

decade, has mobilized a partnership of private and public research groups to focus on the development of new technologies to support *T. cacao* genetic improvement (http://www.ars-grin.gov/ars/SoAtlantic/Miami/ngr/cacao_genetics_meeting_summary.pdf) with a degree of coordination by the International Group for Genetic Improvement of Cocoa (INGENIC), which promotes the exchange of information and international collaboration on topics related to cocoa genetics. This initiative comes at a time of unprecedented new genomic tools that dramatically increase the genetic knowledge to inform breeding strategies and to accelerate notoriously slow tree breeding programs. Over the next decade, *T. cacao* is likely to become a model of an essentially wild crop being transformed through genome-based breeding. It will be interesting to see whether these new tools can provide the basis simultaneously to address the improvement of cocoa productivity while maintaining the quality characteristics of the bean that are crucial for chocolate production.

References

- Dillinger, T.L. *et al.* (2000) Foods of the gods: cure for humanity? A cultural history of the medicinal and ritual uses of chocolate. *J. Nutr.* 130, 2057S–2072S
- Kris-Etherton, P.M. and Keen, C.L. (2002) Evidence that the antioxidant flavonoids in tea and cocoa are beneficial for cardiovascular health. *Curr. Opin. Lipidology* 13, 41–49
- Eskes, B. (2001) Introductory notes. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 8–11
- Lockwood, R. (2003) Who needs clothing? *INGENIC Newsl.* 8, 2–5
- Lanaud, C. *et al.* (1995) A genetic linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 91, 987–993
- Risterucci, A.M. *et al.* (2000) A high-density linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 101, 948–955
- Cheesman, E.E. (1944) Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop. Agri.* 21, 144–159
- Cuatracasas, J. (1964) Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib. US Herb.* 35, 379–614
- De la Cruz, M. *et al.* (1995) Origins of cacao cultivation. *Nature* 375, 542–543
- Motamayor, J.C. *et al.* (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89, 380–386
- Crouzillat, D. *et al.* (2000) Quantitative trait loci analysis in *Theobroma cacao* using molecular markers: inheritance of polygenic resistance to *Phytophthora palmivora* in two related cacao populations. *Euphytica* 114, 25–36
- Flament, M-H. *et al.* (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. *Genome* 44, 79–85
- Queiroz, V.T. *et al.* (2003) Identification of a major QTL in cocoa (*Theobroma cacao* L.) associated with resistance to witches' broom disease. *Plant Breed.* 122, 268–272
- Lanaud, C. *et al.* (2001) Implications of new insight into the genetic structure of *Theobroma cacao* L. for breeding strategies. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 89–107
- Saunders, J.A. *et al.* (2001) USDA DNA fingerprinting program for identification of *Theobroma cacao* accessions. *Proceedings of the International Workshop on New Technologies for Cocoa Breeding, Kota Kinabalu, Malaysia*, INGENIC, London, UK, pp. 108–114
- Jones, P.G. *et al.* (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta* 216, 255–264

1360-1385/\$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tplants.2003.10.004

Why are there so many carbohydrate-active enzyme-related genes in plants?

Pedro M. Coutinho, Mark Stam, Eric Blanc and Bernard Henrissat

Architecture et Fonction des Macromolécules Biologiques, UMR 6098, Centre National de la Recherche Scientifique, Universités Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France

Plants contain far more carbohydrate-active enzyme-encoding genes than any other organism sequenced to date. The extremely large number of glycosidase and glycosyltransferase-related genes in plant genomes can be explained by the complex structure of the plant cell wall, by ancient genome duplication and by recent local duplications, but also by the recent emergence of novel and unrelated protein functions based on widely available pre-existing scaffolds.

In plants, carbohydrates in the form of glycosides are central to many biological pathways, from cell wall structure to energy, signalling and defence. Glycosides are made from activated sugars by glycosyltransferases

and are degraded by glycoside hydrolases (glycosidases). Genes encoding glycosidases and glycosyltransferases in all organisms (all data available from the Carbohydrate-Active enZymes server at <http://afmb.cnrs-mrs.fr/CAZY/>) are currently analysed and listed based on the classifications of glycosidases [1] and glycosyltransferases [2] in sequence- and structure-based families. Now that the genomes of many organisms have been completely sequenced, it is possible to analyse the content of genomes from a global glycobiological perspective.

Content of genomes in carbohydrate-active enzyme-related genes

One of the intriguing features of the analysis of the sequenced genomes is that for bacteria and for eukaryotes there appears to be a global correlation between the number of glycosidase and glycosyltransferase-related genes and the

Corresponding author: Bernard Henrissat (bernard.henrissat@afmb.cnrs-mrs.fr).

total number of genes in the organism (Figure 1). The archaea do not show such a correlation and it is likely that many (if not all) of their glycosidases have been acquired by horizontal transfer (i.e. by transfer of genetic material between organisms other than by descent). Figure 1 shows that *Arabidopsis* is a clear outlier with almost 800 glycosidase and glycosyltransferase-related genes comprising >3.3% of its genes [3]. By comparison, the human genome has only ~350 glycosidase and glycosyltransferase-related genes.

Origins of the multiplicity of carbohydrate-active enzyme-related genes in plants

A survey of the sequences deposited in GenBank shows that the rice genome is likely to show a similar bias. So why do plants have so many glycosidase and glycosyltransferase-related genes? There are several complementary explanations to this, the most obvious being that many different glycosidases and glycosyltransferases are probably required for the biosynthesis and degradation of the complex polysaccharides of the plant cell wall. For instance, the biosynthesis of the complex structure of pectin requires the action of at least 53 different enzymatic activities [4]. Another partial answer is with the ancient global genome duplication that was recently established for *Arabidopsis* [5]. Genome duplication is an obvious way to increase the number of genes. An investigation of the chromosomal location of glycosidase and glycosyltransferase-related genes in *Arabidopsis* reveals that, as for other multi-gene families, many glycosidase and glycosyltransferase-related genes occur as clusters on the genome. Interestingly, the genes within clusters also cluster together on phylogenetic trees comprising either only *Arabidopsis* or with all known sequences within a family, indicating that the clusters have been created by extensive recent local duplications. To date, this feature has been encountered only in plants.

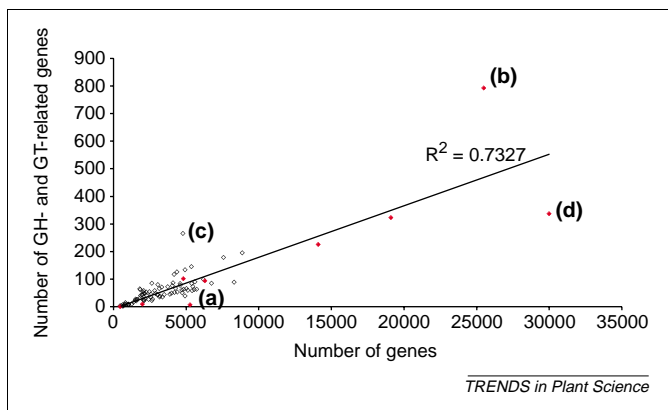


Figure 1. Correlation between the number of glycosidase- (GH) and glycosyltransferase- (GT) related genes and the total number of genes in 86 bacterial and nine eukaryotic genomes. Bacteria are indicated by open diamonds and eukaryotes are indicated by red diamonds. Owing to massive gene loss, parasites tend to appear below the average, for instance, *Plasmodium falciparum* (a), whereas 'excess' is found in *Arabidopsis* (b) and in *Bacteroides thetaiotamicron* (c). Analysis of the complete repertoire of *Bacteroides thetaiotamicron* indicates that it is probably involved in the degradation of otherwise indigestible dietary plant polysaccharides in the human gut. Humans (d) show a deficit, reflecting reliance on gut microorganisms for the digestion of plant polysaccharides. Further details can be obtained from the authors or from the Carbohydrate-Active enZymes database (<http://afmb.cnrs-mrs.fr/CAZY/>).

<http://plants.trends.com>

Although this explains the multiplicity of glycosidase and glycosyltransferase-related genes in plants, another factor is becoming apparent: the recent emergence of novel protein functions, based on glycosidase scaffolds. Several such cases have been identified recently:

- The *chrk1* gene of tobacco encodes a modular protein with an N-terminal domain displaying similarities to chitinases, followed by a transmembrane segment and a C-terminus resembling serine/threonine kinase [6]. Examination of the chitinase-like domain shows that the ChrK1 protein is likely to be devoid of enzymatic activity because it lacks the catalytic amino acid of the related chitinases. Such a protein could, upon recognition of a signal yet to be identified, trigger a cascade of events within the cell, starting with a protein phosphorylation.
- A protein capable of selective inhibition of fungal xylanases has been isolated from wheat [7]. Sequence similarity places it also in the chitinase family and its 3-D structure, recently determined, confirms the structural resemblance to chitinases [8]. Interestingly, the structure shows that the active site cleft of the related chitinases is partially filled by bulky residues resulting in a mis-orientation of the catalytic machinery. The xylanase inhibitor and its chitinase ancestors are believed to be produced by the plant as part of their defence system against fungi. It is tempting to speculate that the novel function emerged based on a class of proteins that was already overexpressed upon fungal attack, and that evolution initially kept the existing signal recognition and expression-regulation pathways.
- The most recently identified example of novel protein function is the soybean hydroxyisourate hydrolase. This enzyme shows extensive sequence similarity to a family of β -glucosidases, including conservation of the known catalytic residues. A recent study elegantly demonstrates that the catalytic residues of hydroxyisourate hydrolase have kept their original roles (acid-base and nucleophile) and yet catalyse a different reaction on a non-carbohydrate substrate [9].

Because the novel proteins are not isofunctional to their glycosidase ancestors, they are hard to characterize and are only now being discovered. Therefore, it is likely that the number of functional plant glycosidases, albeit large, is less than previously thought. The novel proteins that have recently evolved from carbohydrate-active enzymes, not only provide an additional reason for the apparently large number of glycosidase-related genes in plants, but also point toward the immense weaknesses of genomic annotations. Even outside of purely automatic annotations, experts in a particular subfield might notice the absence of catalytic residues, but the example of hydroxyisourate hydrolase shows that the novel function not only has recruited the scaffold, but also the original catalytic machinery. Without functional characterization this protein would have been annotated as a putative β -glucosidase.

The recent acquisition of novel functions based on common scaffolds that is beginning to emerge is general

and not restricted to glycosidases. For instance plant endo- β -1,3-glucanase inhibitors, produced in response to certain fungal pathogens, display remarkable similarity to serine proteases [10]. The acquisition of novel functions based on common scaffolds is not restricted to plants, and might be found in all genomes, including mammalian genomes. An example is *klotho*, a gene first isolated in mouse and whose defect produces a syndrome related to ageing [11]. The protein encoded by *klotho* shows similarities to β -glucosidases (i.e. the same family that also contains the soybean hydroxyisourate hydrolase). Interestingly, the catalytic acid–base residue of β -glycosidases is absent in the *klotho*-encoded protein. Because there is no obvious relationship between the acquired novel function(s) and the pre-existing folds, the recent evolutionary events discussed here make predicting the function of *klotho* (and of many other genes from a variety of organisms including plants) impossible and illustrate the serious limitations of automated computer-based function predictions in genomes.

Acknowledgements

We are particularly thankful to Gideon J. Davies (University of York), Stephen G. Withers (University of British Columbia) and Birte Svensson (Carlsberg Laboratory, Copenhagen) for their tireless encouragements to write this paper. This work was funded by grants QLK1-CT-2000–00811 (GEMINI) and QLK5-CT-2001–00443 (EDEN) from the European Commission.

References

- 1 Henrissat, B. and Davies, G.J. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644
- 2 Coutinho, P.M. *et al.* (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* 328, 307–317
- 3 Henrissat, B. *et al.* (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Mol. Biol.* 47, 55–72
- 4 Mohnen, D. (1999) Biosynthesis of pectins and galactomannans. In *Comprehensive Natural Products Chemistry* (Vol. 3) (Pinto, B.M., ed.), pp. 497–527, Elsevier
- 5 Bowers, J.E. *et al.* (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438
- 6 Kim, Y.S. *et al.* (2000) CHRK1, a chitinase-related receptor-like kinase in tobacco. *Plant Physiol.* 123, 905–915
- 7 Elliott, G.O. *et al.* (2002) Functional identification of the cDNA coding for a wheat endo-1,4- β -D-xylanase inhibitor. *FEBS Lett.* 519, 66–70
- 8 Payan, F. *et al.* (2003) Structural analysis of xylanase inhibitor protein I (XIP-I), a proteinaceous xylanase inhibitor from wheat (*Triticum aestivum*, var. Soisson). *Biochem. J.* 372, 399–405
- 9 Raychaudhuri, A. and Tipton, P.A. (2003) A familiar motif in a new context: the catalytic mechanism of hydroxyisourate hydrolase. *Biochemistry* 42, 6848–6852
- 10 Rose, J.K. *et al.* (2002) Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens. *Plant Cell* 14, 1329–1345
- 11 Kuro-o, M. *et al.* (1997) Mutation of the mouse *klotho* gene leads to a syndrome resembling ageing. *Nature* 390, 45–51

1360-1385/\$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tplants.2003.10.002

Unraveling the functions of glycosyltransferase family 47 in plants

Ruiqin Zhong and Zheng-Hua Ye

Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

The function of glycosyltransferases (GTs) from family GT47 was first identified in animal exostosins as β -glucuronyltransferase involved in the synthesis of heparan sulfate. Two recent papers report the functions of two plant members in this family as a pectin β -glucuronyltransferase and a xyloglucan β -galactosyltransferase. These findings greatly extend our understanding of the biological functions of family GT47 and also represent an important leap toward the molecular dissection of cell wall biosynthesis.

Glycosyltransferases (GTs) (EC 2.4.x.y) are enzymes that catalyze the transfer of sugar moieties from donor molecules (usually nucleoside diphospho-sugars) to specific acceptor molecules, thus forming glycosidic bonds. Based on amino acid sequence similarities, GTs have been classified into at least 65 families [1,2]. Of these, 35 families representing 408 putative GT genes have been identified in the genome of *Arabidopsis thaliana*

(<http://afmb.cnrs-mrs.fr/CAZY/>). The biochemical activities and biological functions of most of the *Arabidopsis* GTs have not been characterized yet [3–5]. GTs catalyze the biosynthesis of disaccharides, oligosaccharides and polysaccharides, and the glycosylation of many proteins and a diverse array of small molecules such as plant hormones, secondary metabolites and herbicides. Plant cell walls are composed of complex polysaccharides including cellulose, hemicelluloses and pectins, which require many GTs for their synthesis. Assigning biochemical activities and biological functions to hundreds of GTs identified in *Arabidopsis* alone will be a huge challenge to plant biologists. The recent identification of two plant members in family GT47 as a pectin β -glucuronyltransferase and a xyloglucan β -galactosyltransferase by Hiroaki Iwai *et al.* [6] and Michael Madson *et al.* [7], respectively, has opened new avenues for investigating the functions of GTs in plants.

Exostosins and family GT47

The family GT47 grouping is based on the β -glucuronyltransferase domain of animal exostosins. In addition to the

Corresponding author: Zheng-Hua Ye (zhye@plantbio.uga.edu).