haemophilus influenza, Bacillus subtillis* and *M. genitalium* [25–28]. However, the gene knockout approach revealed that absolute evolutionary conservation does not necessarily equate with gene essentiality. To rigorously assess the gene essentiality in *M. genitalium*, both individually and in network, the complete chemical synthesis and assembly of the *M. genitalium* genome has been performed [29]. Although this work represents a considerable technical achievement, the authors have yet to demonstrate that their synthesized genomes will support the life of a bacterium. Information on the basic functions of these genes can be found in the Database of Essential Genes (DEG [30]: http://tubic.tju.edu.cn/deg/). This minimal synthetic organism will possess the complete system for translation, transcription and replication but with other systems such as the repair machinery,

the metabolic pathways, the signal transduction apparatus, and the molecular chaperones reduced to a bare minimum. Only 35% of the bacterial minimal gene set could be identified in yeast, as assessed by COGs (Cluster of Orthologous Group). The COG approach is based on the concept that any group of at least three proteins from distant genomes that are more similar to each other than to any other protein within the same genomes most probably belongs to a family of orthologs. This indicates that even among unicellular organisms, the composition of the COGs in the minimal gene set will be divergent between prokaryotes and eukaryotes. This may be due to selective gene loss from clades and horizontal gene transfer (HGT) occurring in prokaryotic evolution whereas HGT is insignificant/implausible in the evolution of eukaryotes. Although the essential house-keeping functions for a prokaryotic or a eukaryotic cell are probably similar, an identification of the genes encoding these functions requires different comparative and functional genomics. Furthermore, defining the minimal gene set without specifying the conditions under which the minimal organism is supposed to survive makes no sense.

The precepts of the minimum gene set in prokaryotes and those of COGs, together with the comparison of eukaryotic genomes, have been applied to the analysis of the Arabidopsis *EMB* genes. The list of the 219 confirmed *EMB* genes present in the SeedGenes database has been used in the present analysis. Although the minimal gene set concept may be intuitive for simple and unicellular organisms, the paradigm requires a reappraisal for multicellular organisms. Similarly to the bacterial functional niches, different plant cell types may have different sets of minimal genes. Due to the low complexity of the Arabidopsis embryo composed of few tissues and cell types, most of these genes will not be identified in screens of seeds arrested in development. The first step of seed development involves cell division probably using some products stored in the egg cell and some that are newly synthesised [2]. The second phase requires the *de novo* production of proteins by transcription and translation. Therefore, we would expect to find among the functions controlled by the Arabidopsis essential genes many common to the essential functions of a prokaryotic cell (described above) with the important caveat that gene duplications have occurred during eukaryotic genome evolution. This implies that the description of the minimal Arabidopsis gene set cannot be solely based on the identification of the *EMB* genes but requires a combined approach of functional and comparative genomics. However, the molecular characterization of the *EMB* genes provides
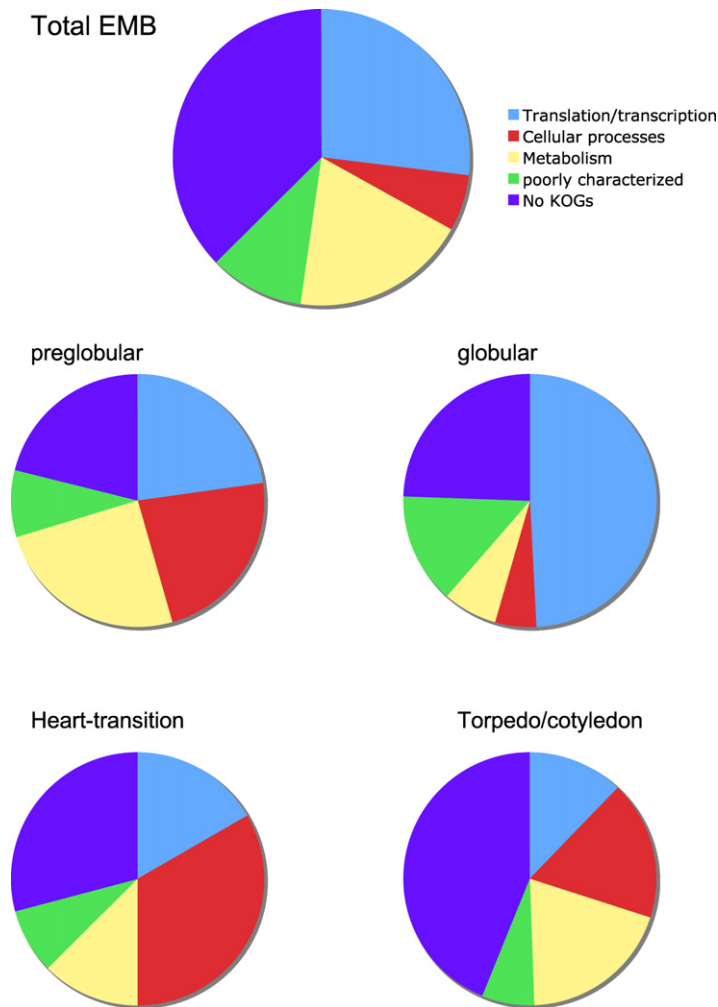
Fig. 2. Essential genes in Arabidopsis in relation to functional categories of KOGs. The KOGs determined for the 219 confirmed *EMB* genes have been grouped by function (Table 1). The repartitions of functional categories have been drawn for the 219 mutants and for each of the four main classes of *emb* mutants. (For colors see the web version of the article.)

the possibility to address the important question of the essentiality of a gene and to establish the criteria on which it is based.

The Arabidopsis genes have been classified in KOGs (euKaryotic Orthologous Group, http://www.ncbi.nlm. nih.gov/COG/grace/shokog.cgi/; [31]), as established by a comparison between seven complete eukaryotic genomes (namely, the genomes of three animals: *Homo sapiens, Drosophila melanogaster, Caenorahabditis elegans*, of the green plant *Arabidopsis thaliana*, of two fungi: *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* and of the microsporidian *Encephalitozoon cuniculi*). We interrogated this KOG database with the list of the 219 Arabidopsis *EMB* genes. A significant proportion of these genes could not be assigned

to any KOGs, a probable consequence of Arabidopsis being the only representative of the photosynthetic eukaryotes. However, there remained 860 KOGs common to the seven eukaryotes. Among these ubiquitous KOGs, 131 are represented by a single orthologous gene. Using this KOG nomenclature, the data presented in Fig. 2 and Table 1 show that the majority of KOGs are involved in translation and transcription machineries, cell division processes, similar to the essential prokaryotic functions. Furthermore, genes involved in metabolism account for a substantial proportion of *EMB* genes. The same data set was classified according to phenotype, showing strong bias such as predominance of KOGs in translation/transcription processes (J and A) for globular arrested seeds and

Table 1
EMB genes and functional classification by KOGs

Information storage and processing	Number of EMB genes[*]
J: Translation, ribosome structure and biogenesis	14
A: RNA processing and modification	15
K: Transcription	6
L: Replication, recombination and repair	9 + 1KR + 1KI
B: Chromatin structure and dynamics	2 + 2BD + 1BDL
Cellular processes and signaling	
D: Cell cycle control, cell division, chromosome partitioning	3 + 1DK + 1DL
Y: Nuclear structure	1 (YU)
V: Defense mechanisms	0
T: Signal transduction mechanisms	3
M: Cell wall/membrane/envelope biogenesis	1
N: Cell mobility	0
Z: Cytoskeleton	2
W: Extracellular structures	0
U: Intracellular trafficking, secretion, vesicle transport	5 + 1UO + 1UR + 1UZ
O: Posttranslational modification, protein turnover, chaperones	19 + 1OPR + 1OT
Metabolism	
C: Energy production and conversion	3
G: Carbohydrate transport and metabolism	5
E: Amino acid transport and metabolism	8 + 1EO
F: Nucleotide transport and metabolism	3
H: Coenzyme transport and metabolism	4
I: Lipid transport and metabolism	5 + 1IQ + 1IR + 1IT
P: Inorganic transport and metabolism	1
Q: Secondary metabolite biosynthesis, transport and metabolism	3
Poorly characterized	
R: General function prediction only	18
S: Unknown function	2
No KOGs	71
Total number of Arabidopsis KOGs	3285
KOGs represented by one ortholog in seven eukaryotic genomes	4/131 (expected 8–9 KOG)

[*] Total number of confirmed EMB: 219 as to May 2008.

of KOGs in cellular processes (O) in heart-transition mutants. The torpedo/cotyledon mutants exhibited the highest proportion of *EMB* genes that do not fall in KOGs (Fig. 2, no KOGs). This may indicate that these genes are involved in seed-specific functions, which are not represented in other eukaryotes. The existence of a correlation between ubiquitous KOGs and essentiality was tested and compared to the results obtained in yeast and *C. elegans* (Fig. 3). Disruption of genes in ubiquitous KOGs has a greater chance to lead to lethality than in other genes. Upon analysis of the 131 ubiquitous KOGs containing only a single orthologous gene, no enrichment in essential genes was observed (Table 1). This observation is somewhat similar to that documented in prokaryotes showing that gene conservation does not imply essentiality. In the case of Arabidopsis, it suggests that most genes belong to ubiquitous KOGs containing several homologous genes within the same genome but with non-redundant functions.

4. How has "essentiality" evolved in the context of whole genome duplication in eukaryotes?

The eukaryotic genomes have evolved by whole genome duplication and subsequent divergence of genes leading to new genes (neofunctionalization) and clade-specific gene losses [32–34]. Orthologs are defined as homologous genes that evolved via vertical descent from a single ancestral gene in the last common ancestor of the compared species. Paralogs are homologous genes that evolved by duplication of an ancestral gene. In addition, the eukaryotic genomes evolved by lineage specific expansion (LSE) [35]. There are 1458 LSEs in the Arabidopsis genome, a number similar to that of human genome (1373), but much higher than that of yeast genome (134). Among the major clusters of Arabidopsis LSEs are the plant-specific kinases, the plant-specific F-box containing proteins and the PPR proteins [36]. LSEs constitute one of the principal sources of organizational and regulatory diversity. The presence of sev-
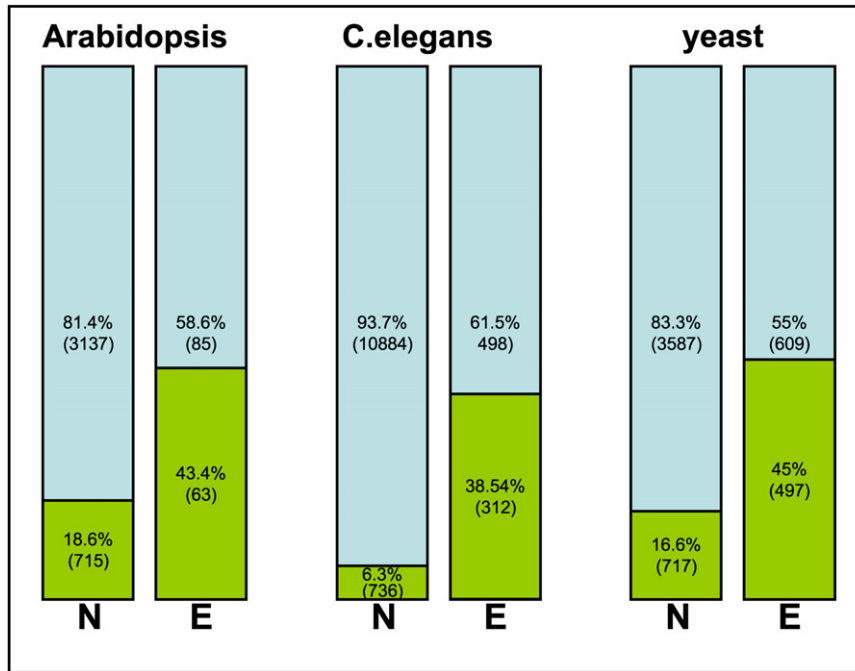
Fig. 3. Essential genes and conservation of KOGs among eukaryotes. Non-essential genes (N) and essential genes (E) in Arabidopsis, worm and yeast respective to KOGs. KOGs common to seven eukaryotic genomes (green) and other KOGs (blue). (For colors see the web version of the article.)

eral *PPR* genes among the *EMB* genes is in agreement with specialization occurring after gene multiplication rather than redundancy [37].

Paleoploidy or whole genome duplication is widespread in flowering plants and is believed to play an important role in the evolution and diversification of species [38,39]. The Arabidopsis genome has also been subjected to several rounds of polyploidy [34,40–42]. Therefore, the question arises as to what extent have gene divergence and gene loss been imposed on the duplication of the ancestral essential genes. Maintaining duplicates of essential genes may be an advantage and now we have to distinguish the concepts of "essential function" and "essential gene" since knock-out of one orthologous gene will no longer lead to lethality while the function still remain essential. Alternatively, duplication may affect gene dosage and gene expression and has a consequence on gene function. Blanc and Wolfe [43] studied the evolutionary effects of polyploidy on plant gene function in Arabidopsis and suggested that gene loss is the most likely outcome. However, these authors found evidence for a non-random pattern of elimination of duplicated genes. It appears that the duplicates of genes associated with transcriptional function have been preferentially retained, whereas duplicates of genes associated with DNA repair have been preferentially lost. In addition, more

than one half of the gene pairs showed significant variation in the pattern of gene expression indicating the possibility of a divergence of function. This observation has been confirmed by a recent work of Ganko et al. [44]. Therefore, one consequence of polyploidy is neo-subfunctionalization. The list of the confirmed Arabidopsis *EMB* genes has been presently analyzed using the web site at http://wolfe.gen.tcd.ie/athal/dup to identify those genes retaining a paralog in the Arabidopsis genome and those that are now unique. This information was correlated to the function encoded by the genes. Of the 219 *EMB* genes, 21 (9.6%) have retained a paralog in the Arabidopsis genome. On the whole Arabidopsis genome, less than 27% of the genes are duplicated between the sister genomic regions. Such low values suggest that duplicated *EMB* genes have been more susceptible to gene loss than any other genes. Although there are some striking examples of duplicated *EMB* genes such as the two genes encoding the catalytic subunit of DNA polymerase epsilon [19,45] and the three genes encoding the ribosomal protein L8/L2 (EMB2296, Meinke/SeedGenes), there seems to be an advantage for the functionality of an EMB protein to be encoded by a single gene.

Studies in yeast revealed that highly "connected" proteins are encoded by few duplicated genes and that deletion of these genes tends to result in lethality. In

agreement with this, the proportion of essential genes is significantly higher among singletons [46]. One aspect of "connectivity" of proteins with essential functions has already been hypothesized in eukaryotes sharing the 131 KOGs [31]. Indeed, nearly all of these functionally characterized KOGs consist of proteins that are subunits of multiprotein complexes. The most prominent of these complexes are those involved in rRNA processing and ribosome assembly together with the pre-40S subunit, as well as the spliceosome and various complexes involved in transcription [31]. The preponderance of multiprotein complex formation among the single gene KOGs is fully compatible with the balance hypothesis [47]. Accordingly, among the list of confirmed Arabidopsis *EMB* genes, several encode proteins participating in the formation of large macromolecular complexes as subunits of the RNA processing complex (SIG5 [48], AtPARN [49], PRP8 splicing factor, PRP4 [50], RNA helicase [51]), of the proteasomes (CUL1 [52], CUL3a and b [53], RPN1 [54]), of ribosomes/translational apparatus (RPL3 and RPL8A, Meinke, Seedgenes, AML1 = RPS5 [55], aminoacyl-tRNA synthetase [56]) and of molecular chaperones (SLP [57], HSP90, Meinke, Seedgenes). In yeast, a correlation has been established between protein connectivity, essentiality and gene uniqueness. Surprisingly, gene essentiality and gene duplication are not correlated in mammals [58]. The difference seems to be due to the fact that in yeast, singleton-encoded proteins tend to have more interacting partners than in mammals, suggesting that they are more intrinsically essential for the organism [59]. In mammals duplicated genes have a higher connectivity than do singletons suggesting that intrinsically they are more essential. In Arabidopsis, some of the *EMB* genes participating in large functional complexes still possess their paralogs. They are particular examples of gene diversification of expression or of function following duplication with high connectivity.

5. What about plant–specific essential genes?

The characterization of genes essential for seed development in Arabidopsis constitutes a foundation for further studies such as genome evolution in relation to functional genomics in photosynthetic organisms. The completion of the list of the Arabidopsis essential genes would considerably benefit from a comparison of appropriate genomes [60]. At present, our understanding is limited by the number of species whose genome has been completely sequenced and for which systematic knockouts have been characterized. As a consequence, a large proportion of the Arabidopsis proteins are not found in KOGs and thus comparisons are limited to prokaryotes and eukaryotes in the non-plant kingdoms. Although we can learn important features dealing with the basic cellular house-keeping functions common to all eukaryotes, we presently have great difficulties in determining the minimal gene set for a plant cell. However, with the increasing number of available genome sequences of photosynthetic organisms (e.g., prokaryotes, algae [61,62], and lower [63] and higher plants [64]) combined with functional genomics (such as in rice [65,66]), the description of the minimal gene set for a photosynthetic cell will become a realistic goal.

Acknowledgements

References

[1] J. Chandler, J. Nardmann, W. Werr, Plant development revolves around axes, Trends Plant Sci. 13 (2008) 78–84.

[2] C. Dumas, P. Rogowsky, Fertilization and early seed formation, C. R. Biologies 331 (2008) 715–725.

[3] S. Baud, L. Lepiniec, Compared analysis of the regulatory systems controlling lipogenesis in hepatocytes of mice and in maturing oilseeds of Arabidopsis, C. R. Biologies 331 (2008) 737–745.

[4] K. Gallardo, R. Thompson, J. Burstin, Reserve accumulation in legume seeds, C. R. Biologies 331 (2008) 755–762.

[5] L. Rajjou, I. Debeaujon, Seed longevity: Survival and maintenance of high germination ability of dry seeds, C. R. Biologies 331 (2008) 796–805.

[6] D. Errampalli, D. Patton, L. Castle, L. Mickelson, K. Hansen, J. Schnall, K. Feldmann, D. Meinke, Embryonic lethals and T-DNA insertional mutagenesis in Arabidopsis, Plant Cell 3 (1991) 149–157.

[7] D.W. Meinke, L.K. Meinke, T.C. Showalter, A.M. Schissel, L.A. Mueller, I. Tzafrir, A sequence-based map of Arabidopsis genes with mutant phenotypes, Plant Physiol. 131 (2003) 409–418.

[8] W. Lukowitz, U. Mayer, G. Jurgens, Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product, Cell 84 (1996) 61–71.

[9] M.H. Lauber, I. Waizenegger, T. Steinmann, H. Schwarz, U. Mayer, I. Hwang, W. Lukowitz, G. Jurgens, The Arabidopsis KNOLLE protein is a cytokinesis-specific syntaxin, J. Cell. Biol. 139 (1997) 1485–1493.

[10] J.M. Alonso, A.N. Stepanova, T.J. Leisse, C.J. Kim, H. Chen, P. Shinn, D.K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C.C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I. Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D.E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W.L. Crosby, C.C. Berry, J.R. Ecker, Genome-wide insertional mutagenesis of *Arabidopsis thaliana*, Science 301 (2003) 653–657.

[11] M. Devic, S. Albert, M. Delseny, T. Roscoe, Efficient PCR walking on plant genomic DNA, Plant Physiol. Biochem. 35 (1997) 331–339.

[12] J. McElver, I. Tzafrir, G. Aux, R. Rogers, C. Ashby, K. Smith, C. Thomas, A. Schetter, Q. Zhou, M.A. Cushman, J. Tossberg, T. Nickle, J.Z. Levin, M. Law, D. Meinke, D. Patton, Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*, Genetics 159 (2001) 1751–1763.

[13] F. Samson, V. Brunaud, S. Duchene, Y. De Oliveira, M. Caboche, A. Lecharny, S. Aubourg, FLAGdb++: a database for the functional analysis of the Arabidopsis genome, Nucleic Acids Res. 32 (2004) D347–D350.

[14] M. Devic, S. Albert, M. Delseny, Induction and expression of seed-specific promoters in Arabidopsis embryo-defective mutants, Plant J. 9 (1996) 205–215.

[15] G. Jurgens, U. Mayer, M. Busch, W. Lukowitz, T. Laux, Pattern formation in the Arabidopsis embryo: a genetic perspective, Philos. Trans. R. Soc. Lond. B Biol. Sci. 350 (1995) 19–25.

[16] D.W. Meinke, I.M. Sussex, Embryo-lethal mutants of *Arabidopsis thaliana*. A model system for genetic analysis of plant embryo development, Dev. Biol. 72 (1979) 50–61.

[17] I. Tzafrir, A. Dickerman, O. Brazhnik, Q. Nguyen, J. McElver, C. Frye, D. Patton, D. Meinke, The Arabidopsis SeedGenes project, Nucleic Acids Res. 31 (2003) 90–93.

[18] A. Ronceret, J. Gadea-Vacas, J. Guilleminot, F. Lincker, V. Delorme, S. Lahmy, G. Pelletier, M.E. Chabouté, M. Devic, The first zygotic division in Arabidopsis requires *de novo* transcription of thymidylate kinase, Plant J. 53 (2008) 776–789.

[19] A. Ronceret, J. Guilleminot, F. Lincker, J. Gadea-Vacas, V. Delorme, N. Bechtold, G. Pelletier, M. Delseny, M.E. Chabouté, M. Devic, Genetic analysis of two Arabidopsis DNA polymerase epsilon subunits during early embryogenesis, Plant J. 44 (2005) 223–236.

[20] S. Lahmy, J. Guilleminot, A.C. Schmit, G. Pelletier, M.E. Chabouté, M. Devic, QQT proteins colocalize with microtubules and are essential for early embryo development in Arabidopsis, Plant J. 50 (2007) 615–626.

[21] R. Yadegari, G. Paiva, T. Laux, A.M. Koltunow, N. Apuya, J.L. Zimmerman, R.L. Fischer, J.J. Harada, R.B. Goldberg, Cell differentiation and morphogenesis are uncoupled in Arabidopsis raspberry embryos, Plant Cell 6 (1994) 1713–1729.

[22] C. Garcion, J. Guilleminot, T. Kroj, F. Parcy, J. Giraudat, M. Devic, AKRP and EMB506 are two ankyrin repeat proteins essential for plastid differentiation and plant development in Arabidopsis, Plant J. 48 (2006) 895–906.

[23] I. Tzafrir, R. Pena-Muralla, A. Dickerman, M. Berg, R. Rogers, S. Hutchens, T.C. Sweeney, J. McElver, G. Aux, D. Patton, D. Meinke, Identification of genes required for embryo development in Arabidopsis, Plant Physiol. 135 (2004) 1206–1220.

[24] E.V. Koonin, How many genes can make a cell: the minimal-gene-set concept, Annu. Rev. Genomics Hum. Genet. 1 (2000) 99–116.

[25] C.A. Hutchison, S.N. Peterson, S.R. Gill, R.T. Cline, O. White, C.M. Fraser, H.O. Smith, J.C. Venter, Global transposon mutagenesis and a minimal Mycoplasma genome, Science 286 (1999) 2165–2169.

[26] B.J. Akerley, E.J. Rubin, V.L. Novick, K. Amaya, N. Judson, J.J. Mekalanos, A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*, Proc. Natl. Acad. Sci. USA 99 (2002) 966–971.

[27] J.I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. Hutchison 3rd, H.O. Smith, J.C. Venter, Essential genes of a minimal bacterium, Proc. Natl. Acad. Sci. USA 103 (2006) 425–430.

[28] K. Kobayashi, S.D. Ehrlich, A. Albertini, G. Amati, K.K. Andersen, M. Arnaud, K. Asai, S. Ashikaga, S. Aymerich, P. Bessieres, F. Boland, S.C. Brignell, S. Bron, K. Bunai, J. Chapuis, L.C. Christiansen, A. Danchin, M. Débarbouille, E. Dervyn, E. Deuerling, K. Devine, S.K. Devine, O. Dreesen, J. Errington, S. Fillinger, S.J. Foster, Y. Fujita, A. Galizzi, R. Gardan, C. Eschevins, T. Fukushima, K. Haga, C.R. Harwood, M. Hecker, D. Hosoya, M.F. Hullo, H. Kakeshita, D. Karamata, Y. Kasahara, F. Kawamura, K. Koga, P. Koski, R. Kuwana, D. Imamura, M. Ishimaru, S. Ishikawa, I. Ishio, D. Le Coq, A. Masson, C. Mauël, R. Meima, R.P. Mellado, A. Moir, S. Moriya, E. Nagakawa, H. Nanamiya, S. Nakai, P. Nygaard, M. Ogura, T. Ohanan, M. O'Reilly, M. O'Rourke, Z. Pragai, H.M. Pooley, G. Rapoport, J.P. Rawlins, L.A. Rivas, C. Rivolta, A. Sadaie, Y. Sadaie, M. Sarvas, T. Sato, H.H. Saxild, E. Scanlan, W. Schumann, J.F. Seegers, J. Sekiguchi, A. Sekowska, S.J. Séror, M. Simon, P. Stragier, R. Studer, H. Takamatsu, T. Tanaka, M. Takeuchi, H.B. Thomaides, V. Vagner, J.M. van Dijl, K. Watabe, A. Wipat, H. Yamamoto, M. Yamamoto, Y. Yamamoto, K. Yamane, K. Yata, K. Yoshida, H. Yoshikawa, U. Zuber, N. Ogasawara, Essential *Bacillus subtillis* genes, Proc. Natl. Acad. Sci. USA 100 (2003) 4678–4683.

[29] D.G. Gibson, G.A. Benders, C. Andrews-Pfannkoch, E.A. Denisova, H. Baden-Tillson, J. Zaveri, T.B. Stockwell, A. Brownley, D.W. Thomas, M.A. Algire, C. Merryman, L. Young, V.N. Noskov, J.I. Glass, J.C. Venter, C.A. Hutchison 3rd, H.O. Smith, Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome, Science 319 (2008) 1215–1220.

[30] R. Zhang, H.Y. Ou, C.T. Zhang, DEG: a database of essential genes, Nucleic Acids Res. 32 (2004) D271–D272.

[31] E.V. Koonin, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, D.M. Krylov, K.S. Makarova, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, I.B. Rogozin, S. Smirnov, A.V. Sorokin, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes, Genome Biol. 5 (2004) R7.

[32] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuvéglise, E. Talla,

N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J.M. Beckerich, E. Beyne, C. Bleykasten, A. Boisramé, J. Boyer, L. Cattolico, F. Confanioleri, A. De Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J.M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G.F. Richard, M.L. Straub, A. Suleau, D. Swennen, F. Tekaia, M. Wésolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, J.L. Souciet, Genome evolution in yeasts, Nature 430 (2004) 35–44.

[33] E.E. Eichler, D. Sankoff, Structural dynamics of eukaryotic chromosome evolution, Science 301 (2003) 793–797.

[34] T.J. Vision, D.G. Brown, S.D. Tanksley, The origins of genomic duplications in Arabidopsis, Science 290 (2000) 2114–2117.

[35] O. Lespinet, Y.I. Wolf, E.V. Koonin, L. Aravind, The role of lineage-specific gene family expansion in the evolution of eukaryotes, Genome Res. 12 (2002) 1048–1059.

[36] N. O'Toole, M. Hattori, C. Andres, K. Iida, C. Lurin, C. Schmitz-Linneweber, M. Sugita, I. Small, On the expansion of the pentatricopeptide repeat gene family in plants, Mol. Biol. Evol. 25 (2008) 1120–1128.

[37] D.A. Cushing, N.R. Forsthoefel, D.R. Gestaut, D.M. Vernon, Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis, Planta 221 (2005) 424–436.

[38] S.P. Otto, J. Whitton, Polyploid incidence and evolution, Annu. Rev. Genet. 34 (2000) 401–437.

[39] J.F. Wendel, Genome evolution in polyploids, Plant Mol. Biol. 42 (2000) 225–249.

[40] C. Simillion, K. Vandepoele, M.C. Van Montagu, M. Zabeau, Y. Van de Peer, The hidden duplication past of *Arabidopsis thaliana*, Proc. Natl. Acad. Sci. USA 99 (2002) 13627–13632.

[41] J.E. Bowers, C. Abbey, S. Anderson, C. Chang, X. Draye, A.H. Hoppe, R. Jessup, C. Lemke, J. Lennington, Z. Li, Y.R. Lin, S.C. Liu, L. Luo, B.S. Marler, R. Ming, S.E. Mitchell, D. Qiang, K. Reischmann, S.R. Schulze, D.N. Skinner, Y.W. Wang, S. Kresovich, K.F. Schertz, A.H. Paterson, A high-density genetic recombination map of sequence-tagged sites for sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses, Genetics 165 (2003) 367–386.

[42] G. Blanc, K. Hokamp, K.H. Wolfe, A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome, Genome Res. 13 (2003) 137–144.

[43] G. Blanc, K.H. Wolfe, Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution, Plant Cell 16 (2004) 1679–1691.

[44] E.W. Ganko, B.C. Meyers, T.J. Vision, Divergence in expression between duplicated genes in Arabidopsis, Mol. Biol. Evol. 24 (2007) 2298–2309.

[45] P.D. Jenik, R.E. Jurkuta, M.K. Barton, Interactions between the cell cycle and embryonic patterning in Arabidopsis uncovered by a mutation in DNA polymerase epsilon, Plant Cell 17 (2005) 3362–3377.

[46] Z. Gu, L.M. Steinmetz, X. Gu, C. Scharfe, R.W. Davis, W.H. Li, Role of duplicate genes in genetic robustness against null mutations, Nature 421 (2003) 63–66.

[47] B. Papp, C. Pal, L.D. Hurst, Dosage sensitivity and the evolution of gene families in yeast, Nature 424 (2003) 194–197.

[48] J. Yao, S. Roy-Chowdhury, L.A. Allison, AtSig5 is an essential nucleus-encoded Arabidopsis sigma-like factor, Plant Physiol. 132 (2003) 739–747.

[49] Y. Chiba, M.A. Johnson, P. Lidder, J.T. Vogel, H. van Erp, P.J. Green, AtPARN is an essential poly(A) ribonuclease in Arabidopsis, Gene 328 (2004) 95–102.

[50] R. Gross-Hardt, C. Kagi, N. Baumann, J.M. Moore, R. Baskar, W.B. Gagliano, G. Jurgens, U. Grossniklaus, LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis, PLoS Biol. 5 (2007) e47.

[51] K. Kobayashi, M.S. Otegui, S. Krishnakumar, M. Mindrinos, P. Zambryski, *INCREASED SIZE EXCLUSION LIMIT 2* encodes a putative DEVH box RNA helicase involved in plasmodesmata function during Arabidopsis embryogenesis, Plant Cell 19 (2007) 1885–1897.

[52] W.H. Shen, Y. Parmentier, H. Hellmann, E. Lechner, A. Dong, J. Masson, F. Granier, L. Lepiniec, M. Estelle, P. Genschik, Null mutation of *AtCUL1* causes arrest in early embryogenesis in Arabidopsis, Mol. Biol. Cell. 13 (2002) 1916–1928.

[53] A. Thomann, V. Brukhin, M. Dieterle, J. Gheyeselinck, M. Vantard, U. Grossniklaus, P. Genschik, Arabidopsis *CUL3A* and *CUL3B* genes are essential for normal embryogenesis, Plant J. 43 (2005) 437–448.

[54] V. Brukhin, J. Gheyselinck, V. Gagliardini, P. Genschik, U. Grossniklaus, The RPN1 subunit of the 26S proteasome in Arabidopsis is essential for embryogenesis, Plant Cell 17 (2005) 2723–2737.

[55] D. Weijers, M. Franke-van Dijk, R.J. Vencken, A. Quint, P. Hooykaas, R. Offringa, An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene, Development 128 (2001) 4289–4299.

[56] M. Berg, R. Rogers, R. Muralla, D. Meinke, Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in Arabidopsis, Plant J. 44 (2005) 866–878.

[57] N.R. Apuya, R. Yadegari, R.L. Fischer, J.J. Harada, J.L. Zimmerman, R.B. Goldberg, The Arabidopsis embryo mutant *schlepperless* has a defect in the chaperonin-60 alpha gene, Plant Physiol. 126 (2001) 717–730.

[58] H. Liang, W.H. Li, Gene essentiality, gene duplicability and protein connectivity in human and mouse, Trends Genet. 23 (2007) 375–378.

[59] X. He, J. Zhang, Higher duplicability of less important genes in yeast genomes, Mol. Biol. Evol. 23 (2006) 144–151.

[60] A.M. Gustafson, E.S. Snitkin, S.C. Parker, S. DeLisi, S. Kasif, Towards the identification of essential genes using targeted genome sequencing and comparative analysis, BMC Genomics 7 (2006) 265.

[61] E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A.Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, Y. Saeys, J. Wuyts, K. Jabbari, C. Bowler, O. Panaud, B. Piégu, S.G. Ball, J.P. Ral, F.Y. Bouget, G. Piganeau, B. De Baets, A. Picard, M. Delseny, J. Demaille, Y. Van de Peer, H. Moreau, Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features, Proc. Natl. Acad. Sci. USA 103 (2006) 11647–11652.

[62] S.S. Merchant, S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, A. Terry, A. Salamov, L.K. Fritz-Laylin, L. Maréchal-Drouard, W.F. Marshall, L.H. Qu, D.R. Nelson, A.A. Sanderfoot, M.H. Spalding, V.V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S.M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.L. Chen, V. Cognat, M.T. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa,

D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P.A. Lefebvre, S.D. Lemaire, A.V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J.V. Moroney, J. Moseley, C. Napoli, A.M. Nedelcu, K. Niyogi, S.V. Novoselov, I.T. Paulsen, G. Pazour, S. Purton, J.P. Ral, D.M. Riaño-Pachón, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S.L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.J. Chen, M. Elias, K. Gendler, C. Hauser, M.R. Lamb, H. Ledford, J.C. Long, J. Minagawa, M.D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A.M. Terauchi, P. Yang, S. Ball, C. Bowler, C.L. Dieckmann, V.N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R.T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y.W. Huang, J. Jhaveri, Y. Luo, D. Martínez, W.C. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I.V. Grigoriev, D.S. Rokhsar, A.R. Grossman, The Chlamydomonas genome reveals the evolution of key animal and plant functions, Science 318 (2007) 245–250.

[63] S.A. Rensing, D. Lang, A.D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P.F. Perroud, E.A. Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, T. Shin-I, Y. Kuroki, A. Toyoda, Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S.

Aoki, N. Ashton, W.B. Barbazuk, E. Barker, J.L. Bennetzen, R. Blankenship, S.H. Cho, S.K. Dutcher, M. Estelle, J.A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K.A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D.R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P.J. Rushton, A. Sanderfoot, G. Schween, S.H. Shiu, K. Stueber, F.L. Theodoulou, H. Tu, Y. Van de Peer, P.J. Verrier, E. Waters, A. Wood, L. Yang, D. Cove, A.C. Cuming, M. Hasebe, S. Lucas, B.D. Mishler, R. Reski, I.V. Grigoriev, R.S. Quatrano, J.L. Boore, The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants, Science 319 (2008) 64–69.

[64] E. Bossolini, T. Wicker, P.A. Knobel, B. Keller, Comparison of orthologous loci from small grass genomes Brachypodium and rice: implications for wheat genomics and grass genome annotation, Plant J. 49 (2007) 704–717.

[65] P. Larmande, C. Gay, M. Lorieux, C. Périn, M. Bouniol, G. Droc, C. Sallaud, P. Perez, I. Barnola, C. Biderre-Petit, J. Martin, J.B. Morel, A.A. Johnson, F. Bourgis, A. Ghesquière, M. Ruiz, B. Courtois, E. Guiderdoni, Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library, Nucleic Acids Res. 36 (2008) D1022–D1027.

[66] J. Zhang, C. Li, C. Wu, L. Xiong, G. Chen, Q. Zhang, S. Wang, RMD: a rice mutant database for functional analysis of the rice genome, Nucleic Acids Res. 34 (2006) D745–D748.