Journal of Plant Physiology 168 (2011) 3–8

Contents lists available at ScienceDirect

Journal of Plant Physiology

journal homepage: www.elsevier.de/jplph

Exploring the genomes: From Arabidopsis to crops

Manuel Spannagl[a], Klaus Mayer[a], Jörg Durner[b], Georg Haberer[a], Andreas Fröhlich[b,*]

[a] Helmholtz Zentrum München, Institute of Bioinformatics and Systems Biology, Ingolstädter Landstrasse 1, 85764 München, Germany
[b] Helmholtz Zentrum München, Institute of Biochemical Plant Pathology, Ingolstädter Landstrasse 1, 85764 München, Germany

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 16 July 2010
Accepted 18 July 2010

Keywords:
Comparative genomics
Gene expansions
Orthologous gene clusters
OrthoMCL

ABSTRACT

Model systems have played a crucial role for understanding biological processes at genetic, molecular and systems levels. *Arabidopsis thaliana* is one of the best studied model species for higher plants. Large genomic resources and mutant collections made Arabidopsis an excellent source for functional and comparative genomics. Rice and Brachypodium have a great potential to become model systems for grasses. Given the agronomic importance of grass crops, it is an attractive strategy to apply knowledge from Arabidopsis to grasses. Despite many efforts successful reports are sparse. Knowledge transfer should generally work best between orthologous genes that share functionality and a common ancestor. In higher plants, however, recent genome projects revealed an active and rapid evolution of genome structure, which challenges the concept of one-to-one orthologous mates between two species. In this study, we estimated on the example of protein families that are involved in redox related processes, the impact of gene expansions on the success rate for a knowledge transfer from Arabidopsis to the grass species rice, sorghum and Brachypodium. The sparse synteny between dicot and monocot plants due to frequent rearrangements, translocations and gene losses strongly impairs and reduces the number of orthologs detectable by positional conservation. To address the limitations of sparse synteny and expanded gene families, we applied for the detection of orthologs in this study orthoMCL, a sequence-based approach that allows to group closely related paralogs into one orthologous gene cluster. For a total of 49 out of 170 Arabidopsis genes we could identify conserved copy numbers between the dicot model and the grass annotations whereas approximately one third (34.7%, 59 genes) of the selected Arabidopsis genes lack an assignment to any of the grass genome annotations. The remaining 62 Arabidopsis genes represent groups that are considerably biased in their copy numbers between Arabidopsis and all or most of the three grass genomes.

Introduction

Model systems have played a crucial role for understanding biological processes at genetic, molecular and systems levels. *Arabidopsis thaliana*, the first higher plant for which a complete finished genome sequence has been published, is arguably the best studied and most important model species for higher plants. Large genomic resources like EST and full length cDNA databases, large collections of characterized mutants from genetic screens and insertion mutants as well as a huge set of expression data covering numerous environmental conditions and developmental stages made Arabidopsis also an excellent source for functional and comparative genomics. Rice and, recently, Brachypodium have a great potential to become model systems for the grasses which contain many of the agronomic most important species like maize, wheat or rice itself. For both species, a high quality complete genome sequence is available and rich genomic resources have been generated or are currently built by the research communities. However, current genomic resources and particularly knowledge in Arabidopsis for a vast number of plant biological processes are unprecedented. Given the agronomic importance of crops, it is an attractive strategy to apply knowledge from Arabidopsis to understand and even enhance crops, particularly for agricultural important traits that are traditionally in the focus of breeding. A striking example for crop improvement by knowledge transfer is the Arabidopsis GAI gene and its wheat ortholog which provided increased grain yields and short growth statures (Silverstone and Sun, 2000). Other attempts targeted genes that confer tolerance to various abiotic stresses, for example drought or osmotic stress (Dubouzet et al., 2003; Zhang et al., 2008). Despite many efforts, however, successful reports are sparse. Knowledge transfer should generally work best between orthologous genes that share by definition functionality and a common ancestor. In higher plants, however, recent genome projects

Abbreviations: DHA, dehydroascorbate; GSH, glutathione; GSSG, glutathione disulphide; GST, glutathione-S-transferase; MDHA, monodehydroascorbate; OGC, orthologous gene cluster; ROS, reactive oxygen species; SOD, superoxide dismutase.

* Corresponding author. Tel.: +49 89 3187 3136; fax: +49 8 89 3187 2726.
E-mail address: andreas.froehlich@helmholtz-muenchen.de (A. Fröhlich).

0176-1617/$ – see front matter © 2010 Elsevier GmbH. All rights reserved.
doi:10.1016/j.jplph.2010.07.008

revealed an active and rapid evolution of genome structure. Rearrangements of genes or chromosomal segments and transpositions are quite common and interrupt co-linearity between genomes of even closely related species. Moreover, poly-ploidizations and segmental and gene duplications are far more prominent in plant genomes compared to mammalian genomes. In combination with selective gene losses, such mutations challenge the concept of one-to-one orthologous mates between two or multiple species. Consequently, expanded gene families impair simple one-to-one functional matches, especially between more distant species.

In this study, we estimate the impact of gene expansions on the success rate for a knowledge transfer from Arabidopsis to three grass species, rice, sorghum and Brachypodium. As our work focuses on plant-pathogen interactions we also focus this study on selected gene families known to play a role in plant–pathogen defense. One of the most prominent features during plant microorganism interaction is the accumulation of reactive oxygen species (ROS) and the integration of this specific redox signal. Therefore we selected genes from the Arabidopsis gene ontology that are involved in redox related processes. These include the ROS producing NADPH-oxidases (GO:0016174), protein families involved in ROS scavenging, namely the superoxide dismutases (GO:0004784), catalases (GO:0004096) and glutathione dehydrogenases (GO:0045174) and the peroxidase (GO:0004601) and glutathione transferase (GO:0004364) families totaling 170 Arabidopsis genes.

NADPH-oxidases

An oxidative burst with accumulation of reactive oxygen species in the extracellular space of plant tissues is characteristic of plant cells exposed to abiotic stress, herbivores, symbiotic organisms and pathogens. The main sources for an oxidative burst in the apoplast are plasma membrane bound NADPH-oxidases and cell wall peroxidases. NADPH-oxidases in plants have been discovered because of their sequence similarity to the mammalian respiratory burst NADPH-oxidase subunit gp91^{phox} (Groom et al., 1996). The *A. thaliana* NADPH-oxidase family consists of 10 putative members, termed AtrbohA to AtrbohJ. The function of these proteins is not well defined. NADPH-oxidases have been shown to be involved in cell death regulation, wound healing and resistance reactions during bacterial and fungal infections (Sagi and Fluhr, 2006; Pogany et al., 2009).

Superoxide dismutases

Superoxide dismutases (SOD, EC 1.15.1.1) are the only enzymes which are able to detoxify superoxide. This highly reactive oxygen species is converted to oxygen and H_2O_2 by a "ping-pong"-mechanism of sequential reduction and oxidation of the enzyme metal center (Abreu and Cabelli, 2010).

Superoxide is produced in all compartments where electron transport chains are active, mainly in mitochondria, chloroplasts and peroxisomes, but also in the apoplast and cytosol. As superoxide readily reacts with sulfur-containing amino acids and all kind of proteins that contain Fe–S-clusters, all these compartments have their own set of SODs to protect proteins from this kind of potentially damaging reactions. The classification of SODs is based on their specific metal cofactor, iron (FeSOD), manganese (MnSOD), copper–zinc (CuZnSOD) or nickel (NiSOD). *A. thaliana* encodes for up to nine different SODs (Alscher et al., 2002).

Catalases

Catalases (EC 1.11.1.6) are tetrameric iron porphyrine proteins that catalyze the decomposition of H_2O_2 to oxygen and water. *A. thaliana* has three known catalase genes, that encode for at least six catalase isoforms. Catalases are mainly localized in the peroxisomes and show distinct spatial and temporal expression patterns (McClung, 1997). H_2O_2 is mainly produced by SODs, during photorespiration in peroxisomes and during fatty acid oxidation in the glyoxysomes. Since catalases are active only at rather high concentrations of H_2O_2, various other detoxification systems are present within the cell and its various compartments which can remove H_2O_2, notably ascorbate reductases, glutathione peroxidases and nonenzymatic antioxidants like ascorbate, glutathione, tocopherol and carotenoids (Gechev et al., 2006).

GSH-dehydrogenases

One of the major antioxidants present in plant cells is ascorbate. Ascorbate accumulates up to millimolar concentrations in plant tissues and reacts with hydroxyl radicals, superoxide and singlet oxygen. In an enzymatic detoxification reaction ascorbate peroxidase uses two molecules of ascorbate to reduce H_2O_2 to water. Oxidation of ascorbate results in monodehydroascorbate (MDHA) and dehydroascorbate (DHA) (Noctor and Foyer, 1998). Even when ascorbate is massively consumed, DHA rarely accumulates in plant cells, hinting at highly efficient ascorbate regeneration systems (Potters et al., 2010). GSH-dehydrogenase (Dehydroascorbate reductase, EC 1.8.5.1) is an enzyme which uses glutathione as a substrate for reduction of DHA to ascorbate, resulting in glutathione disulphide (GSSG). The so called ascorbate-glutathione cycle is then completed by glutathione reductase which re-reduces GSSG in a recovery process using NADPH.

Peroxidases

Peroxidases (EC 1.11.1.x) are a diverse group of proteins whose members can be found in all kingdoms. In general peroxidases catalyze reactions in which a peroxide is reduced and a substrate is oxidized. Peroxidases can be divided in haem and non-haem peroxidases, totalling to at least 10 different (super)families with up to several hundred members each (Oliva et al., 2009). A major plant specific haem-peroxidase family is represented by the class III peroxidases. In *A. thaliana* this family contains 73 members of conserved gene structure and protein size. Most reactions catalyzed by these proteins involve the reduction of H_2O_2 and the oxidation of various secondary metabolites. Byproducts of these reactions are often reactive oxygen species, mainly hydroxyl and peroxyl radicals (Passardi et al., 2005). The proteins of this family are secreted into the apoplastic space and play roles in processes as diverse as cell wall loosening and cross linking, lignification and suberinisation, nodulation in legumes and mycorrhization as well as root elongation and senescence (Cosio and Dunand, 2009). The largest portion of non-haem peroxidases is made up by the family of thiol peroxidases, divided into peroxiredoxins and glutathione peroxidases (Oliva et al., 2009). Both types of enzymes catalyze the reduction of a peroxide to the corresponding alcohol and water. While peroxiredoxins have to be transformed again into their reduced form after this reaction by reducing agents like thioredoxins or cyclophilines, glutathione peroxidases directly use reduced glutathione as an electron donor. Thiol peroxidases are involved in peroxide detoxification, chloroplast and mitochondria metabolism and various stress responses such as plant defense reactions (Tripathi et al., 2009).

Glutathione-S-transferases

Glutathione-S-transferases (GSTs, EC 2.5.1.18) are ubiquitous cytosolic proteins present in all prokaryotes and eukaryotes. Each soluble GST consists of a dimer of two 26 kDa subunits. *A. thaliana*