UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

ECOLE DOCTORALE BIOLOGIE SANTE LILLE NORD DE FRANCE

Thèse de Doctorat / Doctorate Thesis

Ugo Pierre CENCI

En vue de l'obtention du titre de :
Docteur de l'Université des Sciences et Technologies Lille 1

Spécialité :
Génétique Microbienne / Microbial Genetics

La déramification des polysaccharides définit un mécanisme universel et polyphylétique pour la synthèse d'amidon chez les bactéries et les eucaryotes

Polysaccharide debranching defines a universal and polyphyletic mechanism for starch accumulation in bacteria and eukaryotes

18th October 2013

Jury:			
Président	:	**Pr. C. D'Hulst**	Université Lille 1
Rapporteurs	:	**Pr. W. Löffelhardt**	Universität Wien
		Dr. B. Henrissat	DR CNRS Université d'Aix-Marseille
Examinateurs	:	**Pr. Y. Nakamura**	Akita Prefectural University
		Pr. D. Bhattacharya	Rutgers University
		Dr. E. Suzuki	Akita Prefectural University
		Dr. J.-L Putaux	DR CNRS Université Joseph Fourrier
		Dr. C. Colleoni	Université Lille 1
Directeur de thèse	:	**Pr. S.G. Ball**	Université Lille 1

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

ECOLE DOCTORALE BIOLOGIE SANTE LILLE NORD DE FRANCE

Thèse de Doctorat / Doctorate Thesis

Ugo Pierre CENCI

En vue de l'obtention du titre de :
Docteur de l'Université des Sciences et Technologies Lille 1

Spécialité :
Génétique Microbienne / Microbial Genetics

La déramification des polysaccharides définit un mécanisme universel et polyphylétique pour la synthèse d'amidon chez les bactéries et les eucaryotes

Polysaccharide debranching defines a universal and polyphyletic mechanism for starch accumulation in bacteria and eukaryotes

18th October 2013

Jury:			
Président	:	**Pr. C. D'Hulst**	Université Lille 1
Rapporteurs	:	**Pr. W. Löffelhardt**	Universität Wien
		Dr. B. Henrissat	DR CNRS Université d'Aix-Marseille
Examinateurs	:	**Pr. Y. Nakamura**	Akita Prefectural University
		Pr. D. Bhattacharya	Rutgers University
		Dr. E. Suzuki	Akita Prefectural University
		Dr. J.-L Putaux	DR CNRS Université Joseph Fourrier
		Dr. C. Colleoni	Université Lille 1
Directeur de thèse	:	**Pr. S.G. Ball**	Université Lille 1

Résumé

A l'état natif, l'amidon s'agrège sous forme de grains insolubles semicristallins, à la différence des particules de glycogène qui sont hydrosolubles. Chez les eucaryotes photosynthétiques, le passage à l'accumulation de l'amidon est survenu après l'endosymbiose du plaste, à partir d'un réseau métabolique synthétisant du glycogène cytosolique chez l'hôte. Ceci a impliqué le recrutement d'une enzyme de débranchement d'origine chlamydienne, un parasite intracellulaire obligatoire. Cette enzyme de débranchement est considérée comme responsable de l'épissage des branches mal placées au sein d'un polysaccharide précurseur qui, sans ce mécanisme, serait soluble dans l'eau. Nous présentons l'implication de l'enzyme de débranchement dans le processus d'édification de l'amidon dans des cyanobactéries unicellulaires qui accumulent à la fois du glycogène et des grains d'amidon semicristallins. Nous montrons que, dans ces organismes, une enzyme de nature analogue à l'enzyme de débranchement de la plante, mais d'origine bactérienne différente, a été recrutée pour la même fonction. Les enzymes de débranchement des cyanobactéries et des plantes ont évolué séparément à partir d'une enzyme impliquée dans le catabolisme du glycogène pour donner de l'amidon. La nature polyphylétique de la transition de l'accumulation de glycogène à celle d'amidon, chez les Archaeplastida et les cyanobactéries, ainsi que l'évolution convergente de mécanismes biochimiques communs, suggèrent que le débranchement des polysaccharides au cours de leur synthèse est une exigence universelle de l'évolution du métabolisme de l'amidon dans les cellules vivantes.

Mots clefs :

Amidon, Isoamylase, Enzyme de débranchement, Glycogène, Cyanobacterie, Evolution, Archaeplastida, Cyanobacterium sp. CLg, Polysaccharide cristallin.

Abstract

Starch, unlike hydrosoluble glycogen particles aggregates into insoluble, semi-crystalline granules. In photosynthetic eukaryotes, the transition to starch accumulation occurred after plastid endosymbiosis from a pre-existing cytosolic host glycogen metabolism network. This involved the recruitment of a debranching enzyme of chlamydial pathogen origin. The latter is thought to be responsible for removing misplaced branches that would otherwise yield a water-soluble polysaccharide. We now report the implication of starch debranching enzyme in the aggregation of semi-crystalline granules of single-cell cyanobacteria that accumulate both glycogen and starch-like polymers. We show that an enzyme of analogous nature to the plant debranching enzyme but of a different bacterial origin was recruited for the same purpose in these organisms. Remarkably both the plant and cyanobacterial enzymes have evolved a novel yet identical substrate specificity from a preexisting enzyme which originally displayed the much narrower substrate preferences required for glycogen catabolism. The polyphyletic nature of the transition from glycogen to starch in Archaeplastida and cyanobacteria together with the convergent evolution of a common biochemical mechanism suggest that polysaccharide debranching defines a universal requirement for the evolution of starch metabolism in living cells.

Key words:

Starch, Isoamylase, Debranching enzyme, Glycogen, Cyanobacteria, Evolution, Archaeplastida, Cyanobacterium sp. CLg1, Cristalline polysaccharide.

Remerciements

Ce qui me plaît dans les remerciements ; c'est qu'ils peuvent être complètement décousus, comme mon esprit ! Alors à ceux qui liront ces lignes si vous ne comprenez pas tout, tant pis ! Pourvu que vous compreniez ce qui est pour vous.

Tout d'abord je remercie tous les membres de mon jury d'avoir bien voulu prendre du temps pour juger mon travail. I would first like thank my jury members to have accepted this work.
Je remercie particulièrement, les Drs J.-L Putaux et B. Henrissat, pour les échanges très intéressants autour de mon manuscrit.

Ensuite, je dois remercier ceux grâce à qui j'ai attrapé une crampe au poignet, un problème aux yeux… Pour rédiger ma thèse bien évidemment !

« C'est le rôle essentiel du professeur d'éveiller la joie de travailler et de connaître. »
Albert Einstein

Je dédie cette phrase à ceux qui ont su éveiller cette joie, aussi bien :
Her Docteur Professeur Steven G. Ball que Her Docteur Christophe Colleoni.
Mes remerciements vont aussi aux différentes personnes que j'ai croisées pendant cette thèse et qui m'ont accueilli, et aidé à avancer dans ma thèse :
Dr JL Putaux et Amandine Durand Terrasson, Pr D. Bhattacharya et toute son équipe, en particulier Dr Huan Qiu pour l'aide en phylogénie, et Dr M.-M. Perrineau pour les pauses midi et les discussions scientifiques.
Toutefois, je n'oublie pas mes professeurs qui m'ont éveillé -voire réveillé !- durant ma scolarité.
Je remercie aussi Maria Cecilia Arias pour son aide en bioinfo et dans la rédaction de ma thèse, ainsi que Catherine Tirtiaux et Dr Jenifer Nirmal Raj qui ont abattu un énorme travail, et permis cette thèse.
Enfin une pensée pour ceux avec qui j'ai travaillé à l'UGSF.

« Un enfant pourra difficilement transmettre plus tard ce qu'il ne reçoit pas de ses parents. »
P.D James

Donc, avant tout cela, il y a eu ceux qui m'ont préparé lorsque je n'étais que petit garçon, et qui m'ont armé pour cette expédition. Ceux qui m'ont appris à marcher droit devant (Mes

parents et ma famille proche : Pépé, Grand Papa, Mémé, Mamé, Papa, Maman, Gilles, Marion, Fabien, Léo). Sans oublier ceux qui m'ont aidé à corriger mon bout de thèse en français.

Je n'oublie évidemment pas toutes les personnes que j'ai croisées plus récemment et qui font maintenant partie de ma famille (ta famille Gillou, et les pièces rapportées en Jul… qui prennent soin des yeux verts).

Je vous remercie, vous, ma famille algérienne d'ici (y en a pas tant que ça : Sara et Fatiha), et d'au-delà des océans et des mers : celle des Etats Unis - sans laquelle mon passage aux USA aurait été sérieusement compromis : Nasser, Aziza et leurs enfants, et, celle d'Algérie - là, ça fait beaucoup de monde… je n'en nommerai que deux : les parents de Malika : Rabah et Megdouda.

« Aux vrais amis tout est commun »
Euripide

Merci à vous mes amis des différents coins du monde et à ceux ici ; aussi bien les sportifs, que les rigolos de l'UGSF, et avec qui j'ai pu partager de la bonne humeur, et passé du bon temps. Je voudrais aussi faire une spéciale dédicace à ceux avec qui j'ai partagé la pipette ici (Mathieu Ducatez, Laura Boyer, Maud Facon, Justin Findinier, Hande Tuncay, Derifa Kadouche, Jenifer Nirmal Raj).

« A deux vous êtes une équipe qui gagne »
Gilles et Marie Françoise Arcis

Et surtout Malika que j'aime.

Je te remercie, ma femme, toi, qui as permis que j'arrive jusqu'ici : tu fais tant de choses si bien pour moi, veillant au moindre détail... y compris de la coupe de cheveux à l'étincelant de ma bague ! Toi, avec qui je discute science, ce qui nous permet de progresser mutuellement.

Table of Contents

Abréviations

ADN	**:**	Acide DésoxyriboNucléique
ADP-glucose (ADP-G)	**:**	Adenosine DiPhospho-glucose
AMP	**:**	Adenosine MonoPhosphate
Ap	**:**	Amylopectine
Ap-WT	**:**	Amylopectine de la souche WT
Ap-X	**:**	Amylopectine des mutants X
ATP	**:**	Adenosine TriPhosphate
BAM	**:**	β-amylase
BE	**:**	Branching Enzyme (Enzyme de Branchement)
c.-à-d.	**:**	C'est à dire
C1	**:**	Carbone 1 du glucose
C3	**:**	Carbone 3 du glucose
C6	**:**	Carbone 6 du glucose
CAZy	**:**	Carbohydrate Active Enzyme
CBM	**:**	Carbohydrate Binding Module
DLC	**:**	Distribution en Longueur de Chaine
DNS	**:**	Acide DiNitroSalicylique
DP	**:**	Degrés de Polymerisation
DPE (1 or 2)	**:**	Disproportioning Enzyme (1/2) (enzyme disproportionnante)
DSP	**:**	Dual Specificity Phosphatases (Phosphatase à deux spécificités)
DTT	**:**	DiThioThreitol
EGT	**:**	Endosymbiotic Gene Transfer (transfert de gene endosymbiotique)
G-1-P (G1P)	**:**	Glucose-1-Phosphate
GBSS	**:**	Granule Bound Starch Synthase
GH	**:**	Glycoside Hydrolase
Glc	**:**	Glucose
GLG	**:**	Glycogenine
Gly	**:**	Glycogène

GS	**:**	Glycogène Synthase
GWD	**:**	Glucan Water Dikinase
HPAEC-PAD	**:**	High Perfomance Anion Exchange Chromatography-Pulse Amperometric Detection
iDBE	**:**	indirect DeBranching Enzyme (enzyme de débranchement indirecte)
Iso or ISA (1 or 2 or 3)	**:**	Isoamylase
Kav	**:**	Partition coefficient (Coefficient de partage)
kDa	**:**	kiloDalton
LGT	**:**	Lateral Gene Transfer (transfert lateral de gene)
LSF (1 or 2)	**:**	Like-Sex-Four (1or 2)
MET	**:**	Microscope éléctronique à transmission
mM	**:**	Millimolar
nano-LC-MS-MS	**:**	Nano-Liquid Chromatography-Mass Spectrometry- Mass Spectrometry
nm	**:**	nanometer
N-ter	**:**	N-terminal
p/v	**:**	Poids/volume
PAGE	**:**	Poly-Acrylamide Gel Electrophoresis
PATAg	**:**	Periodic Acid Thiosulfate Silver (Ag) proteinate
PCR	**:**	Polymerization Chain Reaction
PEP	**:**	Pyruvate water dikinase
PHO	**:**	Phosphorylase
PHY	**:**	Phytoglycogène
PL-Ap	**:**	Phosphorylase Limit Amylopectin (Phosphorylase Limite Amylopectine)
PL-Gly	**:**	Phosphorylase Limit Glycogen (Phosphorylase Limite Glycogène)
PPDK	**:**	Pyruvate phosphate dikinase
PPS	**:**	Pyruvate phosphate synthetase
PWD	**:**	Phosphoglucane Water Dikinase
RMN	**:**	Résonnance Magnétique Nucléaire
SAR	**:**	Stramenopiles-Alveolées-Rhizaria
SEX4	**:**	Starch Excess 4

SS	**:**	Starch Synthase
SS-ADP	**:**	Starch Synthase utilisant l'Adenosine DiPhospho-Glucose
SS-UDP	**:**	Starch Synthase utilisant l'Uridine DiPhospho-Glucose
UDP	**:**	Uridine DiPhosphate
UDP-G	**:**	Uridine DiPhospho-Glucose
UV	**:**	UltraViolet
WSP	**:**	Water Soluble Polysaccharide (polysaccharide soluble dans l'eau)
WSP-WT	**:**	Water Soluble Polysaccharide from WT
WSP-X	**:**	Water Soluble Polysaccharide from X mutant
WT	**:**	Wild Type
λmax	**:**	longueur d'onde au maximum d'absorbance
µm	**:**	micromètre

Abbreviations

ADP-glucose (ADP-G)	**:**	Adenosin DiPhospho-glucose
AMP	**:**	Adenosin MonoPhosphate
Ap	**:**	Amylopectin
Ap-WT	**:**	Amylopectin from WT strain
Ap-X	**:**	Amylopectin from X mutants
ATP	**:**	Adenosin TriPhosphate
BAM	**:**	β-amylase
BE	**:**	Branching Enzyme
C1	**:**	Carbon 1 of glucose
C3	**:**	Carbon 3 of glucose
C6	**:**	Carbon 6 of glucose
CAZy	**:**	Carbohydrate Active Enzyme
CBM	**:**	Carbohydrate Binding Module
CL	**:**	Chain length
CLD	**:**	Chain Length Distribution
DNA	**:**	DesoxyriboNucleic Acid
DNS	**:**	DiNitroSalicylique Acid
DP	**:**	Degree of Polymerization
DPE (1 or 2)	**:**	Disproportioning Enzyme (1or 2)
DSP	**:**	Dual Specificity Phosphatases
DTT	**:**	DiThioThreitol
EGT	**:**	Endosymbiotic Gene Transfer
G-1-P (G1P)	**:**	Glucose-1-Phosphate
GBSS	**:**	Granule Bound Starch Synthase
GH	**:**	Glycoside Hydrolase
Glc	**:**	Glucose
GLG	**:**	Glycogenin
Gly	**:**	Glycogen
GS	**:**	Glycogen Synthase
GWD	**:**	Glucan Water Dikinase
HPAEC-PAD	**:**	High Perfomance Anion Exchange Chromatography-Pulse Amperometric Detection

iDBE	**:**	indirect DeBranching Enzyme
Iso or ISA (1 or 2 or 3)	**:**	Isoamylase
Kav	**:**	Partition coefficient
kDa	**:**	kiloDalton
LGT	**:**	Lateral Gene Transfer
LSF (1 or 2)	**:**	Like-Sex-Four (1or 2)
mM	**:**	Millimolar
nano-LC-MS-MS	**:**	Nano-Liquid Chromatography-Mass Spectrometry- Mass Spectrometry
nm	**:**	nanometer
NMR	**:**	Nuclear Magnetic Resonance
N-ter	**:**	N-terminal
PAGE	**:**	Poly-Acrylamide Gel Electrophoresis
PATAg	**:**	Periodic Acid Thiosulfate Silver (Ag) proteinate
PCR	**:**	Polymerization Chain Reaction
PEP	**:**	Pyruvate water dikinase
PHO	**:**	Phosphorylase
PHY	**:**	Phytoglycogen
PL-Ap	**:**	Phosphorylase Limit Amylopectin (Phosphorylase Limit Amylopectine)
PL-Gly	**:**	Phosphorylase Limit Glycogen (Phosphorylase Limit Glycogène)
PPDK	**:**	Pyruvate phosphate dikinase
PPS	**:**	Pyruvate phosphate synthetase
PWD	**:**	Phosphoglucan Water Dikinase
TEM	**:**	Transmission Electronic Microscopy
SEX4	**:**	Starch Excess 4
SS	**:**	Starch Synthase
SS-ADP	**:**	Starch Synthase using Adenosine DiPhospho-Glucose
SS-UDP	**:**	Starch Synthase using Uridine DiPhospho-Glucose
UDP	**:**	Uridin DiPhosphate
UDP-G	**:**	Uridin DiPhospho-Glucose
UV	**:**	UltraViolet
WSP	**:**	Water Soluble Polysaccharide

WSP-WT	**:**	Water Soluble Polysaccharide from WT
WSP-X	**:**	Water Soluble Polysaccharide from X mutant
WT	**:**	Wild Type
w/v	**:**	Weight/volume
λmax	**:**	Length wave at maximum absorbance
µm	**:**	micrometer

Foreword: Description of my work

This thesis is composed of 5 distinct parts: a general introduction and summaryof the results in French, a general introduction in English, an experimental description in English, a general discussion in English, annexes.

The English general introduction is entitled "Transition from glycogen to starch metabolism in Archaeplastida". It consists of the manuscript version that was accepted by Trends in Plant Sciences for which I am first author and that is planned for publication in the December 2013 issue. The goal of this work was to find out if acquisition of starch metabolism in Archaeplastida can be viewed as a single event by the common ancestor of Archaeplastida or not. Turning glycogen into starch metabolism entails two distinct events that must have occurred simultaneously. The first is to generate amylopectin into starch granules. The second was to gain the ability to degrade such structures (eukaryotic enzymes of glycogen metabolism are notoriously unable to do so). My thesis supervisors (Pr Steven Ball and Dr Christophe Colleoni) sent me to Rutgers University in Pr Debashish Bhattacharya laboratory to make a more detailed and focused phylogeny not only of the GlgX–derived enzymes, but also of iDBE (indirect debranching enzyme) and of the dikinases. The scenario detailed in Trends in Plant Sciences calling for a unique transfer and gain of isoamylase from chlamydia intracellular pathogens was born out of the discussions between me and Pr. Ball that followed this phylogenetic analysis. This scenario assumes that the mechanism of amylopectin crystallization is the same in red algae and glaucophytes as in Chloroplastida (green algae and land plants). The gene distribution of isoamylase and the description of a multimeric high mass enzyme with analogous substrate preferences in the glaucophyte cytosol (Plancke *et al.,* 2008) are certainly in favor of such an assumption. However this very reasonable hypothesis remains to be proven by experiments analogous to those described in this thesis for cyanobacterial starch metabolism which entails several years work. The second part of the Trends in Plant Sciences paper dealing with the appearance of the glucan water dikinases was entirely written under the responsibility of Pr Martin Steup (together with Berge A Minassian and Felix Nitschke) and is in complete agreement with the dikinases phylogeny. Since the establishment of such enzyme (GWD and PWD) phylogenies was central to proposing a unified simple and parsimonious explanation for the appearance of starch metabolism in Archeaplastida and complete, with the debranching enzyme, the story of starch raises in eukaryotes, Pr Steven Ball proposed that would be the first author of this collaborative piece of work. The opinion format is considered by Trends in Plant Sciences as

a mixture of literature review and original and interesting speculations unfit for reviews. Since 3 different literature surveys have recently appeared on the topic described in this thesis we thought that the Trends in Plant Sciences opinion would be a fitting thesis introduction that is indeed tightly related to our experimental topic (evolution of starch from glycogen metabolism).

The experimental part of the thesis is the original first version of our submission to the Plant Cell augmented by two additional small paragraphs. The first deals with a sequence comparison revealing the facultative nature of a particular loop in determining the GlgX versus isoamylase substrate specificities (p.110). The second covers the analysis of the CL distribution of glycogen or amylopectin incubated overnight with purified *Cyanobacterium* CLg1 GlgX2 enzyme (p.105). This first version, presented here, has been rejected but a second version was accepted thanks to the additional data performed from Pr Yasunori Nakamura's laboratory describing the substrate preferences of recombinant isoamylases and GlgX proteins (demanded by the reviewers for resubmission). In this manuscript version, I did all the experimental work described with the assistance of many and thus with the following restrictions:

1) The mutants were selected by Catherine Tirtiaux-Ball and Dr Jenifer Nirmal Raj by iodine vapor selection. The mutagenesis screening and clone purification was a 4 year work effort. I took the purified clones for all characterizations reported.
2) The genome sequence of Cyanobacterium CLg1 was determined by the laboratories of Dr Eiji Suzuki and Pr Yasunori Nakamura (Akita Prefectoral University-Japan) and was previously published by these researchers (Suzuki *et al.,* 2013). This sequence was central to our success in finding the precise nature of the defective gene in our glycogen accumulating mutants. Dr Maria-Cecilia Arias helped me in analyzing this genome sequence with respect to starch-glycogen metabolism.
3) The NMR samples were prepared by myself but were analyzed in detail by Dr Emmanuel Maes.
4) The TEM analysis of cells and starch or glycogen samples were done either by Amandine Durand Terrasson or by Dr Jean-Luc Putaux (CERMAV Grenoble). I went to Grenoble to actively help them with the sample preparation and staining.

5) The purification of the GlgX2 enzyme and the trypsic digest of the pure sample were performed entirely by myself but its analysis by Mass Spectroscopy were carried out by Dr Anne-Sophie Vercoutter Edouart (UGSF Lille).
6) Molecular modeling of the enzyme structure was performed by Dr Lyann Sim and Pr Monica Palcic (Carlsberg laboratory) who were kind enough to share their knowledge of the structure of the Chlamydomonas ISA1 enzyme.

In addition to this I included two papers in annex where I appear as coauthor:

The Darwin review published in Journal of Experimental Biology. Pr Steven Ball did all the writing on this one and coauthorship was given to acknowledge the active theoretical discussions during our lab seminars. My contribution on this effort is thus marginal but I have nevertheless enclosed the article as it gives a very useful overview of the evolution of storage polysaccharide metabolism to the thesis readers.

The Perspective published in "The Plant Cell" is an important contribution proposing that chlamydiales have participated directly in a tripartite symbiosis that triggered plastid endosymbiosis. This perspective was also the topic of a 3 page-long editorial by David Baum the editor of The Plant Cell responsible for evolutionary biology. In this MS I have participated (together with Dr Christophe Colleoni) to the cloning of the full or near to full *Cyanophora paradoxa* isoamylases whose sequences were very partial in the *C. paradoxa* genome sequence. This manuscript (and the editorial) is also given as this work was not covered by the "Darwin" review and is of importance to our understanding of the debranching enzyme phylogeny.

Finally there are important experiments that I carried out and which are not described in this thesis document. These concern the characterization of mutants defective for soluble starch synthase in *Cyanobacterium*. These experiments will be pursued by Derifa Kadouche and Matthieu Ducatez. These results are too preliminary to be disclosed in full in a publically accessible document.

Aim of this study:

A major breakthrought in our understanding of the starch crystalization process in the Chloroplastida (green algae and plants) arose from the characterization of chlamydomonas mutants, which substitute starch granules by glycogen biosynthesis. Surprinsly this phenotype was correlated with the lost of a α-1,6 hydrolytic enzyme or isoamylase-type debranching enzyme (Mouille et al., 1996 and Ball et al., 1996), and similar mutants were described in different plant species such as maize, rice and arabidopsis. Ball and collaborators published in 1996 the "glucan trimming model". This model proposed that isomylases are implicated in the splicing of the short branched glucans that interfer with the formation of α-glucan double helices of and, consequently, with the crystallization of amylopectin. Although this model was debated for a long time, it is now widely accepted.The recent discovery of starch-accumulating cyanobacteria rises the following question: Does glucan trimming model operate in these organisms? (Nakamura et al., 2005 and Deschamps et al., 2008). Indeed, if early studies suggest that isoamylase found in Archaeplastida had a cyanobacterial origin (Deschamps et al., 2008), which would explain the transmission of starch metabolism through the primary plastid endosymbiosis, recent phylogenetic analysis have shown that Achaeplastida isoamylases of are originated from chlamydia species (Moustafa et al., 2008 and Ball et al., 2013).

In order to understand the synthesis of crystalline polysaccharides in these cyanobacteria, UV mutagenesis was carried out and mutants were screened after iodine vaporization of cell patches. Mutants of the *Cyanobacterium* sp. CLg1 allow us to dissect the mechanism of synthesis of semi amylopectin in these organisms. In addition, an enzymology approach, coupled with a phylogenetic analysis of enzymes responsible for polysaccharide crystallization in organisms, will determine the common mechanism for synthesizing a crystalline polysaccharide. Through this study, we were able to determine the key mechanisms for cyanobacterial starch synthesis and trace the evolution of crystalline polysaccharides metabolism in these organisms.

RESUME EN FRANCAIS

4. Introduction générale

L'amidon est la forme principale de mise en réserve du dioxyde de carbone atmosphérique chez les plantes terrestres. Localisé dans les tubercules, les graines et les légumineuses, l'amidon représente la base de l'alimentation humaine, et fut un pivot pour le développement et l'évolution de l'espèce humaine (Carmody et Wangham 2009). Ce qui par voie de conséquence, a entraîné l'augmentation des cultures céréalières, qui aujourd'hui sont une des ressources les plus importantes de l'alimentation humaine. Ces ressources qui atteignaient, déjà en 2010, 2,775 milliards de tonnes produites à travers les cultures de maïs, blé, riz, manioc et pomme de terre (FAOSTAT data, 2010 : http://faostat.fao.org), devraient, au regard des prévisions atteindre les 3 milliards de tonnes en 2050 (*Charvet J.P.* Produire pour nourrir les hommes 2009). Par ailleurs, l'amidon est aussi nécessaire dans l'industrie; notamment du papier, ou encore dans la production d'éthanol. Mais s'il est primordial dans nos civilisations, il n'est pas le seul polysaccharide de réserve retrouvé dans le monde vivant. En effet, le polysaccharide de réserve le plus largement distribué chez les êtres vivants est le glycogène. Chez l'Homme des déficiences dans la synthèse du glycogène peut entraîner de nombreuses pathologies telles que le syndrome de Mc ARDLE, qui affecte le glycogène présent dans le muscle (Lucia *et al.,* 2012).

Nous comprenons donc, à travers tous ces exemples, l'importance de la recherche sur l'amidon et plus largement sur les polysaccharides de réserve

Dans l'optique de comprendre les métabolismes de synthèse du glycogène et de l'amidon, d'identifier les mécanismes nécessaires à la synthèse « correcte » de ces deux polysaccharide, et leurs variations métaboliques entrainant les différences structurales, nous nous sommes intéressés, durant ma thèse, à décrire la voie de biosynthèse de l'amidon, et à comprendre son évolution, chez les cyanobactéries ou ses deux mécanismes sont intriqués. Si la synthèse des polysaccharides de réserve semble avoir été en grande partie résolue dans certaines classes d'organismes, les processus impliqués dans certaines classes plus exotiques accumulant de l'amidon sont encore mal connu, et pourraient par la même permettre de résoudre certaines controverse sur des liens phylogénétique entre différents groupes. En effet, l'amidon n'est retrouvé que dans un groupe restreint d'organismes issus d'endosymbioses plastidiales, notamment, le supergroupe des *Archaeplastida* dont les plantes vertes et algues vertes font partie. Ces organismes eucaryotes photosynthétiques proviennent de la mise en

place d'une endosymbiose dite « primaire » entre une cyanobactérie et une cellule eucaryote hétérotrophe. Cette théorie corrélait avec l'observation de grains d'amidon chez les cyanobactéries de l'ordre des C*hroococales* (Nakamura *et al.,* 2005, Deschamps *et al.,* 2008c), l'un des ordres les plus proches phylogénétiquement du plaste retrouvé chez les *Archaeplastida* (Deschamps *et al.,* 2008c, et Gupta et Mathews 2010). Cette description par Nakamura *et al.* 2005 nous a permis, au début de ma thèse, de nous poser la question du métabolisme de synthèse de l'amidon chez ces dernières, et, de l'implication des cyanobactéries dans le métabolisme de l'amidon chez les *Archaeplastida,* afin de comprendre quels sont les facteurs déterminants pour retrouver l'amidon chez les organismes vivants. Cependant une théorie linéaire de l'héritage du métabolisme de l'amidon a été destabilisé, récemment, par des analyses phylogénétiques semblant montrer qu'un ancêtre de l'ordre bactérien des *Chlamydiae* aurait participé à l'établissement de cette endosymbiose, en agissant sur le métabolisme des polysaccharides de réserves chez l'eucaryote hôte (Moustafa *et al.,* 2008 et Ball *et al.,* 2013).

Dans le but d'éclairer ces travaux, nous nous proposons de réaliser un bref état de l'art qui sera construit ainsi :

- Tout d'abord, nous exposerons les principales théories sur la phylogénie des organismes issus d'endosymbiose du plaste.
- Puis, une introduction structurale des deux polysaccharides de réserves les plus abondants dans le monde vivant : le glycogène et l'amidon.
- Ensuite, nous discuterons du modèle cyanobactérien, en nous appuyant sur la classification des cyanobactéries et l'origine du plaste, afin de comprendre les régulations et la dépendance de la synthèse des polysaccharides vis-à-vis d'autres métabolismes.
- Enfin, nous réaliserons un résumé des connaissances sur le métabolisme de l'amidon en ciblant deux types d'enzymes particulièrement importantes dans la synthèse d'amidon.

5. Chapitre : évolution

2.1) Les endosymbioses et leurs rôles dans l'évolution du vivant

L'endosymbiose est le terme donné à l'événement évolutif ayant permis l'incorporation d'un être vivant par un autre être vivant, très probablement par phagocytose (Bhattacharya *et al.,* 2012). L'être vivant interné (endosymbionte) sera alors retenu à l'intérieur de l'hôte et intégré à des niveaux différents, allant jusqu'à des niveaux permanents, donnant alors naissance à un organite tel que la mitochondrie ou le plaste, nous parlerons alors d'organellogénèse. La caractéristique principale retenue pour parler d'organite sera alors la capacité d'envoyer des protéines codées par l'hôte à l'endosymbionte grâce à différents systèmes (pour revue Gould *et al.,* 2008), comme le système de translocation Tic-Toc chez les Archaeplastida (pour revue Jarvis et Soll 2001). Ce sont d'ailleurs ces systèmes de translocation des protéines particulièrement difficiles à mettre en place qui sont considérés comme limitants dans l'établissement d'endosymbiose. D'autres échanges peuvent toutefois exister, comme les échanges de métabolites et de gènes (EGT : endosymbionte gene transfert), mais ils ne définissent pas un critère pour démontrer une organellogénèse.

Il existe plusieurs endosymbioses référencées, comme celle ayant donné naissance à la mitochondrie chez les eucaryotes, qui est considérée comme un événement unique (Andersson et Kurland, 1999), mais les endosymbioses du plaste existent également. Celles-ci ont permis à différentes lignées d'eucaryotes d'acquérir, entre autres, la photosynthèse (Mereschkowsky 1905), et sont au moins au nombre de 5 ; deux endosymbioses primaires du plaste différentes et au minimum trois secondaires, ainsi que les tertiaires, qui pour certaines pourraient être juste de la kleptoplastidie[1] (Keeling, 2013).

[1] Kleptoplastidie : Capacité à récupérer un plaste après phagotrophie d'un organisme eucaryote photosynthétique

2.2) L'endosymbiose primaire du plaste ayant donné naissance aux *Archaeplastida*

Cette endosymbiose la plus connue, et surtout la mieux référencée, est surement l'une des plus importantes en terme d'évolution, puisqu'elle a permis l'acquisition de la photosynthèse par les eucaryotes et en partie le développement de la vie comme nous la connaissons aujourd'hui. Mais, cette endosymbiose unique qui a eu lieu il y a 1,6 milliard d'années (Yoon *et al.,* 2004), a été pendant de nombreuses années source de débat, comme on peut le lire dans les deux études publiées -dans le même numéro de la même revue en 1981-, par Whatley qui soutient l'hypothèse de multiple endosymbiose, et par Gibbs soutient un événement unique d'endosymbiose. Il faut bien évidemment prendre en compte, qu'à l'époque, et pendant longtemps, les analyses essentiellement morphologiques induisaient une confusion de certains organismes, causant des conclusions erronées. Mais, c'est l'avènement du séquençage, et de la phylogénie qui mis un terme à ce débat et révéla le caractère unique de cette endosymbiose (Moreira *et al.,* 2000, Rodriguez-Ezpeleta *et al.,* 2005). Ces méthodes ont ainsi permis d'asseoir, de manière certaine, les trois lignées issues directement de cette endosymbiose, d'ailleurs considérée comme le moment clé dans leur évolution. Ces trois lignées sont les Glaucophytes, les Rhodophytes et les Chloroplastida qui forment tous les trois le phylum des *Archaeplastidae*. Si le caractère unique de cet événement a été mis en évidence, il reste toutefois à comprendre les mécanismes mis en jeu lors de l'établissement de cette endosymbiose. Dans ce contexte, de nombreuses études sont développées autour de transporteurs ancestraux servant à échanger des molécules de première importance comme l'ADP-glucose (métabolite typiquement bactérien) ; permettant l'import de l'énergie de la photosynthèse, effectuée dans la cyanobactérie, vers l'organisme hôte (Weber *et al.,* 2006 et Colleoni *et al.,* 2010). D'autres études portant sur les différentes voies métaboliques dans les potentiels ancêtres communs, sont également menées. Des récents travaux réalisés sur le métabolisme des polysaccharides de réserves, chez les *Archaeplastida*, ont montré une mosaïque d'origine phylogénétique pour ces enzymes. En effet, Deschamps *et al.,* (2008c) montraient qu'une partie des gènes du métabolisme des polysaccharides était d'origine eucaryote, mais qu'une autre était d'origine cyanobactérienne. Parallèlement, cette étude démontrait que certaines cyanobactéries pouvaient synthétiser un polysaccharide cristallin, avançant déjà l'idée du rôle pivot du métabolisme des polysaccharides de réserves dans l'établissement de l'endosymbiose primaire du plaste. Malheureusement, cette mosaïque a été pendant longtemps simplifiée, ne faisant entrer dans l'équation que les deux partenaires évidents : l'ancêtre du plaste (l'ancêtre de certaines cyanobactéries actuelles) et l'eucaryote

hôte (un ancêtre des amibes), faussant ainsi les analyses qui pourtant montraient un autre signal bactérien différent du signal cyanobactérien. Ce signal a été, tout d'abord, révélé dans trois études successives indépendantes (Tyra *et al.,* 2007, Huang et Gogarten 2007, Moustafa *et al.,* 2008). Par la suite, les nouvelles analyses ont réattribué l'origine phylogénétique de nombreux transporteurs et de nombreuses protéines, comme l'isoamylase retrouvée chez les chloroplastida, (Ball *et al.,* 2011) aux chlamydiales - des bactéries parasites intracellulaires obligatoires. Ces réattributions ont fait et font encore débat, mais l'étude du génome de *Cyanophora paradoxa* par Price *et al.,* 2012, et celles permises par la disponibilité de ce génome, devraient y mettre peu à peu un terme en confirmant la viabilité de cette empreinte chlamydienne chez les Archeaplastida. Même si les études historiques (Huang et Gogarten 2007, Moustafa *et al.,* 2008) expliquaient déjà la présence des chlamydiae et leur rôle dans l'établissement de l'endosymbiose primaire du plaste, ce sont les études récentes, et la coopération de plusieurs groupes qui semblent, aujourd'hui, démontrer que l'établissement de l'endosymbiose n'a pu se faire que grâce à un ménage à trois : de la cyanobactérie ancestrale, de l'eucaryote hôte, et du chlamydiae (Ball *et al.,* 2013). Cette hypothèse repose sur le détournement du métabolisme des polysaccharides de réserves, et donc de l'énergie par les chlamydiales lors de l'infection de la cellule hôte. Ceci conduisant à la nécessité de garder pour l'hôte, un organisme capable de fournir de manière importante de l'énergie par la photosynthèse : la cyanobactérie. Cette hypothèse pour expliquer l'établissement de l'endosymbiose bien que simple et séduisante doit encore être étayée et fait ainsi l'objet de différents projets.

2.3) Les endosymbioses secondaires

L'endosymbiose secondaire du plaste est un événement phototrophe entre deux eucaryotes, un eucaryote photosynthétique (issu de l'endosymbiose primaire du plaste) étant internalisé par un autre eucaryote. Pour l'endosymbiose secondaire, de nombreuses controverses divisent les spécialistes, ne les laissant s'accorder que sur très peu de points. D'une part, au moins trois endosymbioses secondaires ont donné un plaste avec comme origine un *Chloroplastidae*, et au moins une autre implique une algue rouge internalisée qui a donné un plaste. D'autre part, les plastes issus de ces endosymbioses possèdent entre 3 et 4 membranes, qui sont en partant du plaste : les deux membranes de la cyanobactérie, la membrane de l'algue internalisée, puis enfin la membrane qui dérive de la vacuole de l'hôte

ayant internalisé l'algue (Keeling, 2004). Ici, nous nous focaliserons uniquement sur la ou les endosymbioses d'algues rouges, qui synthétisent des alpha glucanes de réserve et pour certaines de l'amidon.

2.4) L'endosymbiose secondaire des algues rouges

La ou les endosymbioses secondaires des algues rouges ont donné naissance à plusieurs classes, qui ne font pas l'unanimité. En effet, plusieurs scénarios évolutifs ont été imaginés. Il existe un ensemble d'organismes dits « issus d'endosymbioses secondaires d'algues rouges », ce qui signifie que ces organismes possèdent un plaste dérivé d'une algue rouge, mais cet ensemble d'organismes est très hétérogène appartient à différents supergroupes. Dans les différents supergroupes, nous retrouvons des lignées avec un plaste dérivé des algues rouges réalisant la photosynthèse comme les diatomées, des lignées avec un plaste dérivé des algues rouges, mais non photosynthétiques (apicomplexe), mais également des lignées ayant soit totalement perdu leur plaste -comme les colpodellids, retrouvés chez les apicomplexes- soit n'ayant, peut-être, jamais possédé de plaste (Figure 1).

Si historiquement ces organismes ont été classés ensemble : Ciliés et Apicomplexes (Alvéolés), Stramenopiles et Haptophytes du fait de la présence de chlorophylle c, et d'un plaste placé au niveau du reticulum endoplasmique rugueux dans certain cas (Cavalier-smith 1981), dans les chromalvéolés (Cavalier-smith 2004, Adl *et al.,* 2005), les nouvelles données apportées, d'une part, par le séquençage haut débit rendant disponible de nombreux génomes, et d'autre part, par les nouvelles méthodes de phylogénie, ont permis de mettre en évidence la proximité des Rhizarias avec les Stramenopiles et les alvéolés (Burki *et al.,* 2007), donnant ainsi naissance au SAR (Stramenopiles, Alvéolés, Rhizaria) (pour revue Adl *et al.*, 2012). Nous savons parallèlement que des études menées par Hackett *et al.,* (2007), ont démontré la faible union des Haptophytes et des Cryptophytes (classé ensemble dans le supergroupe des Hacrobia) avec les SAR, et nous soulignons, avec intérêt, que la classification regroupe les ciliés avec les apicomplexes, les chromerides et les dinoflagellés, alors qu'ils ne sont pas pour autant toujours classés comme organismes ayant subi une endosymbiose secondaire (Reyes Prieto *et al.,* 2008).

Ainsi la théorie des Hacrobia est mise en exergue (figure 1), pour déstabiliser l'hypothèse des chromalvéolés initiales, impliquant peut-être plusieurs endosymbioses secondaires, puisque les Hacrobia sont potentiellement proches des Archeaplastides dans l'arbre de la vie (figure 1 pointillés noirs). Mais ce supergroupe n'est pas encore parfaitement

établi ; en effet, l'hypothèse des Hacrobia repose, à la base, principalement sur le lien entre le plaste des Haptophytes et des Cryptomonades (Rice et Palmer 2006, Patron *et al.,* 2007). De plus, les Hacrobia ne sont pas toujours situés loin du supergroupe des SAR, indiquant peut-être une origine commune.

Une autre donnée troublante vient aussi déstabiliser notre compréhension de l'endosymbiose secondaire des algues rouges (Moustafa *et al.,* en 2009), démontrant une empreinte de gènes de Chloroplastida dans le génome des diatomées. Cette même étude explique ce signal par la présence cryptique d'une algue verte. Il nous faut toutefois tempérer ces conclusions, car la possibilité d'une sous-représentation des données sur les algues rouges ayant pu conduire à une mauvaise conclusion ne peut être écartée (Deschamps et Moreira 2012). Il est toutefois intéressant de considérer que ce signal permet différentes hypothèses.

La plus répandue est basée sur un regard neuf de l'évolution qui n'implique pas un maximum de parcimonie à une endosymbiose unique rouge, mais plutôt à une endosymbiose unique verte qui se serait effectuée de manière précoce, chez l'ancêtre de tous les organismes ayant subi une endosymbiose secondaire du plaste (Hacrobia et SAR) (Dorrell et Smith 2011). Ainsi la présence d'une empreinte de gènes de Chloroplastida chez les lignées issues de l'endosymbiose secondaire (Moustafa *et al.,* 2009), mais aussi les proximités phylogénétiques des Hacrobia et des SAR pourraient être due à la présence d'un plaste vert, acquit par endosymbiose secondaire, dans l'ancêtre commun de ces deux lignées (Hacrobia et SAR). Cette endosymbiose précoce aurait facilité ainsi, par des systèmes d'import-export de protéines et de métabolites déjà présents, le remplacement par l'endosymbiose tertiaire d'une algue rouge, possédant grâce à sa photosynthèse un avantage sélectif pour les périodes géologiques avec une basse concentration en CO_2.

Une autre question reste aussi en suspens dans les nombreux organismes appartenant à la même classe, mais ne possédant pas de plastes: ont-ils subi une endosymbiose secondaire ou sont-ils juste très proches de ces organismes l'ayant subi ? Cette question longtemps débattue pourrait avoir un intérêt crucial pour déterminer si l'endosymbiose a été un pivot majeur de ces différents supergroupes, ou au contraire, si elle n'est qu'un avantage sélectif pour une partie des organismes.

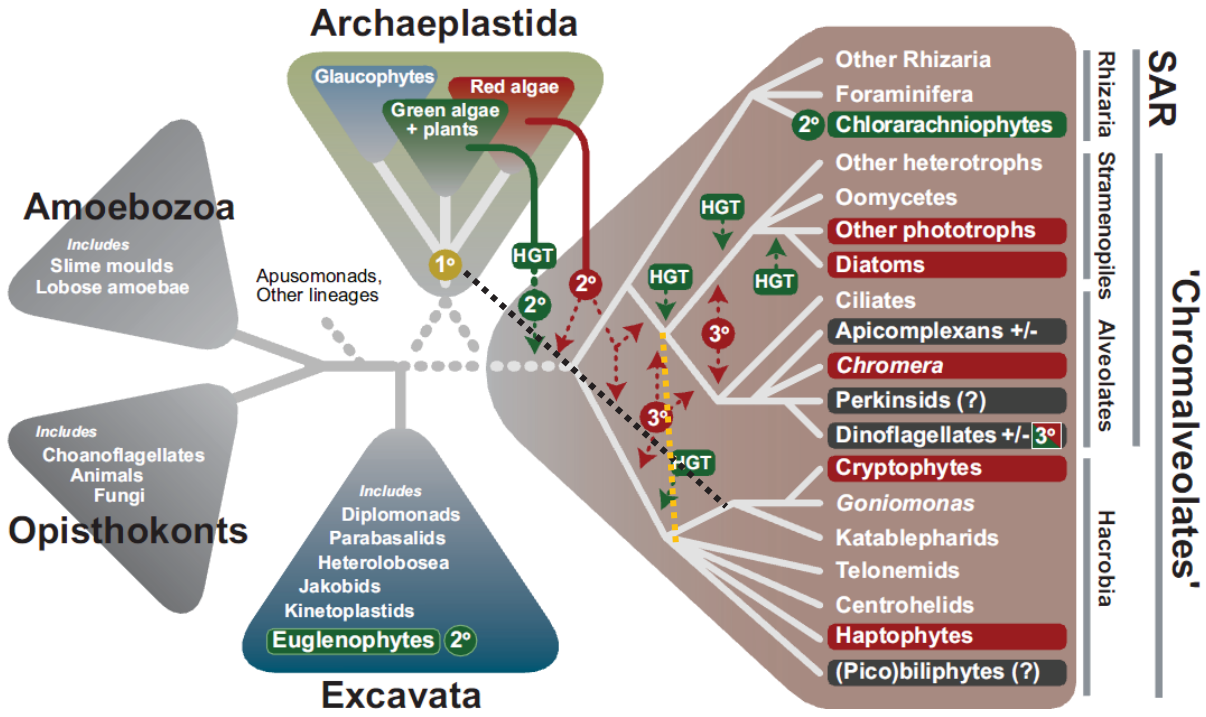


Figure 1 : Arbre du vivant des eucaryotes avec les différents supergroupes. L'origine de la photosynthèse chez la majorité des eucaryotes peut être reliée à l'endosymbiose primaire du plaste ayant donné naissance aux *Archaeplastida* (sauf pour Paulinella). Les supergroupes sont liés entre eux et particulièrement les organismes issus des endosymbioses plastidiales. L'arbre montre le regroupement ancien des chromalvéolés et le nouveau groupe bien accepté des SAR. Les pointillés gris indiquent une direction d'évolution possible. Les pointillés noirs indiquent une proximité probable entre *Archaeplastida* et Hacrobia, principalement cryptophyte, tandis que les pointillés orange indiquent un lien possible entre Haptophytes et SAR (Burki *et al.*, 2012). 1°,2°,3° indiquent respectivement les endosymbioses primaires, secondaire ou tertiaires, et la couleur indique l'origine du plaste, les HGT (horizontal gene transfer) indiquent des empreintes de gènes verts dans les génomes. (?) indiquent une possible présence de plaste. +/- indiquent que ces lignées sont représentées par des espèces possédant de temps en temps un plaste, et de temps en temps n'en présentant pas. Le 3° souligné en vert et rouge indique la présence d'un plaste issu de l'endosymbiose tertiaire pouvant être rouge ou vert. La longueur des branches ne tient pas compte des distances évolutives. (Adapté de Elias et Archibald, 2009)

6. Chapitre : structures des alpha-glucanes de réserve

3.1) Introduction générale sur les alpha-glucanes de réserves

Il existe de nombreuses formes de polysaccharide de réserves chez les eucaryotes et les bactéries, en particulier, nous trouvons des β-glucanes (Paramylon, β-glucane) et des α-glucanes comme le granulose (retrouvé chez les *Clostridium*), le glycogène et l'amidon. Ici, nous nous attarderons sur ces deux derniers. Remarquons que de nombreuses classes d'organismes phylogénétiquement distinctes, ainsi que de nombreuses espèces de bactéries accumulent du glycogène, alors que seules les lignées issues de l'endosymbiose primaire, secondaire ou tertiaire du plaste semblent présenter de l'amidon. Le glycogène et l'amidon se caractérisent tous les deux par un sucre identique, le glucose, et des liaisons entre ces glucoses identiques : les liaisons en alpha 1-4, qui formeront des glucanes linéaires et les liaisons en alpha 1-6, appelées point de branchements ou ramifications qui formeront des glucanes ramifiés (figure 2). Nous pourrions donc penser que ces deux structures ont des propriétés proches, mais en réalité leurs propriétés physico-chimiques diffèrent. Par conséquent, il est normal de s'intéresser aux différences de structure fine responsable de cela conduisant aux différences de propriétés de ces molécules.

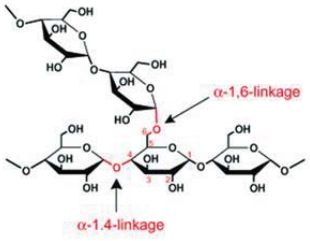


Figure 2 : Formule semi-développée d'un α-1,4 glucane ramifié une fois en α-1,6. La nature des liaisons entre les résidus de glucose est indiquée (Roach *et al.*, 2012).

3.2) Le glycogène

Le glycogène est un polysaccharide composé de glucose lié en position alpha 1-4 et ramifié en alpha 1-6, retrouvé sous forme de petite particule hydrosoluble possédant une structure très facilement métabolisable, due à son organisation. Le glycogène est un polysaccharide ayant des chaînes de glucose liées en alpha 1-4. Leurs tailles moyenne est de 12 ou 13 résidus glucose, c'est-à-dire d'un degré de polymérisation de 12 ou 13, réparti en 12 niveaux théoriques (figure 3A). Ces chaînes se classifient, ensuite, en deux sous-fractions : les

chaînes β qui supportent deux points de branchement en moyenne par chaîne, situées sur les résidus 5-6 et 9-10 (ces deux chaînes ayant une direction opposée grâce au résidu ou se fixe les points de branchement). Ces chaînes β se retrouvent à l'intérieur de la molécule, tandis que les chaînes α, qui ne supportent pas de points de branchement, se localisent dans la couche externe du glycogène.

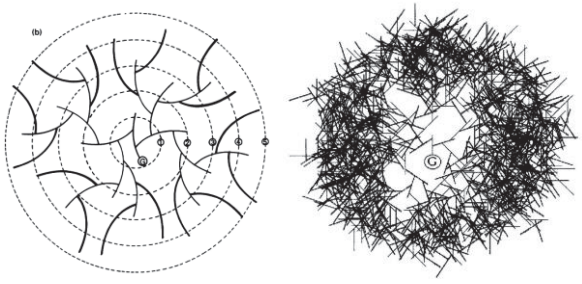


Figure 3 : A. Schéma de l'organisation en couche du glycogène selon Mélendez *et al.,* 1993. Ici sont représentées les cinq premières couches, délimitées par les pointillés, avec au centre la glycogénine. B. Schéma de l'augmentation exponentielle du nombre de chaînes dans les couches externes, provoquant l'encombrement stérique au niveau de la dernière couche comparativement au centre de la molécule qui est plus « aéré » avec la glycogénine.

Cette organisation théorique, basée sur une approche mathématique, conduit ainsi au fur et à mesure que l'on s'éloigne du centre de la particule (figure 3B), à un encombrement stérique empêchant l'activité des enzymes de l'anabolisme du glycogène (glycogène-synthases, enzyme de branchement…) et empêchant ainsi une croissance infinie. Cette dernière couche saturée ne présente donc que des chaînes linéaires, c'est-à-dire directement hydrolysables par les enzymes de type hydrolytique et de type phosphorylasique. De plus, la molécule de glycogène bien qu'ayant une taille limitée, du fait de l'encombrement de la dernière couche, à 42 nm théoriquement, contient 55 000 résidus de glucose, dont 34,6% dans la dernière couche ne soutenant aucun point de branchement, permettant ainsi une libération rapide de l'énergie, lorsque le besoin s'en fait ressentir par les organismes.

Notons aussi qu'il existe environ 20 à 25 tétramères de phosphorylase (Madsen et Corri 1958) sur la surface du glycogène qui dispose de 2100 extrémités non réductrices accessibles sur cette dernière couche (Goldsmith *et al.,* 1982). Ces quelques chiffres permettent aisément d'imaginer l'aspect optimisé de cette structure pour la dégradation rapide par les phosphorylases en glucose-1-phosphate (G1P), qui sera ensuite transformé en énergie. Mais le point de vue dynamique n'est pas le seul intérêt de cette structure. En effet, la distribution des points de branchement et la concentration des chaînes linéaires au niveau de la couche externe permettent un stockage important du glucose (55000 résidus glucose) dans un espace restreint de 42 nm. Cette structure autorise, non seulement de garder le polysaccharide sous forme soluble, donc facilement digestible (contrairement à de l'amidon),

mais tout en diminuant fortement la pression osmotique, qui pourrait être liée à une aussi grande quantité de glucoses sous formes libres. Ainsi, dans leur étude en 1993 Mélendez *et al.,* montraient, par une approche de modélisation mathématique, que pour obtenir une structure optimale avec ces trois restrictions (une grande quantité de glucose dans un minimum de volume, le tout directement accessible sans nécessiter l'intervention d'enzymes de débranchement et un nombre important d'extrémités non réductrices, afin de permettre une libération d'énergie plus rapide obtenue par l'attaque des phosphorylases), celle du glycogène était la plus appropriée. Toutefois, la modélisation mathématique doit être relativisée pour plusieurs points : le glycogène prédit ainsi ne prend pas en compte l'aspect dynamique du vivant, puisque lors de la synthèse il existe un ensemble d'anabolisme et de catabolisme se déroulant en même temps. Par ailleurs *in vitro* il est possible de former des polysaccharides avec une taille bien supérieure à 42 nm (Grimaud *et al.,* 2013 et Roussel *et al.*, 2013). De plus, il a été montré depuis bien longtemps qu'il n'existe qu'une partie minoritaire (29% chez l'homme) du glycogène qui se retrouve sous la forme « optimale » d'après la modélisation (Shearer et Graham 2002), laissant ainsi de nombreuses questions ouvertes sur le rapport structure/fonction.

3.3) L'amidon

L'amidon est utilisé depuis plusieurs millénaires pour divers aspects, et sa structure a fait l'objet de nombreuses controverses, jusqu'au milieu du XXème siècle (pour revue voir Seatharaman et Bertoft 2012a, b, c, d, 2013a, b). Celle-ci ont eu cours jusqu'à ce que le développement de techniques suffisamment puissantes, tant physiques, chimiques que biologiques y mettent partiellement un terme. Ces méthodes ont permis de comprendre que l'amidon était un polymère de glucose lié en position alpha 1-4 et alpha 1-6, constitué de deux fractions majoritaire : l'amylose et l'amylopectine, et d'autres fractions dont ne connait pas leur fonctions exactes (protéine, lipides). Ainsi, dans la prochaine partie, nous décrirons la structure de l'amidon, en essayant de mieux comprendre le rapport entre structure et fonction.

3.3.1) La morphologie des grains d'amidon

L'amidon selon son origine (espèce ou fond génétique de l'organisme) peut avoir des formes différentes, mais les grains présentent toujours des morphologies similaires (http://amidotheque.cermav.cnrs.fr/amidotheque/). Une propriété universelle du grain d'amidon est sa capacité à diffracter la lumière polarisée conduisant à la visualisation des stries de l'amidon (figure 4 B, C et D), et à une biréfringence menant à l'apparition d'une croix de malte (figure 4). Cette croix de malte permet alors de visualiser l'extrémité proximale pointue du grain, permettant de localiser le hile, et la direction de croissance des chaînes de glucose. Même si les grains sont généralement de forme ronde (figure 4 A) comme dans le tubercule de pomme de terre (Zeeman *et al.,* 2002), il existe aussi des amidons en forme d'ogive (figure 4B), comme chez *Phajus grandifolius* (Chanzy *et al.,* 2006), ou bien en forme de disque comme l'amidon retrouvé chez les endospermes de céréales. Ceci dit, ces amidons peuvent aussi avoir des formes plus variables comme en témoigne les figures 4C et 4D où les grains présentent des protubérances, ou bien certains ne présentent pas qu'un seul hile, mais deux (figure 4E) (Chanzy *et al.,* 2006).

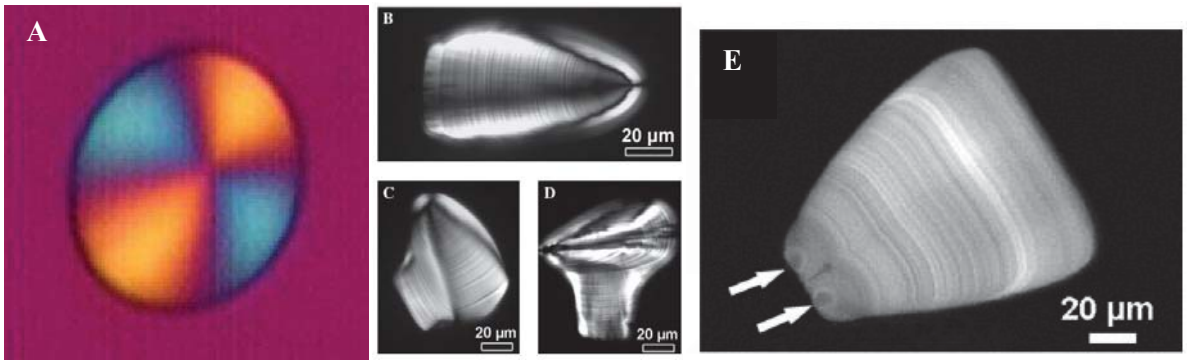


Figure 4 : Grains d'amidon observés au microscope en lumière polarisée : de pomme de terre (*S. tuberculosum*) (A) (Zeeman *et al.,* 2002), et de *P. grandifolius* présentant des formes d'ogives (B), ou des protubérances (C et D). Grain d'amidon observé au microscope confocale présentant deux hiles, indiqué par les flèches blanches (E) (Chanzy *et al.,* 2006).

Ainsi les différences morphologiques observées sont sans doute le reflet de différences fines du métabolisme de l'amidon entre les espèces. Cependant, le mécanisme et la raison biologique de ces modifications ne sont pas encore connus. Toutefois, les études de plus en plus larges sur de nombreux organismes permettent malgré tout d'y voir de plus en plus clair, comme pour les grains retrouvés chez *Guillardia theta* qui auront une forme de « gants de base-ball », dû à la localisation du grain d'amidon autour du pyrénoïde, permettant de

recevoir directement le flux de carbone créé par la photosynthèse (figure 5 A et B) (Deschamps *et al.*, 2006).

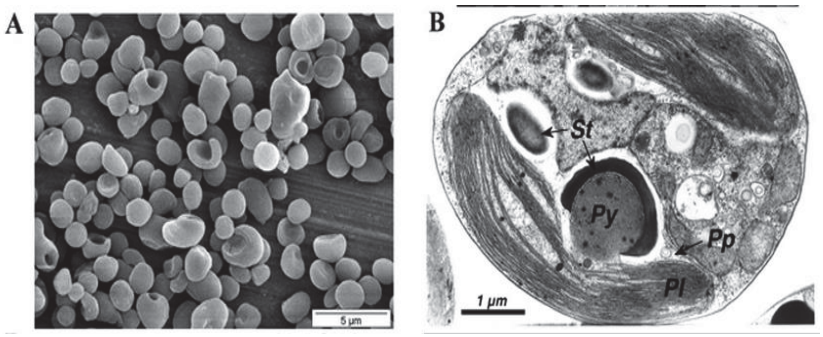


Figure 5 : Visualisation au microscope électronique à balayage des grains d'amidon de *Guillardia theta* en forme de gant de base-ball (A). Cette morphologie particulière s'explique par une croissance du grain d'amidon autour du pyrénoïde de la cellule, comme le montre la figure B après coloration PATAg et visualisation par microscope électronique à transmission(B). Py : pyrénoïde, St : Amidon, Pp : périplaste, Pl : plaste (Deschamps et al 2006)

3.3.2) L'amylopectine

L'amylopectine est le polysaccharide majoritairement retrouvé dans le grain d'amidon avec au minimum 70% de la quantité de glucose (dépendant des espèces et du fond génétique), représentant entre 10^5 et 10^6 résidus glucose avec un poids compris entre 10^7 et 10^9 daltons. C'est l'amylopectine qui est responsable de la cristallinité du grain d'amidon. Ce polysaccharide est, comparativement au glycogène, modérément ramifié, puisque son taux de branchements est compris entre 4 et 6% (pour revue voir Buléon *et al.*, 1998). L'amylopectine est ainsi constituée de milliers de chaînes liées en alpha 1-4 et connectées entre elles par des liaisons alpha 1-6. Notons que ces liaisons en alpha 1-6 forment des grappes grâce à leurs concentrations dans certaines régions, s'opposant ainsi à la distribution régulière des points de branchements dans le glycogène, et provoquant des différences structurales entrainant des propriétés différentes. Dans cette organisation, nous distinguons alors plusieurs types de chaînes, appelées type A, B ou C. Les chaînes sont ainsi nommées en commençant par les chaînes externes et en allant progressivement jusqu'au centre de la molécule. Les chaînes A ne supporteront aucune autre chaîne, et seront soutenues par les chaînes B. Ces chaînes B seront alors supportées par d'autres chaînes B, jusqu'à ne reposer plus que sur une unique chaîne C disposant d'une extrémité réductrice libre. Les chaînes B, en fonction de leurs positions se désigneront : chaînes B1 pour celles soutenant les chaînes A puis au fur et à mesure que l'on

s'approchera du centre de la particule d'amidon les chaînes se nommeront B2, B3… Les chaînes B et les chaînes A comprises entre 6 et 35 résidus glucose liés en position alpha 1-4, forment des grappes grâce à l'organisation des points de branchements, alors que les chaînes B (ou C) au-dessus de 35 sont des chaînes qui sont impliquées dans la connexion des grappes qui pourront varier en taille et en forme. Ainsi le ratio des chaînes courtes (6-35 glucoses) sur les chaînes longues (>35 glucoses) (Hanashiro *et al.,* 2002) s'applique couramment pour déterminer la taille des grappes, même si ce modèle ne permet pas de déterminer exactement la taille de toutes les grappes de façon précise (Bertoft *et al.,* 2011). Cette organisation en grappe conduit à la formation de zones cristallines et amorphes. Les parties constituées de chaînes de glucose linéaire des grappes forment les doubles hélices qui sont cristallines, et les zones amorphes se situeront an niveau des parties où se condenseront les points de branchements (figure 6). Cette succession de zones amorphe et cristalline d'une taille de 9nm est caractéristique de l'amylopectine, quel que soit l'organisme (figure 6).

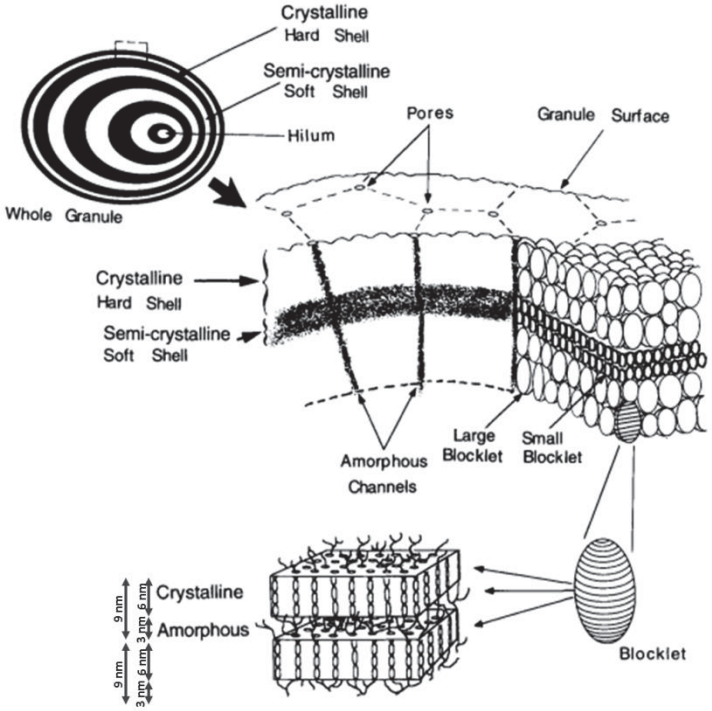


Figure 6 : Représentation d'un grain d'amidon avec l'alternance des couches cristallines et semi cristallines (en haut à gauche), puis un agrandissement est réalisé pour mieux apprécier les couches cristallines et amorphes, mais aussi situer, à l'intérieur de ces couches, les structures en blocklet. Enfin l'agrandissement des blocklets montre la succession des couches amorphes (région où se situe les points de branchements) et couches cristallines (régions constituées de double hélices de glucose) qui auront une taille de 9 nm (adapté de Gallant *et al.* 1997).

Au niveau des grappes, cette succession amorphe cristalline n'est pas responsable de la succession des couches du grain d'amidon, comme observé après hydrolyse partielle (figure

6). En effet, cette structure observable au microscope est une alternance de couches « molles », facilement hydrolysées, et de couches plus dures difficilement hydrolysables, et, de ce fait, encore présentes même après hydrolyse (figure 6). Il existe donc un autre niveau de structure, au niveau des couches observables en microscopie électronique comme dans la figure 6, qui est appelé blocklet (Gallant *et al.,* 1997). Ces blocklets sont de petites structures constituées de plusieurs grappes possédant une alternance de structure amorphe et de structures cristalline de 9 nm. Cette structure de 9 nm comprend la zone de ramification de la grappe (3 nm), qui permettra alors la formation de chaîne de glucose linéaire suffisamment longue pour aboutir à une formation en double hélice de glucose (figure 6), qui chasseront l'eau et provoqueront l'effondrement en un solide semi-cristallin. La formation de l'amidon en grappe sera réalisée de cette manière, couche après couche, dans toutes les directions (à partir du hile), jusqu'à obtenir un grain d'amidon qui théoriquement pourrait avoir une taille illimitée.

Si tous les grains d'amidon ont une organisation commune pour les grappes et les blocklets, il existe des différences entre eux, selon les espèces ou le fond génétique. En effet, selon l'organisation des doubles hélices de glucane, il existe l'amylopectine de type A, B ou C (Buléon *et al.,* 1997). Ainsi, dans l'endosperme des céréales nous retrouverons de l'amidon de type A (figure 7A), où les hélices sont encastrées les unes dans les autres et entourées de 8 molécules d'eau par unité, et forment une unité monoclinique (Imberty *et al.,* 1988). Tandis que, dans les tubercules ou bien dans l'amidon transitoire des feuilles, nous observons du type B (figure 7B), où six doubles hélices de glucose, forment un hexagone, entourant 36 molécules d'eau (Imberty et Perez 1988). Enfin, l'amidon de type C ne représente pas une structure distincte, mais sera un mélange de type A et B qui se retrouve majoritairement dans les légumineuses. Cependant la structure et la cristallinité de l'amidon ne dépendent pas uniquement de l'amylopectine, car il a été vu sur les grains d'amidon de *Chlamydomonas reinhardtii* qu'un enrichissement en amylose peut influencer le type cristallin et créer des cristallites de type B (Wattebled *et al.,* 2002).

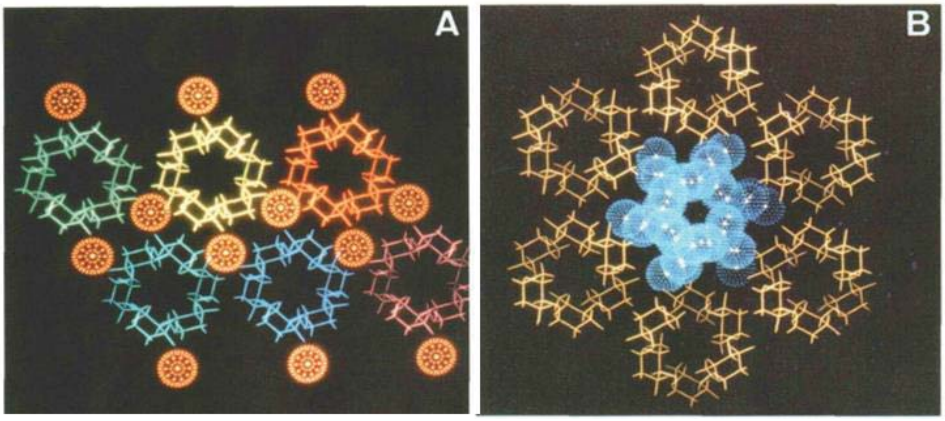


Figure 7 : Organisation des doubles hélices de glucose dans l'amylopectine de type A (A) et de type B (B). Dans l'amylopectine de type A on observe clairement les hélices enchâssées entre elles entourées de 8 molécules d'eau (ronds orange) (A), tandis que l'amylopectine de type B est organisé autour des molécules d'eau en son centre (ronds bleu) (B)(Buléon *et al.,* 1997)

3.3.3) L'amylose

La deuxième fraction majoritaire du grain d'amidon, composée de glucose, est l'amylose. C'est une fraction de 600 à 1000 résidus glucose, essentiellement constituée de liaison alpha 1-4 avec un taux de ramification très bas d'environ 1% (Takeda *et al.,* 1993). Cette fraction est non-liée à l'amylopectine, en témoigne la séparation des deux fractions sans hydrolyse des liaisons glycosidiques, mais simplement insérée entre les grappes d'amylopectine. L'amylose est considérée comme non nécessaire à la biosynthèse d'un grain d'amidon cristallin (Ball *et al.,* 1998). De plus, elle ne pourrait pas exister en l'absence d'amidon cristallin, dû à une nécessité d'un grain semi-cristallin pour l'activité de la GBSS (Granule bound starch synthase) -seule famille d'enzyme théoriquement responsable de la synthèse d'amylose (Dauvillée *et al.,* 1999)-. Toutefois plus récemment a été mise en évidence, dans des mutants de l'orge, déficients pour toutes les enzymes de branchements, une structure cristalline uniquement constituée d'amylose, synthétisée peut être par la GBSS, mais aussi peut-être par les amidon synthases solubles -toutes capables de synthétiser de longues chaînes de glucose liées en alpha 1-4-(Carciofi *et al.,* 2012), ce qui pourrait remettre en cause une partie des idées reçues. L'amylose peut ensuite être associée à d'autres éléments et former de simples hélices avec des molécules hydrophobes, telles que les lipides (Buléon *et al.,* 1998). Cette fraction, bien qu'assez bien étudiée, pourrait ainsi cacher des propriétés intéressantes comme le sous-entend l'étude de Wattebled *et al.,* 2002, qui montre l'influence de l'amylose sur le type cristallin des grains d'amidon.

3.3.4) La semi amylopectine

A travers les précédentes parties, nous avons montré que l'amidon possède de nombreuses facettes influençant sa structure et ses propriétés physico-chimiques. Ainsi parmi toutes les formes d'amidon, il existe aussi des formes limites, que nous pourrions définir comme intermédiaire entre le glycogène et l'amidon ; c'est le cas de la semi amylopectine. C'est une molécule cristalline, parfois appelée « amidon cyanobactérien » (appellation historique puisqu'ayant été décrite en premier chez les cyanobactéries) (Deschamps *et al.,* 2008c) ; toutefois elle se distingue de l'amidon « commun » par une quantité de chaînes courtes moins importantes (DP≤8) par rapport aux chaînes longues (DP≥37) (Nakamura *et al.,* 2005). Ces différences structurales vont avoir plusieurs conséquences, notamment sur les grappes ou sur la température de gélatinisation. (Suzuki *et al.,* 2013). Il est à noter qu'il existe aussi des structures intermédiaires chez d'autres organismes que les cyanobactéries, comme chez les algues rouges, avec le cas de *Cyanidioschyzon merolae* (Hirabaru *et al.,* 2010), ou chez d'autres espèces comme il a été observé par Shimonaga *et al.*,2008.

Nous remarquons que l'amidon de la souche *Cyanobacterium* sp. CLg1 est de la semi-amylopectine (Suzuki *et al.,* 2013). La raison biologique pour laquelle les cyanobactéries accumulent de la semi amylopectine, plutôt que de l'amylopectine, pourrait être liée à la dégradation, mais aussi à des voies métaboliques légèrement différentes, d'un point de vue fonctionnel, mais aussi d'un point de vue phylogénétique, qui seront vus plus tard. Il est aussi à noter que l'amidon de CLg1 semblerait être modifié selon les conditions de cultures. En effet les grains retrouvés après croissance en milieu pourvu d'azote (ASNIII) présentent une cristallinité de type A (Suzuki *et al.*,2013), tandis que ceux retrouvés dans les cellules cultivées dans un milieu dépourvu d'azote (ASO) présentent une cristallinité de type C (Putaux J-L communication personnelle).

7. Chapitre : Les cyanobactéries

4.1) Les cyanobactéries et leurs classifications

Les cyanobactéries sont des organismes procaryotes capables de réaliser la photosynthèse oxygénique. Elles ont, par ailleurs, joué un rôle important dans l'oxygénation de l'atmosphère il y a 2.45-2.32 million d'années (Schirrmeister *et al.,* 2013, Rasmussen *et al.,* 2008). Cette capacité fut aussi transmise aux plantes lors de l'endosymbiose primaire du plaste, décrite précédemment. Les milieux de vie sont principalement l'eau douce et l'eau de mer, mais elles peuvent aussi être retrouvées dans des environnements chauds (déserts), ou hyper-salins. Même si tout l'embranchement des cyanobactéries forme, de manière solide, un clade monophylétique (Groupe 11, Holt *et al.*, Chapitre : Oxygen phototrophic bacteria, Bergey's Manual of determinative bacteriology, 9th edition, 1994), il existe néanmoins de grandes variations au niveau morphologique, mais aussi au niveau de leur division cellulaire, sans oublier les nombreux autres mécanismes différents entre les espèces, tels que la fixation d'azote. De ces observations morphologiques, et des caractères aisément observables, chaque cyanobactérie a été classée dans l'une des 5 grandes classes (Rippka *et al.,* 1979). Mais la disponibilité des génomes et les nouveaux outils phylogénétiques ont permis de réaliser des classifications différentes sur bases, par exemple, des ARN 16S (Honda *et al.,* 1999), permettant ainsi de séparer les cyanobactéries en 7 groupes distincts. D'autres classifications ne sont pas à exclure comme celle de Gupta qui sépare les cyanobactéries en 3 clades (Gupta et Mathews 2010). Encore plus récemment, les études basées sur des reconstructions d'arbre d'espèces à partir de plusieurs arbres de gènes (superarbre) ont appuyé certaines classifications ou en ont déstabilisé d'autres (Szöllosi *et al.,* 2012). Ces classifications qui prêtent toutes à discussions ne sont pas étonnantes, puisque les cyanobactéries forment un très vieil embranchement de procaryotes constituant un groupe qui existerait depuis au moins 2 à 2.5 milliards d'années, d'après les données fossiles et les biomarqueurs utilisés (Summons *et al.,* 1999) et même peut être depuis 2.7 ou 3.5 milliards d'années (Falcon *et al.*, 2010). Ainsi les transferts latéraux de gènes, courant chez les procaryotes, ont pu, au fur et à mesure, augmenter et interférer avec des classifications précises (Dagan *et al.,* 2013). De plus, des événements multiples menant à des évolutions morphologiques convergentes ont aussi eu lieu, comme la capacité à se trouver sous forme unicellulaire ou multicellulaire (Schirrmeister *et al.,* 2011) (figure 8).

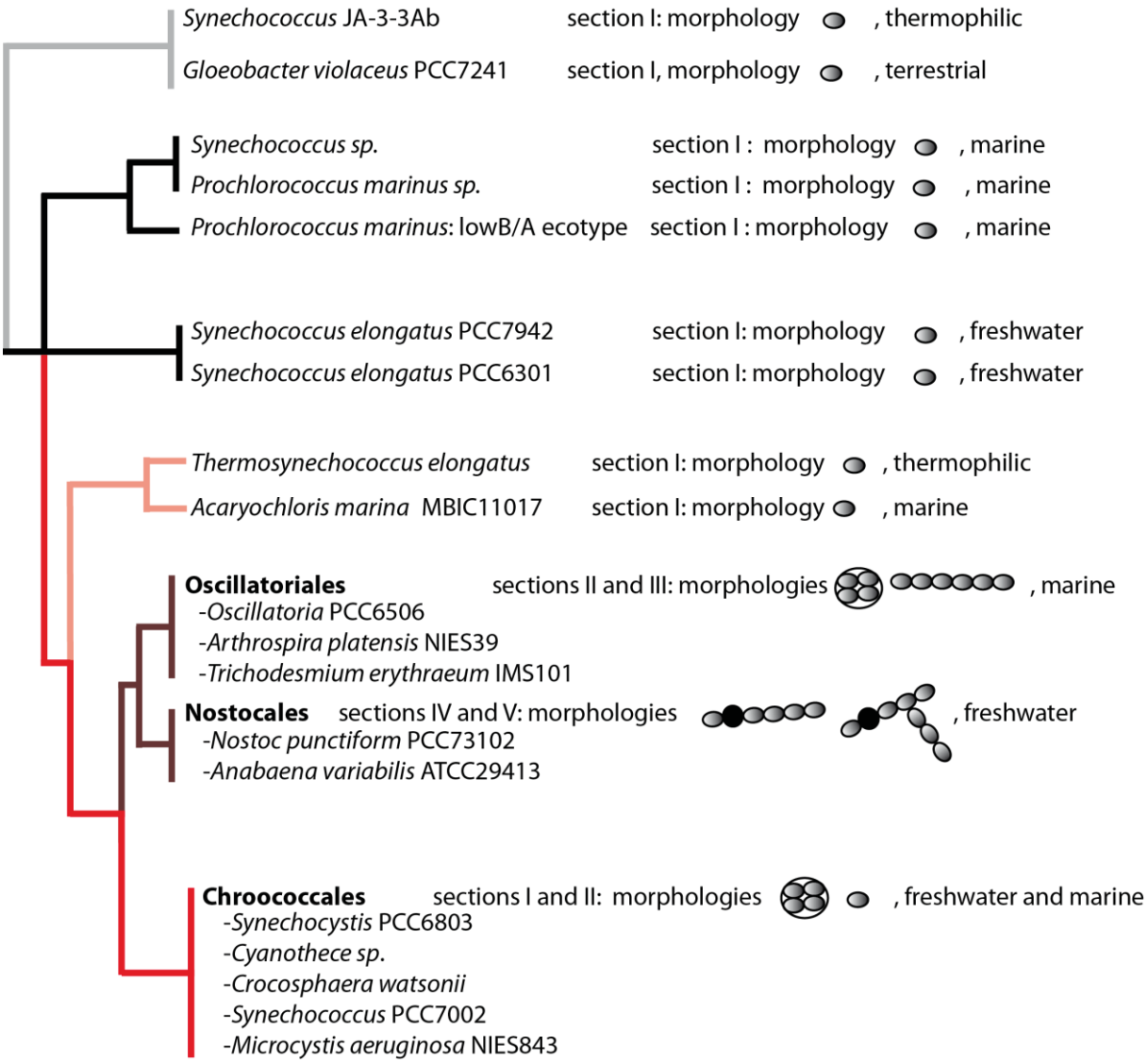


Figure 8 : Représentation schématique de la phylogénie des cyanobactéries, montrant les différents ordres, selon leurs ARN 16S. La morphologie des cyanobactéries retrouvées dans ces ordres est aussi présentée, et démontre la grande variété morphologique chez les cyanobactéries, mais aussi au sein même des différents ordres (tiré de Colleoni et Suzuki 2013).

L'un des rôles majeurs accordés à ce groupe d'organismes est l'oxygénation de l'environnement terrestre il y a 2.5 milliards d'années. Cependant, les cyanobactéries ont aussi permis l'acquisition de la photosynthèse par les eucaryotes lors de l'endosymbiose primaire du plaste, il y a 0.7 à 1.5 milliard d'années (Van den Eynde *et al.*, 1988). Si l'endosymbiose primaire du plaste est l'événement évolutif certainement le plus décrit et le plus étudié dans la littérature, les autres symbioses impliquant les cyanobactéries (lichen…) sont aussi particulièrement intéressantes. Elles sont d'ailleurs de plus en plus étudiées dans un but écologique mais aussi dans le but de déterminer les mécanismes permettant de passer de deux organismes distincts à une endosymbiose obligatoire. L'endosymbiose primaire permis

aux eucaryotes, non seulement de réaliser la photosynthèse pour fixer le carbone atmosphérique et produire de l'énergie, grâce au sucre produit par l'endosymbionte (Weber *et al.,* 2006), mais aussi la capacité à libérer de l'O_2 qui aurait permis un meilleur fonctionnement de la chaîne respiratoire de la mitochondrie (Martin et Müller 1998), et également la capacité à synthétiser certains acides aminés (Reyes-Prieto et Moustafa, 2012).

A travers cet événement, les cyanobactéries ont, ainsi, participé de manière majeure, aux génomes des eucaryotes ayant internalisé la cyanobactérie par transfert endosymbiotique de gènes. En effet, environ 18 % du génome pour *Arabidopsis thaliana*, 7 % pour les algues rouges, et 6 % pour les glaucophytes, sont des gènes présentant une origine cyanobactérienne (Price *et al.,* 2012) dont les protéines correspondantes interviennent dans de nombreuses voies métaboliques : cycle de Calvin, synthèse des acides aminés (Reyes Prieto et Moustafa 2012). Plus récemment, des études ont montré que des enzymes d'origines cyanobactériennes intervenaient dans le métabolisme des polysaccharides de réserves et que ces enzymes étaient, d'un point de vue phylogénétique, reliées à un groupe de cyanobactéries unicellulaires marines fixatrices d'azote (Deschamps *et al.,* 2008c, Nakamura *et al.,* 2005, Scheengurt *et al.,* 1994). Ces études indiquaient que l'ancêtre potentiel du plaste pouvait appartenir à ce groupe, toutefois, si des études vont en ce sens (Gupta et Matthews 2010), d'autres font penser le contraire (Deusch *et al.,* 2008, Dagan *et al.,* 2013), puisque ces dernières accordent l'origine du plaste aux cyanobactéries filamenteuses hétérocystes, des cyanobactéries du sous-groupe IV ou V selon la classification de Rippka.

Dans toutes les études, des propriétés communes de l'ancêtre du plaste existent, comme par exemple la diazotrophie. Cet élément plaide ainsi, en faveur d'un plaste ancêtre des Chroococales (cyanobactérie diazotrophe unicellulaire) qui, bien que longtemps sous-estimé par rapport aux cyanobactéries hétérocystes, est considéré pour jouer un rôle important dans le cycle de l'azote en rendant disponible l'azote pour la croissance du phytoplancton, même si ce sont les cyanobactéries filamenteuses non hétérocystes du genre *Trichodesmium* qui seraient les principales cyanobactéries fixatrices d'azotes dans les écosystèmes (pour revue Zehr 2011).

De manière intéressante, plusieurs façons de coupler la photosynthèse oxydative productrice d'oxygène avec la réduction d'azote par la nitrogénase (qui craint l'oxygène) a été démontré. Il peut exister :

- Soit des cellules spécialisées qui permettent la séparation dans l'espace des deux fonctions incompatibles, comme les cyanobactéries hétérocystes. Certaines cellules réaliseront la photosynthèse, et fourniront énergie et substrat carboné aux cellules

réalisant la fixation d'azote, ne réalisant pas la photosynthèse et appelées cellules hétérocystes. Ces cellules possèderont une large paroi cellulaire protégeant la nitrogénase de l'oxygène.

- Soit, nous pourrons avoir une séparation dans le temps, c'est le cas des cyanobactéries unicellulaires, où la photosynthèse se déroulera pendant la phase lumineuse tandis que la fixation d'azote se déroulera pendant la phase obscure. De plus, pendant la phase obscure la respiration conduira à une chute de l'oxygène accumulée pendant la journée, préservant ainsi la nitrogénase de l'oxygène produite pendant la journée. Ceci sous-entend la nécessité d'accumuler pendant la phase lumineuse une grande quantité d'énergie afin de permettre la fixation d'azote (puisque cette réaction consomme 16 molécules d'ATP pour une molécule d'azote fixée), mais aussi la division cellulaire de ces organismes. C'est pourquoi le métabolisme du glycogène aurait évolué en métabolisme de l'amidon qui est inerte osmotiquement et capable d'accumuler de grande quantité de glucose. C'est en tout cas l'explication proposée dans plusieurs études (Deschamps *et al.,* 2008c, Falcon *et al.,* 2010) pour justifier la présence d'amidon chez les cyanobactéries mis en évidence par Nakamura *et al.,* 2005.

4.2) Métabolisme des polysaccharides de réserves chez les cyanobactéries

Les connaissances accumulées sur le mécanisme des polysaccharides chez les bactéries sont surtout issues de l'étude de l'entérobactérie *E. coli* (revue Ball et Morell 2003), modèle couramment utilisé pour les procaryotes. Mais si pour *E. coli* nous avons un nombre d'isoformes de protéines restreints, cela n'est pas valable pour tous les organismes procaryotes. En effet, des cyanobactéries, comme *Cyanobacterium* MBIC 10216, *Cyanothece* ATCC 51142, ou *Cyanobacterium* sp.CLg1, présentent plusieurs isoformes aussi bien pour les activités amidon-synthases, que pour les activités de débranchement (Colleoni et Suzuki, 2012). Ces dernières, ne sont présentes qu'en un exemplaire chez *E.coli*, qui est réservée au catabolisme, supposant d'autres fonctions chez ces cyanobactéries. En effet, certaines cyanobactéries présentent 4 enzymes potentiellement impliquées dans le débranchement des polysaccharides comme *Cyanobacterium* sp. CLg1. Cette redondance pourrait ainsi révéler une optimisation de la synthèse de polysaccharides. Par ailleurs, il faut souligner que la

synthèse de polysaccharides, chez les cyanobactéries est fortement reliée aux autres activités de la cellule, ce que nous verrons par la suite.

4.3) Les polysaccharides de réserves de type alpha-glucane chez les cyanobactéries

Les premières observations de granules polysaccharidiques ont été montrées par le laboratoire de Sherman (Schneegurt *et al.,* 1994, 1997), chez la cyanobactérie unicellulaire diazotrophe *Cyanothece* ATCC5142, qui est capable d'accumuler, pendant la nuit, des granules de polysaccharide. Ces analyses menées par Sherman, aussi bien par microscopie du granule de carbohydrate, que par colorations spécifiques des polysaccharides, révélaient, seulement, une structure anormale du glycogène. Quant à l'amidon, il fallut attendre que l'équipe du Professeur Y. Nakamura mette en évidence sa présence (Nakamura *et al.,* 2005). Ce polysaccharide ainsi révélé chez différentes cyanobactéries se fit appeler semi-amylopectine du fait de sa faible quantité en longues chaînes, empêchant la distribution polymodale de l'amidon comme retrouvée chez les plantes supérieures, et d'autres espèces appartenant aux *Archaeplastida*. De plus, aucun grain ne présentait de présence d'amylose.

Parallèlement à cette étude, une cyanobactérie diazotrophe unicellulaire marine, la cyanobactérie *Cyanobacterium* sp. CLg1, a été isolée et a révélé la présence de granules de polysaccharide contenant de l'amylose (Deschamps *et al.,* 2008c). Cette cyanobactérie isolée dans l'atlantique nord (Falcon *et al.,* 2004) est phylogénétiquement proche de la cyanobactérie *Crocospherae watsonii* selon la séquence de l'ARN 16S. L'un des aspects les plus intéressants dans la description de cet amidon est sa proximité phylogénétique avec l'ancêtre du plaste des *Archaeplastida* d'après les études de Gupta et Matthews 2010 et de Deschamps *et al.,* 2008c).

Mais l'aspect phylogénétique n'est pas le seul à prendre en compte pour comprendre le métabolisme des synthèses des polysaccharides. En effet, une approche fonctionnelle est nécessaire, mais sans perdre de vue que ce métabolisme est intimement lié à certains mécanismes, et à la physiologie de la cellule.

4.4) Un métabolisme dirigé par l'horloge circadienne

L'horloge circadienne chez les cyanobactéries comme chez les plantes dirige de nombreux processus biologiques comme la division cellulaire, mais aussi le métabolisme des polysaccharides de réserves. Ce rythme chez les cyanobactéries est particulièrement important car dépendant uniquement de trois protéines KAI A, KAI B et KAI C. Ces trois protéines sont ainsi capables de se phosphoryler et de se déphosphoryler, en présence d'ATP, selon un rythme particulier et cela même *in vitro* (pour revue voir Johnson *et al.,* 2008). L'ATP étant particulièrement important puisque des études semblent montrer que ces protéines sont régulées par le ratio ATP/ADP (Rust *et al.,* 2011).

Ainsi les protéines KAI A et KAI B vont phosphoryler et déphosphoryler la protéine KAI C sur un rythme d'environ 24h, ensuite cette protéine sera en mesure de jouer sur l'expression de certains gènes de manière directe ou indirecte. Cette régulation se fera au niveau transcriptionnel sur une grande partie du génome des cyanobactéries (Aoki *et al.,* 2002 et Liu *et al.,* 1995) et permettra ainsi de séparer différentes voies métaboliques incompatibles, telles que la photosynthèse oxygénique et la fixation de l'azote par la nitrogénase (Welsh *et al.,* 2008), mais aussi régulera le métabolisme des polysaccharides de réserves. Une étude a ainsi montré un schéma de régulation des gènes *glgP* et *glgA* (Mohr *et al.,* 2010) deux protéines impliquées dans le métabolisme des polysaccharides de réserve chez *C. watsonii*. En effet, durant la nuit, une transcription élevée du gène *glgP*, impliqué dans le catabolisme a été observé, tandis que le gène *glgA* était lui transcrit pendant la journée. Des expériences, menées par Jenifer Nirmal Raj chez la cyanobactérie *Cyanobacterium* sp. CLg1, ont aussi montré une activité opposée entre l'activité synthase et l'activité phosphorylase (Thèse de Nirmal Raj J. 2010). Cette dernière pourrait être sujette à une régulation plus fine telle qu'une modification post-traductionnelle, comme celle induite par la protéine Histine-phosphocarrier protein (HPr) chez *E.coli* qui permet de réaliser une oligomérisation entre HPr déphosphorylé et GlgP permettant son activation (Seok *et al.,* 1997).

4.5) Un métabolisme qui dépend directement de la photosynthèse

De manière instinctive, si nous devions déterminer de quel mécanisme peut dépendre le métabolisme des polysaccharides de réserves, nous penserions tout de suite à le relier à celui qui produit de l'énergie. Dans les plantes et les cyanobactéries, qui réalisent la photosynthèse oxygénique, il est donc naturel de retrouver une relation intime entre

photosynthèse et synthèse des polysaccharides de réserves. De plus, ce lien est d'autant plus fort que l'ADP glucose pyrophosphorylase, enzyme permettant la transformation du glucose-1-phosphate en ADP glucose, premier intermédiaire de l'anabolisme des polysaccharides chez les cyanobactéries est directement activé par un effecteur provenant de la photosynthèse, le 3-PGA (Plaxton et Preiss, 1987). Il faut souligner que le 3-PGA et sa concentration dépendent de la photosynthèse, puisque c'est lors de la première étape du cycle de Calvin que le 3PGA est synthétisé, à partir du dioxyde de carbone (CO_2), par la RuBisCo (Ribulose-1,5-biphosphate carboxylase oxygénase). Il est ainsi représentatif de l'efficacité photosynthétique, et donc de la quantité de glucose pouvant être métabolisé en polysaccharide de réserve.

4.6) Un métabolisme dépendant de la disponibilité d'azote directement assimilable

Il est connu que le manque d'un composé important pour la cellule peut provoquer une augmentation de la quantité de glycogène dans la cellule, chez les procaryotes hétérotrophes (Preiss *et al.,* 1983), mais il est aussi possible d'obtenir le même phénotype chez les cyanobactéries non fixatrices d'azote comme avec *Synechocystis* PCC 6803 (Yoo *et al.,* 2007). Toutefois ces changements de condition de culture doivent être contrôlés puisqu'ils entraineront plusieurs changements : une dégradation de leurs phycobiliproteines (de Loura *et al.,*1987), mais aussi des modifications structurales du polysaccharide, puisque dans certains cas, nous aurons une modification de la taille moyenne des chaînes. D'autres études sur des cyanobactéries fixatrices d'azote semblent aussi montrer une augmentation des polysaccharides en absence d'azote (Reddy,*et al.,*1993).

8. Chapitre : Mécanisme de synthèse des alpha-glucanes de réserves

Si les structures et les propriétés de l'amidon et du glycogène sont très différentes alors que leur composition est semblable, nous savons que leurs métabolismes, nous savons que leurs métabolismes, quel que soit l'organisme, ne varient que très modestement. Dans ce chapitre, nous tenterons donc de comprendre les variations significatives conduisant aux différences majeures observées en réalisant un état de l'art général du métabolisme de l'amidon et du glycogène.

Le mécanisme de synthèse de tous les alpha-polysaccharides de réserve recensés, est initié à partir du glucose-1-phosphate (G-1-P) qui est transformé en nucléotide sucre, c'est-à-dire soit transformé en UDP-glucose comme chez les opisthokontes, soit transformé en ADP-glucose (ADP-glc) comme chez les bactéries en général ou chez les Chloroplastida.

Glucose-1-phosphate + NTP ➔ NDP-glucose + PPi

Il est à noter que si la synthèse d'ADP-glucose est finement régulée chez les bactéries ou les plantes, par différents effecteurs ; soit en induisant, comme le 3PGA chez les *Chloroplastida* ou les cyanobactéries qui reflète directement l'activité photosynthétique, soit en inhibant l'activité ADP-glucose pyrophosphorylase comme l'orthophosphate (Plaxton and Preiss, 1987). Ce type de régulation chez les organismes utilisant l'UDP-glucose est toutefois impossible puisque ce dernier est au carrefour de nombreuses voies métaboliques. Par conséquent, la biosynthèse du glycogène s'effectue essentiellement sous le contrôle des activités glycogène-synthases qui sont sous formes actives ou inactives en fonction de leur état de phosphorylation. A la suite de cette transformation en nucléotide-sucre, nous retrouverons l'étape de l'initiation des chaînes.

UDP-glucose + glycogénine-Tyr ➔ glycogénine-Tyr-O-$(glc)_n$+ UDP

Chez la levure, cette initiation est connue pour être catalysée par la glycogénine, une enzyme appartenant à la famille GT 8 (Glycosyl Transferase selon la classification CAzy), qui grâce aux résidus de tyrosine permet la glucosylation avec le nucléotide sucre. Le glycogène viendra alors se former autour de cette enzyme (Lomako *et al.*, 2004). Ainsi si le mécanisme

est bien connu et établi pour la synthèse du glycogène chez les opisthokontes, l'initiation chez les organismes ne possédant pas de glycogénine comme les Chloroplastida ou les bactéries, n'est que très peu décrit dans la littérature. Toutefois de récentes études chez *Arabidopsis thaliana* (Szydlowski *et al.,* 2009) et chez *Agrobacterium tumefaciens* (Ugalde *et al.,* 2003) montrent que certaines glycogène/amidon synthase à ADP-glucose, appartenant à la famille GT 5 (Cantarel *et al.*, 2009), seraient capable d'initier la synthèse de polysaccharide. Mais le mécanisme exact d'initiation n'a pas encore été défini, d'autant plus que certaines études montrent que le complexe phosphorylase enzyme de branchement chez les chloroplastida pourrait être capable d'initier *in vitro* (Nakamura *et al.,* 2012). Suite à l'initiation des chaînes de glucanes les enzymes de type glycogène/amidon synthase à ADP-glucose ou à UDP-glucose (famille GT 5), allongent ensuite les chaînes nouvellement synthétisées.

$$\text{NDP-glucose} + \text{alpha,1-4,(glucose)}_n \rightarrow \text{alpha,1-4,(glucose)}_{n+1} + \text{NDP}$$

Ce mécanisme d'allongement des chaînes est bien conservé à travers les organismes, mais la spécificité des formes de synthase paraît toutefois différentes, et pourrait bien provoquer des changements significatifs dans la structure du polysaccharide comme sur le grain d'amidon (Zhang *et al.,* 2008, Szydlowski *et al.,* 2011). De plus, pour les opisthokontes par exemple, des effecteurs, comme le glucose-6-phosphate (G-6-P), ou des mécanismes de phosphorylation (Roach *et al.*, 2012) régulent finement ces activités enzymatiques appartenant à la famille GT3 (Cantarel *et al.*, 2009), impliquant des différences majeures dans leur mode d'action.
Les chaînes de glucose sont ensuite transformées en polysaccharide ramifié par l'action des enzymes de branchements.

Glucanes linéaires liés en alpha 1-4 ($\geq$12) ➔ Glucanes linéaires liés en alpha 1-4 avec un glucane linéaire lié en alpha 1-6

Les enzymes de branchements (BE) ont des propriétés assez proches d'un organisme à l'autre, et se définissent par leurs capacités à cliver une chaîne linéaire et transférer le malto-oligosaccharide en position alpha 1-6 sur le glucose d'une autre chaîne ou de la même chaîne. Toutefois pour réaliser cette réaction, il est nécessaire que la chaîne de glucose linéaire soit assez longue, pour qu'il puisse rester 6 résidus de glucose sur la chaîne initiale après

l'hydrolyse, et pour que la chaîne à transférer soit d'une taille minimale de 6-7 résidus de glucose (Guan *et al.,* 1997, Nakamura *et al.,* 2010).

En parallèle l'enzyme possédant l'activité inverse n'est autre que l'enzyme de débranchements. En effet, c'est cette enzyme qui permettra selon le « trimming modèle » présenté par Ball *et al.,* en 1996, la formation des grappes d'amylopectine, par un débranchement spécifique des points de branchement mal positionnés, empêchant la formation de double hélices, au niveau d'un précurseur soluble, permettant ainsi la cristallisation de l'amidon (Figure 9).

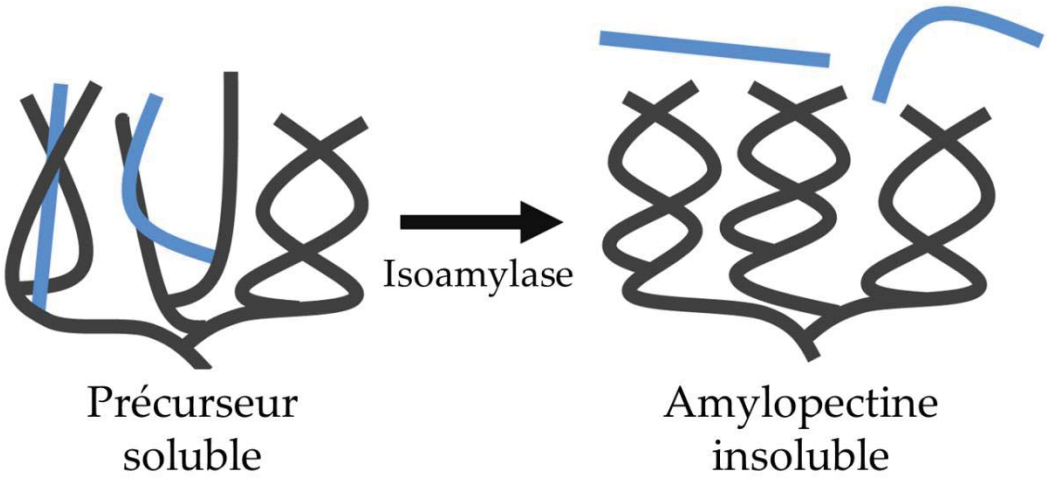


Figure 9 : Schéma du mécanisme d'épissage par l'isoamylase des Chloroplastida présenté par Ball *et al.,* en 1996. L'isoamylase est nécessaire pour le passage d'un polysaccharide soluble (à gauche), où certaines chaînes empêchent la formation de double hélice (en bleu), à un polysaccharide de type amylopectine (à droite), formé par une zone où se concentreront les branchements afin de permettre la formation de double hélices, et la cristallisation.

L'enzyme de débranchement présente chez les *Archaeplastida*, et qui a été décrite comme nécessaire à la cristallisation de l'amidon (Mouille *et al.,* 1996), est une enzyme de débranchement directe du même type que celle retrouvée chez les procaryotes, et différente de celle retrouvée couramment chez les eucaryotes, qui elle est une enzyme de débranchement indirecte (iDBE). Il est important de dire que si leur activité conduit dans les deux cas à l'hydrolyse d'une liaison alpha 1-6 le schéma d'action reste très différent. En effet, si l'enzyme de type directe peut cliver directement la liaison alpha 1-6, l'enzyme de débranchement indirecte devra tout d'abord hydrolyser la liaison alpha 1-4 du glucose possédant la liaison du point de branchements, puis transférer cette chaîne sur l'extrémité non réductrice d'une autre chaîne, pour enfin hydrolyser la liaison alpha 1-6 et libérer le dernier glucose (figure 10).

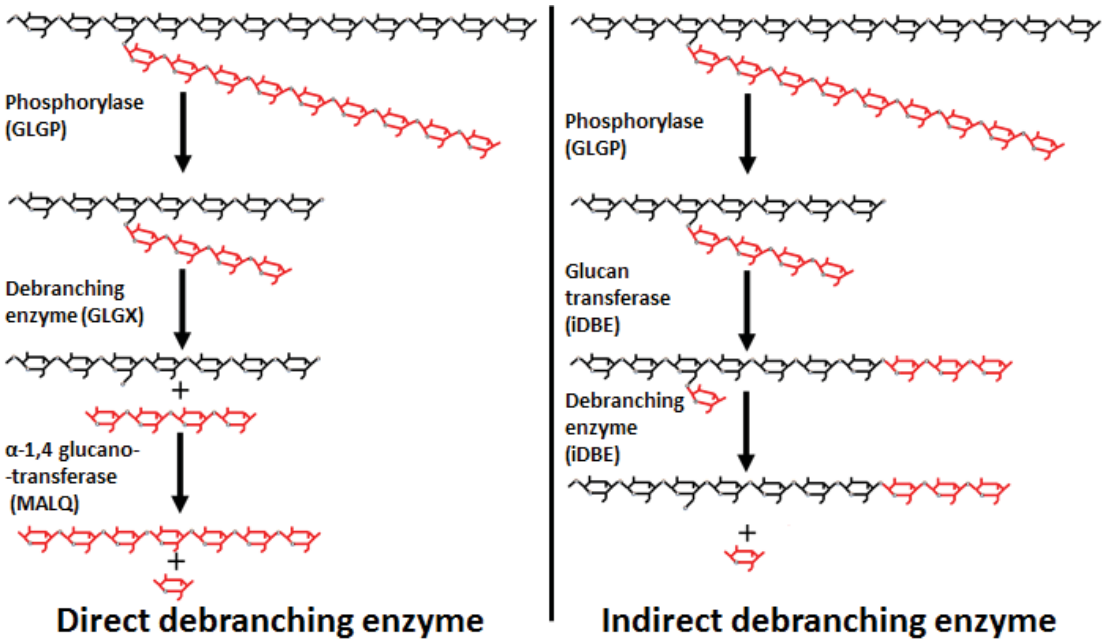


Figure 10: Schéma récapitulatif de la différence d'action entre une enzyme de débranchement directe et indirecte. On observe ici que les deux enzymes doivent théoriquement agir sur des glucanes rognés par des phosphorylases, puis l'enzyme de débranchement directe hydrolysera sans étape préalable la liaison alpha 1-6 et libérera un maltooligosaccharide. Le maltooligosaccharide sera, ensuite, recyclé par une enzyme disproportionnante de type MalQ, tandis que l'enzyme de débranchement indirecte devra d'abord hydrolyser la chaîne ramifiée et transférer l'oligosaccharide avant d'hydrolyser la liaison alpha 1-6.

Ces enzymes capables de débrancher directement ont aussi une spécificité de substrat. En effet, comme il a été décrit chez *E. coli* et *Synechococcocus* PCC7942, les enzymes directes de types bactériennes paraissent actives uniquement sur un substrat ayant déjà subi l'action de dégradation du glycogène par la phosphorylase, c'est-à-dire dont les chaînes externes ont une taille de 4 résidus de glucose, comme observé dans la figure 10 (Dauvillée *et al.*, 2005 ; Suzuki *et al.*, 2007). A l'inverse l'enzyme de débranchement directe retrouvée chez certaines bactéries comme *Pseudomonas* sp., et chez les *Archaeplastida* est capable d'hydrolyser des polysaccharides avec de longues branches de type amylopectine (Dauvillée *et al.*, 2000). Mais si ces études semblent montrer une activité non spécifique de cette enzyme envers le glycogène et l'amylopectine, des études plus récentes basées sur l'isoamylase de *S. tuberosum* semblent montrer une activité forte sur amylopectine et très réduite sur le glycogène (Hussain *et al.*, 2003). Ajoutons que la spécificité de l'isoamylase pour les polysaccharides avec de longues branches externes est, pour l'instant, expliquée par la présence d'une boucle entre les domaines II et III (Katsuya *et al.*, 1998). Si cette activité responsable de la cristallisation de l'amidon chez les Chloroplastida est toutefois encore peu

décrite de manière enzymologique, elle pourrait révéler des propriétés différentes, comparativement aux autres enzymes de débranchements, permettant de comprendre le mécanisme de cristallisation des polysaccharides de réserve.

Si les enzymes de débranchements sont responsables chez les Chloroplastida et surement chez tous les *Archaeplastida* de la cristallisation de l'amidon, grâce à la libération, au niveau d'un précurseur soluble, des chaînes mal positionnées, interdisant la formation de doubles hélices. Chez les procaryotes, les enzymes de débranchements sont impliqués dans le catabolisme du glycogène après action des phosphorylases permettant de libérer du glucose-1-phosphate plus riche en énergie que du glucose simple.

Au contraire, chez les *Archaeplastida* la première étape de dégradation de l'amidon cristallin est tout autre. En effet, il a été montré qu'elle n'était autre qu'une étape de déstabilisation de la structure du grain, par phosphorylation, sur les glucoses des carbones 6 et 3, respectivement, par la GWD (Glucan Water Dikinase) et la PWD (PhosphoWater Dikinase) (Hejazi *et al.*, 2012a). La caractérisation de mutants chez Arabidopsis montre que ces deux enzymes agissent séquentiellement et de manière non redondante : la GWD phosphoryle le C6 afin que la PWD puisse phosphoryler le C3 (Kötting *et al.*, 2005, Yu *et al.*, 2001). Notons que la GWD parait avoir un rôle primordial dans la dégradation du grain tandis que PWD semble avoir une activité plus secondaire. Malgré l'absence de redondance fonctionnelle, ces deux enzymes partagent le même domaine PPDK, contenant un domaine phosphohistidine (figure 11). La phosphorylation des glucoses s'effectue, dans les deux cas en deux étapes ; tout d'abord l'ATP sera clivé, afin de libérer le gamma phosphate, par la suite le phosphate sera transféré au glucane par l'intermédiaire du résidu histidine (Hejazi *et al.*, 2012b).

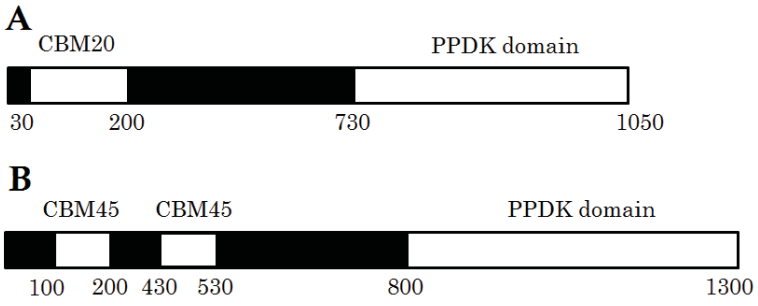


Figure 11: Schéma des différents domaines présents chez les enzymes PWD (A) et GWD (B). Il est facilement observable les deux domaines identiques PPDK, et les deux domaines de liaison au polysaccharide (CBM : carbohydrate binding module) qui diffèrent dans les deux enzymes.

La différence la plus marquante, au niveau des domaines présents dans ces protéines, et qui permet la spécificité d'action de ces deux enzymes, est sans aucun doute située au niveau des « carbohydrate binding module » (CBM), puisque la PWD possède un CBM20, tandis que la GWD possède deux CBM45. Ces CBM sont particulièrement intéressants à analyser, puisqu'ils ne sont pas considérés comme ayant une forte affinité avec les alpha-glucanes (Glaring *et al.,* 2011, Christiansen *et al.,* 2009), et que leur distribution, en particulier du CBM 45, parait limitée à certains organismes. Ainsi, si la PWD peut avoir été acquise à la suite d'un réarrangement entre une enzyme avec un module PPDK, et une enzyme possédant déjà un CBM20, la question reste ouverte pour la GWD, qui de plus semble présenter un domaine PPDK très proche phylogénétiquement de la PWD. Toutefois, les récentes recherches sur la phosphorylation du glycogène chez les eucaryotes, comme l'implication de la laforine dans la déphosphorylation du glycogène nécessaire à la bonne dégradation de celui-ci (Tagliabracci *et al.,* 2008), pourrait faire penser que ces deux enzymes, présentes chez les *Archaeplastida*, auraient pu évoluer naturellement, dès la mise en place de la structure cristalline, à partir du métabolisme de phosphorylation du glycogène présent dans la cellule hôte eucaryote (Cenci *et al.*, 2013). Par ailleurs, une enzyme de la même famille que la laforine existe aussi chez les *Archaeplastida* et est appelée SEX4 (Kötting *et al.*, 2009). Cette enzyme impliquée dans la déphosphorylation nécessaire des glucanes, permet l'hydrolyse des chaînes par les enzymes de type beta-amylase (Silver *et al.,* 2012). Ceci tend à démontrer que le mécanisme de dégradation de l'amidon dépend, chez les *Archaeplastida*, du mécanisme d'origine hôte de phosphorylation des glucanes, et n'a que peu de chance d'avoir une origine procaryote, puisque ces organismes ne possèdent pas, de manière référencée, un mécanisme de phosphorylation des polysaccharides.

Suite à ce mécanisme de phosphorylation des polysaccharides chez les eucaryotes les enzymes capables d'hydrolyser les polysaccharides rentrent en jeu. Alors que le glycogène paraît être dégradé, majoritairement, par des phosphorylases permettant de récupérer un glucose-1-phosphate, au lieu d'un glucose avec une amylase classique (Wilson *et al.,* 2010), l'amidon, lui, parait être dégradé, majoritairement, par des enzymes de type β-amylase qui libéreront des molécules de maltoses (Smith *et al.,* 2005). Toutefois, s'il est certain que ces différentes méthodes de catabolisme ont des intérêts biologiques différents pour les organismes, leurs intérêts relatifs exacts sont pour l'instant assez peu connus.

INTRODUCTION: TRANSITION FROM GLYCOGEN TO STARCH METABOLISM IN ARCHAEPLASTIDA

1. Starch and glycogen metabolism

Starch and glycogen define storage polysaccharides composed solely of glucosyl moieties that are predominantly linked by α-1,4 and α-1,6 bonds [1]. Starch, unlike hydro-soluble glycogen, is deposited as granules that have a semi-crystalline structure and unlimited size, and that are temporarily unavailable to hydro-soluble enzymes (Figure 1). Typically, two types of polyglucans are present in starch: amylopectin, the major branched component, is largely responsible for building the internal structure of the granule, whereas the minor moderately branched amylose is dispensable for granule formation and is synthesized by the granule-bound starch synthase (GBSS), the sole enzyme working in a starch-bound environment [1,2]. However, GBSS activity requires a preformed solid starch granule [3].

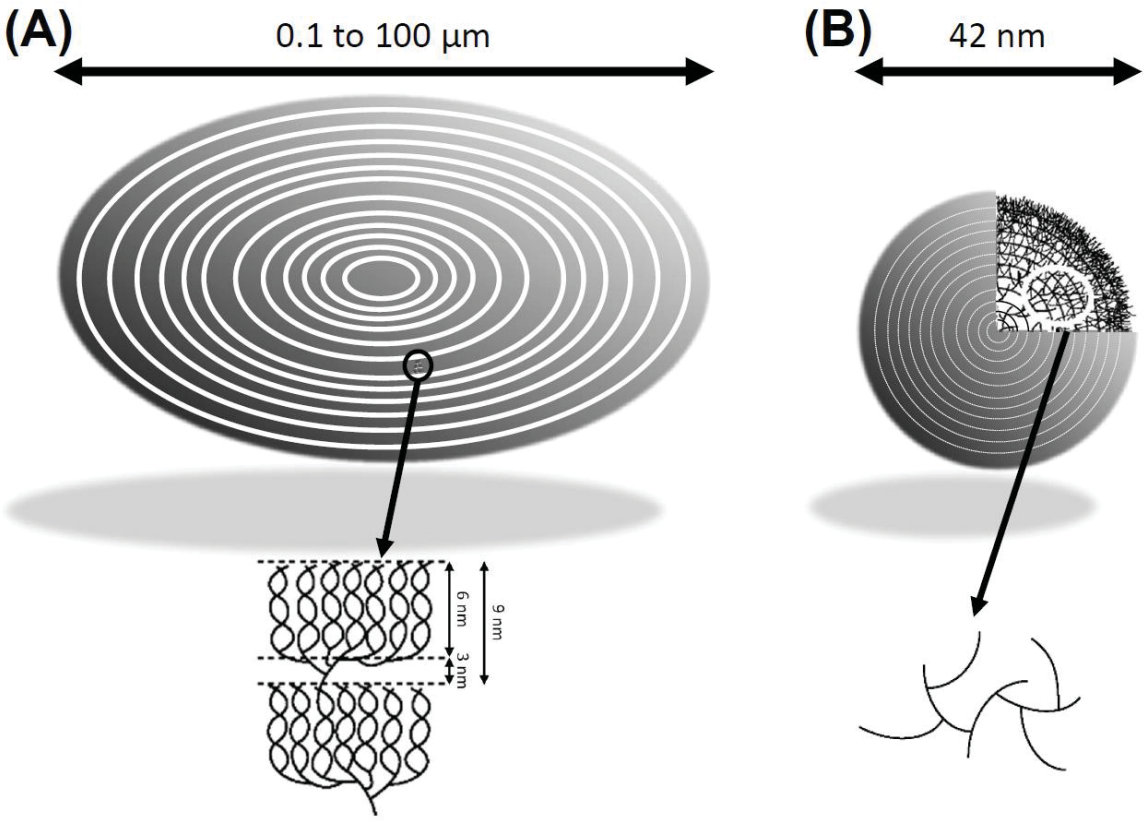


Figure 1. Comparative structure of glycogen and starch

Two different kinds of polysaccharide granules are compared. Solid semi-crystalline starch (A) is compared to a hydrosoluble glycogen particle (B). Sections of both types of organization are enlarged and shown below. In these enlarged sections the branching pattern of the chains is displayed. With an average of two α-1,6 linkages per glucan, a predictable maximum of 42 nm is obtained for the glycogen granule shown in (B) because of the crowding of the chains at the particle's periphery. Two typical amylopectin clusters are shown in the enlarged starch section (A). The concentration of α-1,6 branches at the root of

each of the two clusters yields a closer packing of glucans because each branch generates a novel chain. The proximity of the chains enables them to intertwine and associate as double helical structures. As shown in the enlarged clusters the chains align in parallel double helices. Furthermore, in the granule, each cluster aligns with the neighboring cluster expelling water from the aggregated structure, which collapses into the solid macrogranular structure of starch (for a review of storage polysaccharide structure (glycogen and starch) see [1]). Two distinct types of crystalline arrangements have been reported in starches called the A and B-types allomorphs.

The generation of starch rather than glycogen can be explained if a mechanism generating the concentrated distribution of branches, in certain regions, can be envisioned. The selective splicing out by debranching enzymes of those chains that do not conform to such a distribution (defined as glucan trimming, see text) is the only such mechanism presently described and supported by all published data. Definitive proof of this proposed mechanism can only be made when *in vitro* synthesis of starch granules will be achieved.

Studies of *Chlamydomonas (Chlamydomonas reinhardtii)*, cereal and *Arabidopsis* (*Arabidopsis thaliana*) mutants have established that during synthesis [4–8] a debranching enzyme is required for normal starch synthesis. It is generally assumed that this enzyme splices out those α-1,6 branches that prevent polysaccharide assembly to yield a crystalline structure. In the absence of this debranching enzyme, mutant plants and algae accumulate glycogen. Because semi-crystalline polysaccharides escape efficient degradation by the hydro-soluble enzymes of eukaryotic glycogen catabolism, we propose in this opinion that enzymes have co-evolved to degrade such structures. In all photosynthetic eukaryotes, the so-called glucan and phosphoglucan water dikinases (GWDs and PWDs) phosphorylate the hydrophobic crystalline structure of amylopectin [9–12]. These enzymes transfer the α-phosphate of ATP to the C3 or the C6 of glucosyl residues within the crystalline arrays [13]. This action appears to loosen hydrophobic crystals by generating more hydrophilic sections of the granule, which become accessible to hydro-soluble enzymes [14]. Mutants with a defective GWD or PWD are impaired in starch degradation and with time exhibit a starch-excess phenotype. Furthermore, both mutants are compromised in growth. In *Arabidopsis*, all these phenotypical features are more pronounced when GWD is lacking [9,15,16,17]. Leaves of a rice mutant lacking functional GWD possess approximately ten-fold higher levels of transitory starch having a low glucosyl 6-phosphate content and reduced grain yield but vegetative growth is similar to that of the wild type [17].

Starch metabolism specifically developed in photosynthetic eukaryotes and is found in the three major Archaeplastida lineages that emerged after plastid endosymbiosis: Glaucophyta, Rhodophyceae (red algae) and Chloroplastida (green algae and land plants) [18,19]. In all these lines, including those evolved from Archaeplastida through secondary

endosymbiosis, starch can be considered as a cytosolic carbohydrate store with no larger α-glucans being synthesized in plastids. However, an important exception is the Chloroplastida, which typically form starch in plastids only [18,19]. Only a few wild-type plants are capable of synthesizing both starch and glycogen-like polysaccharides inside plastids. For example, in *Cecropia peltata* starch is metabolized in photosynthesis-competent cells but plastidial biosynthesis of glycogen-like polysaccharides is restricted to a small population of morphologically and functionally highly specialized cells [20–22].

There is now a large body of evidence that suggests that the ancient pathway of starch metabolism evolved initially in the eukaryote host cytosol shortly after endosymbiosis and that the whole pathway was selectively directed to chloroplasts as the green algal lineage emerged [23,24]. Given that all Archaeplastida are monophyletic [25–26], as are most enzymes of starch metabolism [18,27], identifying the most parsimonious hypothesis explaining the distribution of starch metabolism enzymes in extant glaucophytes and in red or green algae is feasible (Figure 2) [25,26]. Heterotrophic eukaryotes unrelated to plastid endosymbiosis synthesize glycogen rather than starch and, hence, it is presumed that at the onset of plastid endosymbiosis glycogen accumulated in the host cytosol (Figure 2A) and that the transition to starch came shortly thereafter (Figure 2B). Reconstruction of cytosolic polysaccharide metabolism in the last common ancestor of all photosynthetic eukaryotes suggests that this pathway was used to export carbon from the cyanobiont to the host cytosol, thereby enabling selection of plastid endosymbiosis (Figure 2A) [18,19]. The results of a recent study indicate that key enzymes involved in the export of photosynthetic carbon from the evolving plastid defined virulence effectors secreted in the cytosol by intracellular *Chlamydia* pathogens [28,29]. Thus, the cyanobiont, the pathogen and the eukaryote were tied in a stable tripartite symbiosis at the onset of plastid endosymbiosis [28,29]. Photosynthetic carbon export to the eukaryotic host cytosol was central to the success of plastid endosymbiosis and a specific *Chlamydia* phylogenomic imprint can still be found in extant Archaeplastida genomes [30–32].

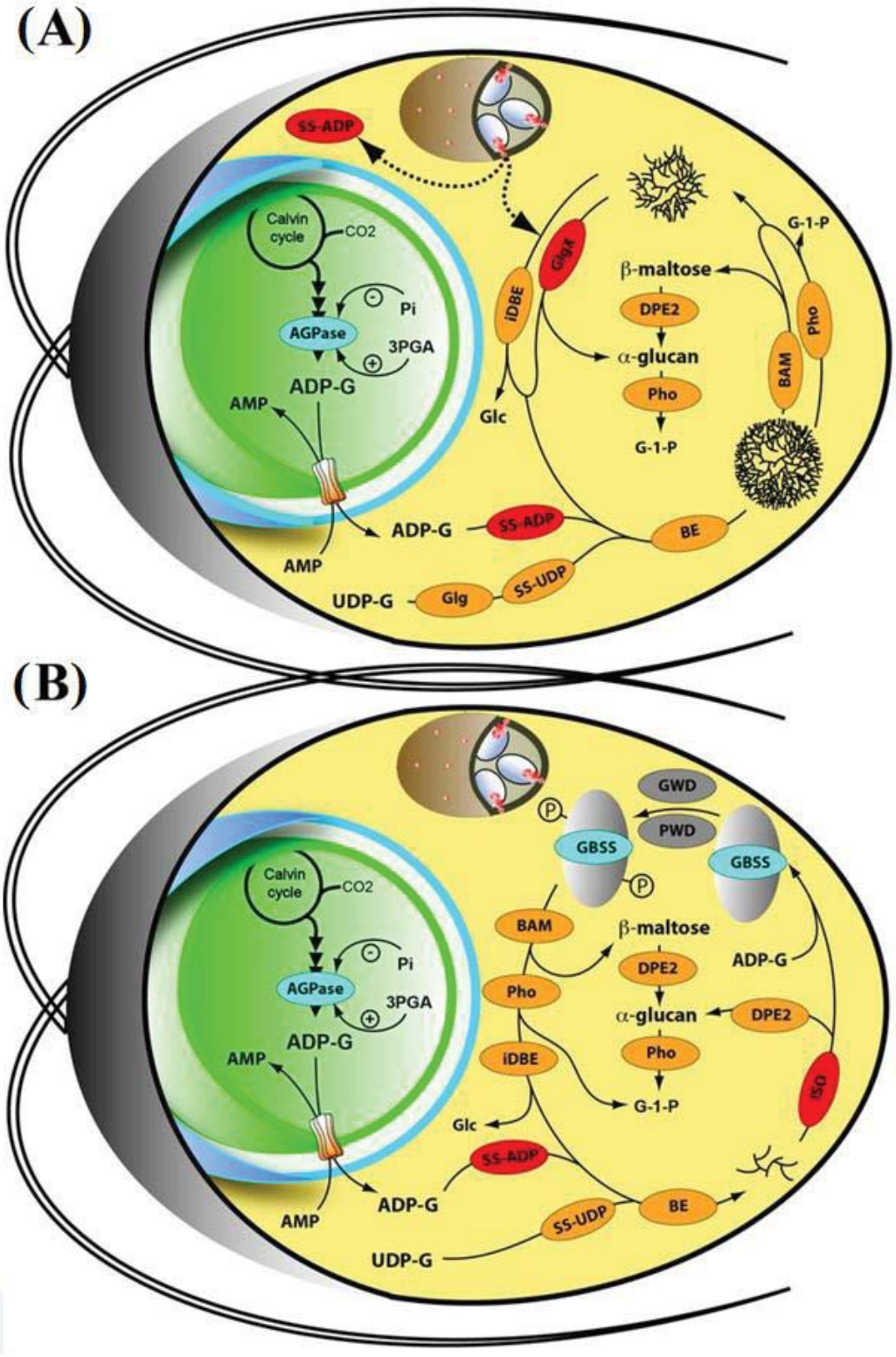


Figure 2. Reconstruction of storage polysaccharide metabolism in the common ancestor of Archaeplastida. **(A)** The storage polysaccharide network at the onset of endosymbiosis. In this reconstruction the transition to starch in the common ancestor cytosol has not yet occurred and glycogen has accumulated. The direct debranching enzyme (labeled GlgX) of bacterial phylogeny has not yet been recruited for amylopectin crystallization and is secreted as a virulence effector by *Chlamydia* pathogens (depicted as a brown inclusion vesicle containing

white pathogens attached to this membrane by their type-three secretion system, shown in orange, which is responsible for secretion of the SS-ADP and GlgX functions. The cyanobiont is shown in blue and green. The bacterial direct debranching enzyme still displays its ancestral bacterial function, detailed in Figure 3. The eukaryotic indirect debranching enzyme has a similar function. However, the chlamydial enzyme releases the maltotetraose outer chains (labeled α-glucan) in the cytosol, which may have been subjected to degradation by a combination of one of the two DPE2 amylomaltase isoforms still found in extant Rhodophyceae and Glaucophyta and of the glycogen phosphorylase. The cytosolic dual substrate pathway of glycogen accumulation relies on both UDP-glucose (UDP-G) generated through host biochemical networks in its cytosol according to host needs and on ADP-glucose (ADP-G) generated by the cyanobacterial ADP-glucose pyrophosphorylase (AGPase), which is activated by 3-phosphoglyceric acid (3-PGA) and inhibited by orthophosphate according to the networks and physiology of the cyanobiont. To be incorporated into cytosolic glycogen this substrate must be exported by a nucleotide-sugar translocator of host origin (shown in beige on the inner membrane of the cyanobiont) that exchanges ADP-glucose with AMP (for a review, see [18]). The ADP-glucose substrate in the cytosol must be incorporated through an ADP-glucose-specific glucan synthase (labeled SS-ADP). By contrast, the host UDP-glucose pools will be directed to glycogen according to the highly regulated eukaryotic UDP-glucose-specific glucan synthase (labeled SS-UDP). This enzyme, unlike the bacterial glucan synthase, requires a primer to elongate a glucan. This primer is defined by glycogenin, an autoglucosylating protein (GLG). The glucans elongated through both glucan synthases will then be branched into glycogen by the branching enzyme (BE). The glycogen outer chains will be degraded through either α-amylase (BAM) or glycogen phosphorylase (PHO) to generate maltose and glucose-1-P, respectively. The maltose will metabolized by the DPE2 amylomaltase. Enzymes of host phylogenetic origin are colored in beige; those of cyanobacterial origin in blue and those of chlamydial origin in red. At this stage the chlamydial genes encoded by the pathogens located in the inclusion vesicle have not been transferred to the host nucleus. **(B)** Cytosolic storage polysaccharide metabolism has been reconstructed by hypothesizing the simplest enzyme set that would explain the distribution of the genes of the starch metabolism network within the three distinct Archaeplastida lineages (for a review, see [18]). This early stage corresponds to the common ancestor after the transition from glycogen to starch has occurred. The glycogen metabolism network illustrated in (A) is a simplification of this reconstruction. The transition to starch required the duplication and evolution of the bacterial direct debranching enzyme into a functional isoamylase (iso), which is now encoded by the nucleus after LGT of the chlamydial gene to the host genome. This enzyme processes the branches generated randomly on the hydrophilic branched polysaccharides generated by branching enzymes. The debranched glucans (labeled α-glucan) are metabolized through a combination of DPE2-like amylomaltases and phosphorylases. Simultaneously a gene fusion between a carbohydrate binding module and a dikinase domain enables the phosphorylation and loosening of the otherwise undegradable amylopectin crystals (depicted as a circled P attached to the white starch granules). This fusion generates the archaeplastidal GWD–PWD novel activities (shown in gray), which were required to initiate starch catabolism through the α-amylase and phosphorylases (see above). The presence of polysaccharides aggregated into semi-crystalline starch granules enabled the binding and function of the cyanobacterial GBSS (shown bound to starch) responsible for amylose synthesis within the polysaccharide matrix. The late transfer of the cyanobacterial GBSS gene to the host nucleus suggests the maintenance of at least part of the cyanobacterial storage polysaccharide metabolism network until the transition to starch occurred. The cyanobiont and chlamydial inclusion vesicle are depicted as in (A). (B) is adapted from Ref. 19.

As intracellular pathogens, *Chlamydia* presumably secreted enzymes that metabolize storage polysaccharides in the host cytosol to induce glycogen synthesis at the beginning of the infectious cycle. At the end of this cycle, stores were mobilized by a secreted chlamydial glycogen phosphorylase, thereby bypassing the highly regulated eukaryotic glycogen phosphorylase (Figure 1) [28]. Hence stores were readily available to the pathogens because glycogen degradation was no longer controlled solely through the host.

In this Opinion article, we will focus on those events that we propose may have been involved in the evolution of starch metabolism from the preexisting glycogen metabolism network active in the host cytosol of plastid endosymbiosis. From detailed phylogenetic and biochemical considerations, we propose that the switch of glycogen to starch metabolism in the Archaeplastida ancestors was unique and required the concomitant appearance of a debranching enzyme generating insoluble crystalline polysaccharide structures and of a dikinase allowing their mobilization.

2. Recruitment of a direct debranching enzyme from *Chlamydia* pathogens may explain the switch from glycogen to starch

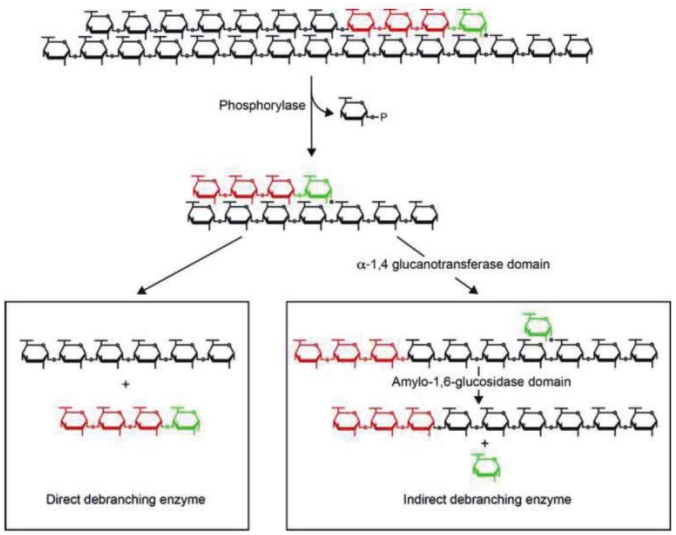


Figure 3. Comparative pathways of glycogen mobilization and debranching in eukaryotes and bacteria

Both heterotrophic eukaryotes and bacteria mobilize glycogen chiefly through the use of glycogen phosphorylase. Figure I shows, an outer branched oligosaccharide, such as those potentially existing at the periphery of a glycogen particle. In this example, two outer chains of 11 and 12 glucose residues, respectively, united by a common branch, are first shortened by glycogen phosphorylase, which generates Glc-1-P through a phosphorolytic reaction

attaching a phosphate on the C1 of the terminal glucose, thereby rupturing the α-1,4 glucosidic linkage. The enzyme remains active until it reaches the glucose residue situated four residues upstream from the α-1,6 branch and stops. Both eukaryotes and bacteria need to proceed by processing the residual polysaccharide using a debranching enzyme. Bacteria and eukaryotes use two different debranching mechanisms. Bacteria use the mechanism highlighted on the left by 'directly' attacking the α-1,6 branch with its two outer chains composed of four Glc residues. Enzymes with a high substrate specificity that will not (or poorly) hydrolyze the branch if the outer chains are longer than four are known as 'GlgX', a name coined after the corresponding *E. coli* locus encoding the best-studied enzyme of this type. The chain above the branch will be released in the form of maltotetraose, which needs to be metabolized by enzymes of malto-oligosaccharide metabolism (including α-1,4 glucanotransferases and malto-oligosaccharide phosphorylase). The chain shown beneath still attached to the glycogen particle will be further shortened by glycogen phosphorylase until the glycogen phosphorylase encounters the next branch where it will stop four Glc residues away. Other direct debranching enzymes that do not require a precise outer chain length to be active are found in bacteria. These are generally secreted to mobilize glycogen or starch-derived substrates present in the external medium. According to their ability to debranch loosely or tightly spaced branches the debranching enzymes are usually named isoamylases or pullulanases, respectively. Direct debranching enzymes have not been found in eukaryotic lineages unrelated to plastid endosymbiosis. In such eukaryotes, the product resulting from glycogen phosphorylase action (defined as glycogen with outer chains consisting of four Glc residues) will be processed as shown on the right-hand side of the drawing illustrated above. An enzyme called 'indirect debranching enzyme' containing two distinct catalytic sites will first hydrolyze the α-1,4 bond preceding the branch on the outermost chain and then transfer three Glc residues (maltotriose depicted in red) to the neighboring glycogen attached chain. This will generate a seven glucose residue long outer chain that can be further shortened by glycogen phosphorylase. The glucose unmasked at the branch (depicted in green) will then be attacked by the second α-1,6 glucosidase active site present on the indirect debranching enzyme and released from the glycogen particles. Importantly, no malto-oligosccharides are released by the indirect debranching enzyme during glycogen catabolism in eukaryotes and no enzymes able to catabolize malto-oligosccharides are therefore present in the cytosol of eukaryotes. This difference may be used by extant *Chlamydia* pathogens, which possibly trigger maltotetraose production through secretion of their own GlgX type of effector in the host cytosol, triggering the formation of a substrate that only they can further metabolize.

To ensure crystallization of vicinal α-glucan chains of amylopectin, clustering of branches is required (Figure 1). It has been proposed that isoamylase isozyme1 (isa1), which encodes the catalytic subunit of the isoamylase debranching enzyme, is chiefly responsible for trimming out the misplaced branches in a pre-amylopectin hydrophilic precursor, thereby leading to cluster formation and polysaccharide crystallization. This hypothesis, known as the "glucan trimming model" is consistent with a very large body of data [33-35]. However, definitive proof of this model would still require the setting up of an *in vitro* starch synthesizing system demonstrating the requirement for glucan trimming by isoamylase to generate starch-like granules. Nevertheless isoamylase has been proven to be required for

aggregation of starch granules *in vivo* in all plant systems tested. Detailed phylogenetic analysis (Figure 4) supports the notion that in all Archaeplastida direct debranching enzymes, including isoamylase but excluding pullulanases, originate from a lateral gene transfer event involving a chlamydial gene [30-32]. This gene is related to the bacterial GlgX debranching enzymes, which appear to be involved in glycogen catabolism [36]. In *E. coli*, GlgX-type enzymes are known to restrict their debranching activity to those outer chains which have already been shortened by glycogen phosphorylase [36,37]. Glycogen phosphorylase stops phosphorolysis four glucose residues upstream from the branch (see Figure 3 for details on debranching enzyme function in bacteria and eukaryotes). GlgX has high substrate specificity and will not debranch longer outer-chains. In bacteria this high specificity prevents futile cycles owing to the presence of branching enzymes that transfer oligosaccharides at the α-1,6 position of at least six glucose residues in length. The restrictive substrate specificity of GlgX will not allow degradation of these novel branches unless the ensuing outer chains have first been shortened by glycogen phosphorylase (Figure 3). In Chloroplastida there are three genes that encode distinct GlgX-derived debranching enzymes [33]. Isa3 may be defined as a catabolic enzyme because it has retained the restricted substrate specificity of GlgX [33]. Indeed, isa3 is involved in the degradation of oligosaccharides released from starch by the action of GWDs in the presence of both α-amylase and phosphorylase [8,38]. Despite all the problems of phylogenetic signal erosion witnessed in the trees, the phylogeny supports that Isa2 may have evolved from duplication of the isa3 gene to generate a non-catalytic 'regulatory' or 'scaffolding' subunit of the large-size heteromultimeric isoamylase enzyme containing the isa1 catalytic subunit [28] (Figure 4). However, the exact position of the isa1 group within the chlamydial isoamylase-GlgX clade is likely to remain obscure because of the high level of phylogenetic signal erosion and the ensuing poor bootstrap support of the relevant nodes. As mentioned above, the starch metabolism network was rewired from the host cytosol to the evolving chloroplast in the emerging green algae [23,24]. This redirection of the whole network into the stroma was a problematic process given that isolated duplications followed by transit peptide acquisition of genes duplicated from the cytosolic starch metabolism network would yield no selective advantage unless other enzymes of this pathway simultaneously underwent similar alterations [24]. See [24] for further discussion of a hypothesis to explain how the rewiring was nevertheless made possible by sequential steps leading to first the synthesis of oligosaccharides followed by that of glycogen. Finally, plastidial starch emerged and cytosolic starch metabolism was lost. Hence this stepwise process suggests that the chloroplastic polysaccharide turnover evolved later and

independently from the acquisition of starch metabolism in the cytosol [24]. It nevertheless made use of the genes encoding the enzymatic toolkit for cytosolic starch synthesis. Do we have any evidence that analogous mechanisms were functional for polysaccharide crystallization in the cytosol of the Archaeplastida common ancestor?

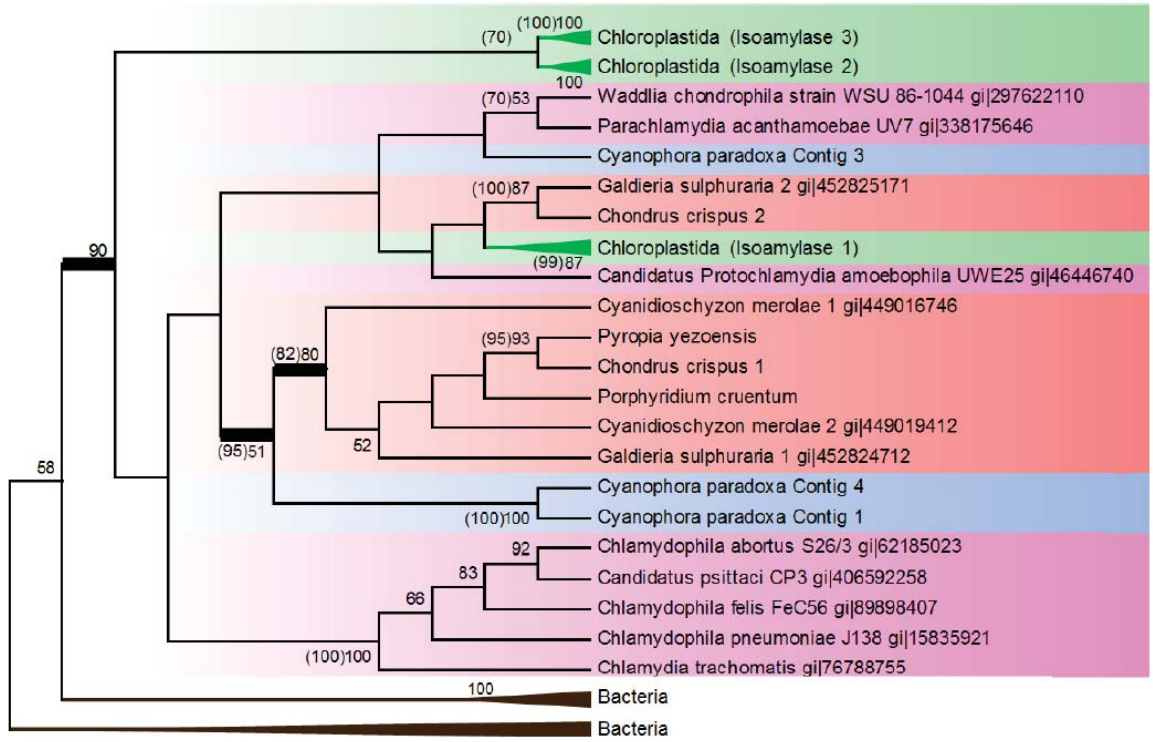


Figure 4. Maximum-likelihood phylogenetic tree of Archaeplastida and bacteria debranching enzymes. The tree shown is similar to those previously published and discussed in detail in [28]. After alignment and block selection, the best-fit model, according to ProtTest 3.2, was LG + G. The tree was calculated with PhyML, with 100 bootstrap replicates; bootstrap values higher than 50 are indicated without parentheses. The same alignment was run with PhyloBayes and the posterior probabilities of each node above 70 are indicated in parentheses. A highly supported node (shown in bold at bootstrap 90) groups all Archaeplastida isoamylase-like sequences and the *Chlamydia* into a monophyletic group in agreement with the proposed chlamydial origin of this gene [28]. Cyanobacterial sequences are not found close to the root of this group. CAZymes (Carbohydrate Active enZymes), such as the GlgX-isoamylase type of glucosyl hydrolases discussed here, are moderately constrained with respect to sequence during evolution. Erosion of the phylogenetic signal following lateral gene transfer is well documented in such cases [28]. Despite this, a well-supported monophyletic group of enzymes from Rhodophyceae at bootstrap 80 is evident, which groups together with Glaucophyta at a significant albeit moderate (52) bootstrap (both of the relevant nodes are highlighted in bold). We propose in this Opinion that this monophyletic group of glaucophyte and red algae enzymes reflects the first LGT of the chlamydial sequence to the nucleus of the common ancestor of Archaeplastida. Given that this ancestor as deduced from the *Cyanophora paradoxa* genome sequence [26] still contained an iDBE (see Figure 3), which plays a similar role in eukaryotes to the role played by GlgX in bacterial glycogen degradation, we believe that this LGT was selected because it conferred a novel desirable trait to these organisms: the ability to accumulate starch. Other possibly

polyphyletic LGTs of chlamydial sequences are shown in the tree. We propose that such events were selected because they either conferred a GlgX-like enzyme function to Rhodophyceae, which have since lost the iDBE, which would have been redundant in glaucophytes, or because they enabled the building of a more complex and efficient heteromultimeric isoamylase involved in polysaccharide degradation. As to the green algae and plant ISAs, phylogenetic signal erosion and the ensuing low bootstrap support prevents us from proposing a possible sequence of events. The high level of signal erosion seen in Chloroplastida can be easily explained because the enzymes of starch metabolism were redirected into a novel environment, the plastid stroma [23,24], thereby generating yet another round of signal erosion in the Chloroplastida [28].

Analysis of the gene content with respect to starch metabolism in Rhodophyceae and Glaucophyta [18,26,39,40] has revealed that genes encoding GlgX-isoamylase-like sequences are present in several distinct copies in Rhodophyceae and Glaucophyta, which suggests the presence of different isoforms with different functions. The starch metabolism network in red algae is extremely simple, unlike that of the green algae and land plants [18,39,40]. Usually, it contains only one gene encoding a single isoform for each type of catalytic activity involved in polysaccharide metabolism. Isoamylase-GlgX and an enzyme similar to the 'disproportionating isozyme 2' (DPE2; also designated as amylomaltase or transglucosidase) are noticeable exceptions in these algae. The presence of an isoamylase would have required a novel α-glucanotransferase to catabolize efficiently those chains released during pre-amylopectin trimming. In green algae and land plants, this function is under the control of the 'disproportionating isozyme 1' (DPE1), which evolved selectively in the emerging green lineage by lateral gene transfer (LGT) of an unknown proteobacterial source [23]. All glaucophyte and rhodophyte genomes lack DPE1 but contain several copies of genes encoding DPE2, an enzyme of eukaryotic affiliation that is hypothesized to be specific for the assimilation of α-maltose produced by the eukaryotic α-amylases in plants [18,26,38,39]. Two groups of DPE2 sequences have been found in Rhodophyceae and Glaucophyta, one of which is closer to DPE2 from green plants and algae [41]. It is tempting to speculate that the other transferase could supply the otherwise missing transferase activity with a function analogous to DPE1.

Biochemical data support the existence of a multimeric isoamylase activity in the model glaucophyte *Cyanophora paradoxa*, the activity of which shows a similar substrate specificity as that seen in the plant enzyme [42]. *Cyanophora paradoxa* also contains an indirect debranching enzyme (Figure 3) of eukaryotic affiliation [26] that has the same biochemical function as GlgX in eukaryotic glycogen metabolism [43]. However, it defines a complex bifunctional enzyme producing glucose and modified polysaccharides, which is

unlikely to supply a function analogous to the bacterial direct debranching enzymes for the trimming of pre-amylopectin (Figure 3). We believe that this enzyme would be redundant with a GlgX-ISA3-type enzyme in glaucophytes. In this Opinion article, we propose the following sequence of events as a theoretical framework for future experimentation and possible validation. At the onset of endosymbiosis (Figure 2A) the storage polysaccharide network contained both an indirect debranching enzyme (iDBE) and a GlgX enzyme acting on glycogen shortened by glycogen phosphorylase and α-amylase. The GlgX enzyme was coded by the chlamydial symbiont or pathogen as an effector protein and was required to import maltotetraose into the Chlamydiae-containing inclusion vesicle but was not *per se* required for glycogen catabolism by the host. A copy of the chlamydial gene was then transferred to the nucleus and through one or several point mutation(s) became able to accommodate substrates with chains longer than four glucose residues. This emerging isoamylase activity could define the ancestor of the monophyletic group (Figure 4) of most of the Rhodophyceae and Glaucophyta sequences. Hence starch metabolism would be a monophyletic acquisition in Archaeplastida. The ancestral isoamylase was moderately efficient at assembling polysaccharides leading to a moderately hydrophobic structure intermediate between glycogen and starch. This has been described for the red alga *Porphyridium sordidum* [44], which may only harbor this unique isoform of GlgX-isoamylase (note that *Porphyridium purpureum* defines the only Archaeplastida with only one isoamylase isoform reported). In a polyphyletic fashion, more efficient isoamylases were generated either by duplicating and subfunctionalizing the ancestral gene or by recruiting analogous LGT from the chlamydial source. Increased efficiency may have been acquired by regulatory or scaffolding subunits akin to the ISA2 gene of plants and green algae, thereby leading to a more efficient heteromultimeric enzyme as suggested by the results obtained with *Cyanophora paradoxa* [42]. In the common ancestor of the red and green algae, loss of iDBEs required their substitution by the chlamydial GlgX function. There again either duplication of the isoamylase regulatory or catalytic subunits or a novel LGT from Chlamydiae could have generated this ancestral GlgX-ISA3-like enzyme. Possibly, such an enzyme could define the ancestor of the ISA2 and ISA3 genes of green algae and plants. The appearance of the ancestral isoamylase required both a novel α-1,4 glucanotransferase (discussed above) and of an of enzyme enabling degradation of the more hydrophobic polysaccharides. This speculative scenario can be readily tested by the purification of the isoamylase and GlgX-like proteins from Rhodophyceae and Glaucophyta. It could further be

ascertained by examining the biochemical properties of the corresponding recombinant proteins.

3. Phosphorylation of starch and glycogen

Starch contains low levels of monophosphate esters (Figure 5) introduced during both starch biosynthesis and degradation but esterification rates are likely to vary [45–48]. All phosphorylating enzymes belong to the small group of dikinases, which, unlike the hundreds of kinases, use ATP as a dual phosphate donor (converting ATP to AMP) and two final acceptors of the two phosphates.

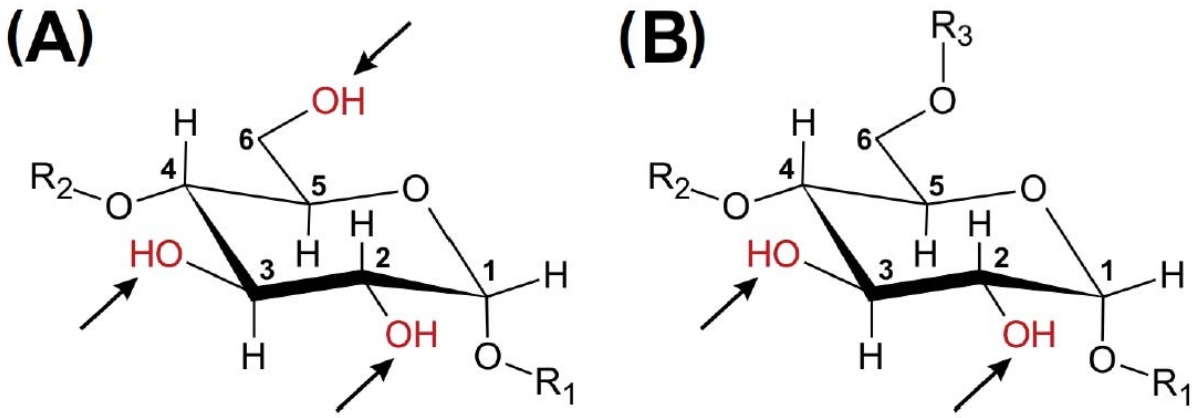


Figure 5. Possible phosphorylation sites of glucose residues within amylopectin and glycogen. An α-1,4-linked glucose residue present within a glycogen or starch chain is shown on the left side and a glucose at the branch is shown on the right. The OH sites available to phosphorylation are highlighted in red.

One important function of starch-related dikinases appears to render an otherwise inaccessible crystalline substrate more hydrophilic and accessible to hydro-soluble starch metabolizing enzymes. Given that linear α-glucans tend to spontaneously interact and crystallize it is not inconceivable that the phosphorylation of highly ordered chains associated with phase transition requires more energy than the phosphorylation of a soluble sugar and, therefore, is mediated by a dikinase rather than by a kinase [49,50]. Crystalline maltodextrin representing either the A- or the B-type allomorph [1, Box1] have been used as a model for native starch granules to analyze the function of glucan phosphorylation. When incubated with recombinant GWD (or PWD) and ATP, phosphorylation rates by far exceeded those of

solubilized substrates. Phosphorylation of the insoluble crystalline maltodextrin resulted in an almost complete dissolution of both phosphorylated and neutral maltodextrins [50]. These results underline the importance of phosphorylation-mediated structural alterations of α-glucans. They do, however, not exclude some more specific effects that the monophosphate esters may exert on distinct carbohydrate-active enzymes under *in vivo* conditions. In addition, phosphorylation of α-glucan chains may favor reactions that require the single-chain state (such as branching) by preventing the spontaneous formation of double helices. If so, phosphorylation also exerts a distinct function during starch biosynthesis [47,48]. Furthermore, more complex down-stream effects of GWD have been reported for wheat endosperm [51] which are not discussed here.

As exo-acting catabolic enzymes, such as β-amylase, cannot bypass phosphorylated glucosyl residues in α-glucans, complete degradation of the storage polyglucans requires an ongoing cycle of phosphorylating and dephosphorylating reactions. This cycle is now widely accepted to occur during starch degradation. Presumably, the starch (de)phosphorylating proteins evolved in the context of polysaccharide crystallization as mediated by the chlamydial GlgX-derived isoamylase. As mentioned above, the dikinases have selectively emerged in Archaeplastida and are present in all starch-storing eukaryotes. Interestingly in those few instances where an ancestrally starch-storing clade has evolved to glycogen metabolism (such as the red alga *Galdieria sulphuraria* which accumulates glycogen while other red algae synthesize starch), dikinases were also lost. [44,52] Hence the correlation between the dikinases and crystallinity of amylopectin remains absolute in all eukaryotes. Furthermore, to date, these enzymes have not been found in bacteria, which suggests that they could have first evolved in the Archaeplastida ancestor cytosol shortly after endosymbiosis. We propose that this innovation was possible because of a pre-existing eukaryotic phosphoglucan metabolism.

In *Arabidopsis*, two genes enzymes have been identified that encode plastidial starch-related dikinases. The two plastidial starch phosphorylating enzymes are the glucan, water dikinase (GWD; *At1g10760*; also designated as GWD 1) phosphorylating at C6 and the so-called phosphoglucan, water dikinase (PWD; *At5g26570*; also named GWD 3) phosphorylating at C3 (for review, see [50]). Both gene products possess an N-terminal transit peptide and, following topogenesis, the two proteins have access to plastidial starch granules. Both dikinases are large multidomain yet monomeric proteins and mediate a series of phosphate transfer reactions thereby converting ATP to AMP. First, they transfer the γ-

phosphate of ATP to the dikinase protein and then to water yielding orthophosphate. Second, they transfer the β-phosphate to a conserved histidine residue of the dikinases. Subsequently, the phosphohistidine formed serves as a phosphate donor for the phosphorylation of either the C6 or the C3 position of the glucosyl residue to be phosphorylated [13,50]. Mutants lacking the conserved histidine residue are unable to phosphorylate α-glucans but capable of transferring the γ-phosphate from ATP to water [53].

The two dikinases share a similar organization given that the C-terminal domains contain an ATP-binding motive and the conserved histidine residue. At least one copy of a carbohydrate-binding module (CBM) is located at the N-terminal region (for details, see [50]). The CBMs of GWD and PWD differ: PWD possesses a CBM20 whereas the GWD belongs to the CBM45 family, members of which have been identified in only a few plant proteins and appear to have a relatively low affinity towards α-glucans [50,54]. The entire (de)phosphorylation cycle, which is essential for undisturbed starch turnover is likely to require reversible binding of the dikinases to starch granules, which, in principle, can be mediated by structural alterations of the polysaccharides and, or proteins.

The occurrence of different types of CBMs in GWD and PWD suggests that the two dikinases interact with different carbohydrate targets. This suggestion is supported by several studies: first, the phenotype of *Arabidopsis* mutants deficient in GWD is more severe than that of PWD-deficient lines, indicating that the two dikinases do not exert redundant functions [55,56]. Second, under *in vitro* conditions recombinant GWD acts on both native starch granules from wild-type plants and on crystalline maltodextrins whereas the enzymatic action of PWD strictly relies on a preceding phosphorylation by GWD (thus acting downstream of GWD [50,57]. Third, PWD forms a major proportion of monophosphorylated α-glucans. This demonstrates that this enzyme is chiefly phosphorylating neutral glucans. This is consistent with the view that the GWD-mediated prephosphorylation alters the structure even of neutral glucan chains, thereby generating a suitable substrate for PWD [50,57].

Little is known about the selective functions of analogous proteins in the cytosol of Glaucophyta and Rhodophyceae. In most, but not all, cases, several isoforms of GWD or PWD candidate genes have been found, suggesting similar functional specialization. [26, 39, 40]

The modular composition of GWDs and PWDs may suggest that these genes could have appeared shortly after endosymbiosis by gene fusions of a domain of a related dikinase [such as pyruvate, orthophosphate dikinase (PPDK) and pyruvate, water dikinase;PEP synthetase (PPS)] of unknown origin and a pre-existing starch-binding domain. We propose that the starch-related dikinases evolved, replacing pre-existing glycogen kinases that were unable to efficiently phosphorylate crystalline regions of amylopectin.

The frequency of such a domain fusion event would have been sufficient to propose that this was generated at the same time the first mutations required to change a preexisting GlgX debranching enzyme into a functional isoamylase (see above) were selected.

Glycogen phosphorylation in eukaryotes is a novel, hotly debated and fast moving field. Monophosphate esters were proposed to result from an unavoidable side reaction of glycogen synthase (GS) [58–60]. However, a recent study has shown that the same phosphorylation sites are used in glycogen and starch (Figure 5) and that monoesterification is not caused by either a side reaction of GS or of glycogen phosphorylase [61]. Thus, it is reasonable to assume that glycogen phosphorylation is due to distinct but as yet unknown enzymes that are functional during both biosynthesis and degradation of glycogen. Glycogen phosphorylating enzymes may not necessarily be dikinases and, therefore, may not be discovered when assuming sequence similarities to plant GWDs or PWDs. Hence Archaeplastida adapted a preexisting eukaryotic machinery of glycogen phosphorylation to allow degradation of crystalline polysaccharides. The function of eukaryotic glycogen phosphorylation (as well as that of starch phosphorylation during biosynthesis) is not fully understood. In glycogen, glucosyl 6-phosphate residues are unevenly distributed and more concentrated in the interior parts of the glycogen [61]. It is, however, uncertain whether subsequent dephosphorylating reactions strongly affect the phosphorylation pattern observed in glycogen. Based on circumstantial evidence [61] one could speculate that phosphorylation of α-glucan chains is indirectly linked to chain branching and, thereby, is relevant for the structure of the polyglucans.

In any case, it is tempting to speculate that the evolving plants have capitalized on the biochemical capacity displayed by the host of plastid endosymbiosis to phosphorylate glycogen and have created a novel pathway for metabolizing the semi-crystalline storage polysaccharides.

4. Dephosphorylation of starch and of glycogen

Although glycogen phosphorylating enzymes have not been identified yet, a mammalian protein dephosphorylating polyglucans has been known since the late 1990s [62]. The protein was named laforin because if it is not functional it is one of the two main causes of Lafora disease. It is a recessively inherited severe epilepsy afflicting approximately 1 in 200,000 individuals worldwide and arguably one of the severest known diseases. Many tissues of patients suffering of Lafora disease accumulate malformed polyglucans which are poorly branched, hyperphosphorylated, and insoluble [63,64]. Lafora disease is caused by loss-of-function mutations of the gene that encodes the laforin glycogen phosphatase [62] or of the gene encoding the malin ubiquitin E3 ligase [65], which appears to function in the regulation of laforin [66].

Laforin belongs to the large group of dual specificity phosphatases (DSPs), which dephosphorylate both phosphotyrosine and phosphoserine and phosphothreonine residues and, in some cases, also act on non-proteinaceous substrates [67]. Laforin is the only known human DSP possessing a CBM (CBM20 family). In plants, a plastidial DSP has been identified that carries both a CBM and a DSP motive. Initially, this protein was named PTPKIS1, assuming that phosphoproteins were the actual target of the phosphatase [68,69]. In *Arabidopsis*, the same locus was identified by screening chemically mutagenized lines for a starch-excess (SEX) phenotype and the gene product was designated as SEX4 [49,69,70]. Despite some phenotypical differences, the results of the screen strongly suggest that in higher plants starch is normally turned over provided both phosphorylating and dephosphorylating enzyme activities are functional [50].

Acting on phosphorylated glucans, laforin and SEX4 are less selective than GWD or PWD. Both phosphatases hydrolyze monophosphate esters at C3 and at C6 and also act on hydrosoluble α-glucans [49,70–73]. Likewise, glycogen from laforin-deficient mice contains elevated levels of glucosyl 6-phosphate residues [60]. Possibly, laforin and, or SEX4 also exert various actions on (phospho)proteins in addition to that on phosphorylated α-glucans [59]. Two other plastidial DSPs from *Arabidopsis* that have been recently characterized will not be discussed here. For review, see [51].

In vivo laforin and SEX4 appear to be functionally similar and functional human laforin complements the SEX4-deficient *Arabidopsis* mutant [72,73]. However, the two proteins are not orthologs. In SEX4, the CBM20 and DSP motifs are located at the C- and N-

termini, respectively, but the intramolecular order of the two domains is reversed in laforin [59,74]. Thus, laforin and SEX4 appear to be generated by independently performed fusions of domains.

True laforin orthologs have been reported for red algae such as *Chondrus crispus*, *Porphyridium cruentum* and *Cyanidioschyzon merolae* (and possibly also *Cyanophora paradoxa*) [18,26,39,40] . Apparently, during the targeting of the starch pathway to the plastid a novel laforin-like protein was generated but the reasons for this novelty remain unclear.

Recently, two additional plastidial starch-related dual-specificity phosphatases have been identified in *Arabidopsis thaliana* that are designated as Like-Sex-Four1 (LSF1) and Like-Sex-Four2 (LSF2). *Arabidopsis* mutants lacking functional LSF1 or LSF2 possess a starch-excess phenotype. For review see [50]. Based on *in vitro* studies, LSF1 does not exhibit a noticeable phosphatase activity when assayed with phosphorylated α-glucans. Likewise, transitory starch of the LSF1-deficient mutants possesses similar glucose-based contents of glucose 6-phosphate and glucose 3-phosphate levels [75]. Currently, the actual biochemical function of LSF1 is not known. By contrast, LSF2 possesses a phosphatase activity selectively hydrolyzing monophosphates at C3 of starch-related glucosyl moieties [76]. As opposed to SEX4 and LSF1, the phosphatase LSF2 does not contain a classical CBM that is separated from the catalytic domain. LSF2 does, however, undergo multiple interactions with amylopectin. Site-specific mutations and structural analyses of LSF2 revealed that all glucan binding sites are located within the catalytically active phosphatase domain [77]. These results have important implications for the search of further phosphatases acting on phosphorylated α-glucans.

As discussed above for the phosphorylating enzymes, we propose that chloroplast-containing eukaryotic cells established novel starch-dephosphorylating enzymes and adapted these enzymes to specific features of starch (and, possibly, to the novel intracellular compartmentalization). However, the novel enzymes were generated using a pre-existing mode of removing monophosphates.

5. Why do Archaeplastida accumulate starch?

When considering the physico-chemical differences of starch and glycogen the question arises as to the possible reasons underlying the evolution of glycogen metabolism into starch. At first one might propose that the vast amounts of carbon made available by photosynthesis required a novel packing of glucose into insoluble stores. If so, one would expect that all cyanobacteria accumulate starch-like polysaccharides. However, there are only a few cyanobacteria synthesizing starch-like granules and the vast majority of cyanobacteria remain *bona fide* glycogen accumulators [78]. Looking closer at the physiology of starch-accumulating cyanobacteria, it appears that those lineages metabolizing semi-crystalline polysaccharides are all unicellular and diazotrophic. Because nitrogenase is inhibited or inactivated by molecular oxygen, unicellular diazotrophy can only be sustained if phototrophy and diazotrophy are temporally separated by tight circadian clock control of cellular metabolism. Polysaccharides synthesized in the light are metabolized in darkness to reach anoxia by respiration of large amounts of glucose. Simultaneously, vast amounts of ATP and reducing power are provided to nitrogenase. Hence cyanobacteria with such a physiology require a larger carbon store compared with other cyanobacteria [18]. However, if more carbon can be stored in starch rather than in glycogen why is not starch a more widespread form of storage in eukaryotes and bacteria? The answer may rest in the ease with which peripheral glucosyl residues may become available to cellular metabolism. One can speculate that the flexibility offered by glycogen breakdown may not be matched by starch and that turnover of carbon in the light may define a desirable fine-tuning mechanism for photosynthesis optimization. Hence there may be a balance between the selective advantages offered by flexible glycogen mobilization and the additional efforts associated with storage of starch.

Although it is easy to understand why unicellular diazotrophic cyanobacteria may have resorted to starch metabolism it is less obvious why the Archaeplastida ancestor evolved semi-crystalline starch. First, the cytosolic localization of the storage polysaccharide may have prevented its use to optimize photosynthesis through carbon turnover in the light. Hence the advantages offered by the flexibility of glycogen degradation may not have applied as in cyanobacteria. Second, because *Chlamydia* pathogens may have been involved by setting up the metabolic link between the protoplastid and its host, it is reasonable to assume that the pathogens may have had the ability to tap directly into the host glycogen stores [26]. Hence the eukaryotic host had lost its monopoly on the mobilization of its own glucose stores.

Hence, the switch from glycogen to starch may have been prompted by the advantage afforded by an increase in carbon sink strength in the cytosol and because solid starch escaped direct mobilization of storage by *Chlamydia* pathogens. Indeed, the initial step of starch mobilization was then defined by the host glucan water dikinases with no direct access to these stores by the unregulated chlamydial effector glycogen phosphorylases.

Finally, the escape of the glucose stores from direct degradation by the glycogen catabolism enzymes enabled the evolution of the first mechanisms integrating storage polysaccharide mobilization in the cytosol, to photosynthate supply by the cyanobiont. The highly integrated pathway of host glycogen catabolism control was not tailored to take into account the timing and amounts of carbohydrates afforded by cyanobacterial synthesis. By the evolution of a novel first step defined by the required phosphorylation of the semi-crystalline glucans, natural selection allowed the evolution of novel circadian clock-mediated controls specifically exerted on this novel biochemical step without interfering with the highly integrated phosphorylation cascades responsible for host glycogen phosphorylase regulation. This innovation also facilitated the evolution of novel diurnal assessment mechanisms informing the host about the extent of the stores supplied by photosynthesis. These would have facilitated the emergence of a novel regulation of storage polysaccharide metabolism, taking into account the amount and activity of the biochemical networks of the cyanobiont.

The results of phylogenetic analysis of the starch debranching enzyme are consistent with a monophyletic acquisition of starch in the cytosol of the Archaeplastida ancestor.[28] However, the detailed biochemical properties of all the distinct GlgX isoamylase-type of debranching enzymes from the Rhodophyceae and Glaucophyta shown in Figure 4 need to be ascertained to confirm this idea.

Acknowledgements

This research was funded by the French Ministry of Education, the Centre National de la Recherche Scientifique (CNRS), the Agence Nationale pour la Recherche (ménage à trois), by the European Union, by the Region Nord Pas de Calais.

References

1. Buléon, A. *et al.* (1998) Starch granules: structure and biosynthesis. *Int. J. Biol. Macromol*. 23, 85-112

2. Ball, S., van de Wal, M. and Visser, R. (1998) Progress in understanding the biosynthesis of amylose. *Trends Plant Sci.* 3, 462-467

3. Dauvillée, D. *et al.* (1999) Novel starch-like polysaccharides are synthesized by a soluble form of granule-bound starch synthase in glycogen accumulating mutants of *Chlamydomonas reinhardtii. Plant Physiol.* 119, 321-330

4. James, M. G, Robertson, D. S. and Myers, A. M. (1995) Characterization of the maize gene sugary1, a determinant of starch composition in kernels. *Plant Cell* 7, 417-429

5. Mouille, G. *et al.* (1996) Preamylopectin Processing: A Mandatory Step for Starch Biosynthesis in Plants. *Plant Cell* 8, 1353-1366

6. Nakamura, Y. *et al.* (1997) Correlation between activities of starch debranching enzymes and α-polyglucan structure in endosperms of sugary-1 mutants of rice. *Plant J.* 12, 143-153

7. Zeeman, S. C. *et al.* (1998) A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. *Plant Cell* 10, 1699-1712

8. Wattebled, F. *et al.* (2005) Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. *Plant Physiol.* 138, 184–195

9. Lorberth, R. *et al.* (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. *Nat. Biotechnol.* 16, 473–477

10. Ritte, G. *et al.* (2002) The starch-related R1 protein is an α-glucan, water dikinase. *Proc. Natl. Acad. Sci.* 99, 7166-7171

11. Kötting, O. *et al.* (2005) Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. *Plant Physiol.* 137, 242–252

12. Baunsgaard, L. *et al.* (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated alpha-glucans and is involved in starch degradation in Arabidopsis. *Plant J.* 41, 595–605

13. Ritte, G. *et al.* (2006) Phosphorylation of C6 and C3 positions of glucosyl residues in starch is catalyzed by two distinct dikinases. *FEBS Lett.* 580, 4872-4876

14. Hejazi, M. *et al.* (2008) Glucan, water dikinase phosphorylates crystalline maltodextrins and thereby initiates solubilization. *Plant J.* 55: 323–334

15. Yu, T. S. *et al.* (2001) The Arabidopsis *sex1* mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. *Plant Cell* 13, 1907–1918

16. Vriet, C. *et al.* (2010) A suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. *Plant Physiol.* 154, 643-655

17. Hirose, T *et al.* (2013) Disruption of a rice gene for α-glucan water dikinase, OsGWD1, leads to hyperaccumulation of starch in leaves but exhibits limited effects on growth. *Frontiers in Plant Sci.* 4, 147

18. Deschamps, P. *et al.* (2007) Metabolic symbiosis and the birth of the plant kingdom. *Mol. Biol. Evol.* 25, 536-548

19. Ball, S. *et al.* (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J. Exp. Bot.* 62, 1775–1801

20. Rickson, F.R. (1971) Glycogen plastids in Müllerian Body cells of Cecropia peltata – a higher green plant. *Science* 173, 344-347

21. Marshall, J.J. and Rickson F.R. (1973) Characterization of the α-D-glucan from the plastids of Cecropia peltata as a glycogen-type polysaccharide. *Carbohydrate Res.* 28, 31-37

22. Bischof, S. *et al.* (2013) *Cecropia peltata* accumulates starch or soluble glycogen by differentially regulating starch biosynthetic genes. *Plant Cell* 25, 1400–1415

23. Deschamps, P. *et al.* (2008) Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. *Genetics* 178, 2373-2387

24. Deschamps, P. *et al.* (2008) The relocation of starch metabolism to chloroplasts: when, why and how. *Trends Plant Sci.* 13, 574-582

25. Rodríguez-Ezpeleta, N. *et al.* (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* 15, 1325-1330

26. Price, D. C. *et al.* (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335, 843–847

27. Patron, N. J. and Keeling, P. K. (2005) Common evolutionary origin of starch biosynthetic enzymes in green and red algae. *J. Phycol.* 41, 1131-1141

28. Ball, S. G. *et al.* (2013) Metabolic effectors secreted by bacterial pathogens; essential facilitators of plastid endosymbiosis? *Plant Cell* 25, 7–21

29. Baum, D. (2013) The Origin of Primary Plastids: A Pas de Deux or a Ménage à Trois? *Plant Cell* 25, 4-6

30. Huang, J. and Gogarten, J. P. (2007) Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8, R99

31. Becker, B., Hoef-Emden, K. and Melkonian M. (2008) Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evolutionary Biology* 8, 20

32. Moustafa, A., Reyes-Prieto, A. and Bhattacharya, D. (2008) Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* 3, e2205

33. Hussain, H. *et al.* (2003) Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. *Plant Cell* 15:133–149

34. Ball, S.G. *et al.* (1996) From glycogen to amylopectin: a model explaining the biogenesis of the plant starch granule. *Cell* 86, 349-352

35. Myers, A. M. *et al*. (2000) Recent progress toward understanding biosynthesis of the amylopectin crystal. *Plant Physiol.* 122, 989-998

36. Dauvillée, D. *et al.* (2005) Role of the *E. coli glgX* gene in glycogen metabolism. *J. Bacteriol.* 187, 1465-1473

37. Jeanningros, R., N. *et al.* (1976) Purification and properties of debranching enzyme from E. coli. *Biochim. Biophys. Acta* 438, 186-199

38. Delatte, T. *et al.* (2006) Evidence for distinct mechanisms of starch granule breakdown in plants. *J. Biol. Chem.* 285, 12050-12055

39. Collén J. *et al.* (2013) Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* 110, 5247-5252

40. Bhattacharya, D. *et al.* (2013) Genome of the red alga *Porphyridium cruentum. Nature Commun.* 4, 1941

41. Arias, M. C. *et al.* (2012) Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants. *Mob. Genet. Elements* 2, 81-87

42. Plancke, C. *et al.* (2008) Pathway of cytosolic starch synthesis in the model glaucophyte *Cyanophora paradoxa*. *Eukaryot. Cell* 7, 247-257

43. Teste M. A. *et al*. (2000) The *Saccharomyces cerevisiae YPR184w* gene encodes the glycogen debranching enzyme. *FEMS Microbiol. Lett.* 193, 105-110

44. Shimonaga, T. *et al.* (2008). Variations in storage α-glucans of the Porphyridiales (Rhodophyta). *Plant Cell Physiol*. 49, 103-116

45. Hizukuri, S. *et al.* (1970) Studies on starch phosphate. 1. Estimation of glucose 6-phosphate residues on starch and the presence of other bound phosphate(s). *Starch/Stärke* 45, 417-420

46. Tabata, S. and Hizukuri, S. (1971) Studies on starch phosphate. Part 2. Isolation of glucose 3-phosphate and maltose phosphate by acid hydrolysis of potato starch. *Starch/Stärke* 23, 267-272

47. Nielsen, T.H. *et al.* (1994) Starch phosphorylation in potato tubers proceeds concurrently with *de novo* biosynthesis of starch. *Plant Physiol.* 105, 111-117

48. Ritte, G. *et al.* (2004) Phosphorylation of transitory starch is increased during degradation. *Plant Physiol.* 135, 2068-2077

49. Hejazi, M. *et al.* (2010) The laforin-like dual-specificity phosphatase SEX4 from Arabidopsis hydrolyzes both C6- and C3-phosphate esters introduced by starch-related dikinases and thereby affects phase transition of α-glucans. *Plant Physiol.* 152, 711-722

50. Hejazi, M. *et al.* (2012) Starch phosphorylation and dephosphorylation. In: *Starch: Origins, Structure and Metabolism*. *(Essential Reviews in Experimental Biology)* (Vol. 5) (Tetlow, I, ed.), pp. 279–310, SEB

51. Ral, J.-P. *et al.* (2012) Down-regulation of glucan, water-dikinase activity in wheat endosperm increases vegetative biomass and yield. *Plant Biotechn. J.* 10, 871-882

52. Schönknecht, G. *et al.* (2013) Gene transfer from bacteria and archea facilitated evolution of an extremophilic eukaryote. *Science* 239, 1207-1210

53. Hejazi, M. *et al.* (2012) The plastidial glucan, water dikinase (GWD) catalyses multiple phosphotransfer reactions. *FEBS J.* 279, 1953-1966

54. Glaring, M.A. *et al.* (2011) Starch-binding domains in the CBM45 family – low affinity domains from glucan, water dikinase and α-amylase involved in plastidial starch metabolism. *FEBS J.* 278, 1175-1185

55. Kötting, O. *et al.* (2005) Identification of a novel enzyme required for starch metabolism in *Arabidopsis* leaves. The phosphoglucan, water dikinase. *Plant Physiol.* 137, 242-252

56. Baunsgaard, L. *et al.* (2005) A novel isoform of glucan, water dikinase phosphorylates prephosphorylated α-glucans and is involved in starch degradation in *Arabidopsis*. *Plant J.* 41, 695-605

57. Hejazi et al. (2009) The two plastidial starch-related dikinases sequentially phosphorylate glucosyl residues at the surface of both the A- and the B-allomorph of crystallized maltodextrins but the mode of action differs. *Plant Physiol.* 150, 962-976

58. Tagliabracci, V.S. *et al.* (2011) Phosphate incorporation during glycogen synthesis and Lafora Disease. *Cell Metabolism* 13, 274-282

59. Roach, P.J. *et al.* (2012) Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* 441, 763-787

60. Gentry, M.S. *et al.* (2013) Laforin, a protein with many faces: glucan phosphatase, adapter protein, *et alii*. *FEBS J.* 280, 526-537

61. Nitschke, F. *et al.* (2013) Hyperphosphorylation of glucosyl C6 carbons and altered structure of glycogen in the neurodegenerative epilepsy Lafora disease. *Cell Metabolism* 17, 756-767

62. Minassian, B.A. *et al.* (1998) Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. *Nat. Genet.* 20, 171-174

63. Sakai, M. *et al.* (1970) Studies in myoclonus epilepsy (Lafora body form). II. Polyglucosans in the systemic deposits of myoclonus epilepsy and in corpora amylacea. *Neurology* 20, 160-176

64. Tagliabracci, V.S. *et al*. (2008) Abnormal metabolism of glycogen phosphate as a cause for Lafora disease. *J. Biol. Chem.* 283, 33816-33825

65. Chan, E.M. *et al.* (2003) Mutations in NHLRC1 cause progressive myoclonus epilepsy. *Nat. Genet.* 35, 125-127

66. Tiberia, E. *et al.* (2012) Increased laforin and laforin binding to glycogen underlie Lafora body formation in malin-deficient Lafora disease. *J. Biol. Chem.* 287, 25650-25659

67. Pulido, R. and Hooft van Huijsduijnen, R. (2008) Protein tyrosine phosphatises: dual-specificity phosphatises in health and disease. *FEBS J.* 275, 848-866

68. Fordham-Skelton, A.P. *et al.* (2002) A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. *Plant J.* 29, 705-715

69. Kerk, D. *et al.* (2006) A chloroplast-localized dual-specificity protein phosphatase in Arabidopsis contains a phylogenetically dispersed and ancient carbohydrate-binding domain, which binds the polysaccharide starch. *Plant J.* 46, 400-413

70. Kötting, O. *et al*. (2009) STARCH-EXCESS4 is a laforin-like phosphoglucan phosphatase required for starch degradation in *Arabidopsis thaliana*. *Plant Cell* 21, 334-346

71. Worby, C.A. *et al.* (2006) Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates. *J. Biol. Chem.* 281, 30412-30418

72. Niittylä, T. *et al.* (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. *J. Biol. Chem.* 281, 11815-11818

73. Gentry, M.S. *et al.* (2007) The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. *J. Cell Biol.* 178, 477-488

74. Gentry, M.S. *et al.* (2009) Lafora disease: Insights into neurodegeneration from plant metabolism. *Trends Biochem. Sci.* 34, 628-639

75. Comparat-Moss *et al.* (2010) A putative phosphatase, LSF1, is required for normal starch turnover in Arabidopsis leaves. *Plant Physiol.* 152, 685-697

76. Santelia et *al*. (2011) the phosphoglucan phosphatase like SEXFour2 dephosphorylates starch at the C3-position in *Arabidopsis*. *Plant Cell* 23, 4096-4111

77. Meekins, D.A. *et al.* (2013) Structure of the Arabidopsis glucan phosphatase LIKE SEX FOUR2 reveals a unique mechanism for starch dephosphorylation. *Plant Cell* 25, 2302-2314

78. Colleoni, C. and Suzuki, E. (2012) Storage polysaccharide metabolism in cyanobacteria. In: *Starch: Origins, Structure and Metabolism*. *(Essential Reviews in Experimental Biology)* (Vol. 5) (Tetlow, I, ed.), pp. 217–253, SEB

MATERIALS AND METHODS

1. Media and culture conditions

Cyanobacterium sp. CLg1 was grown in Artificial Sea Water medium in the absence (AS0 medium) or in the presence of nitrogen source provided by 0.88 mM of sodium nitrate (ASNIII medium) (Rippka *et al.,* 1979). The axenic strain was grown at 22°C and subjected to light/dark cycle (12h/12h) under 0.12 µmol of photon.m^{-2}.s^{-1}. Wild type and mutant strains were maintained on solid ASNIII medium (0.8% noble agar) and transferred onto new plates every three months.

2. Mutagenesis campaign on the *Cyanobacterium sp.* CLg1 and iodine screening

The mutagenesis campaign was performed on exponential growth cultures of *Cyanobacterium* CLg1 sp. plated on solid ASNIII medium (0.9 % noble agar). Cells (> 10^5 cells/plate) were then subjected to Ultra-Violet irradiation at 0.5; 3; 6 cm from the source (Trans-illuminator TS-15 with a peak intensity of 7 mW.cm^{-2} at 254 nm, Ultra-violet Products, Inc., San Gabriel, Calif.) for various exposure times (15; 30; 60; 120 s) and immediately incubated in darkness during 24 hours to avoid the activation of photo damage reparation system (Golden, 1988). After one month of growth in day-night cycle (12h-12h), all surviving colonies belonging to one mutagenesis condition (*e.g*. 30 s, 1 cm) were gently scratched by adding ASNIII medium over the agar slants and transferred in a 1L flask containing 300 mL of ASNIII. Mutagenized cells were diluted and plated on solid agar ASNIII medium. After one month, single colonies were transferred to 96-wells plates containing 300 µL of liquid ASNIII medium and incubated for another month. 20.000 cell patches were performed in duplicate by loading 40 µL of cell suspension on two solid ASNIII media. After one month of growth, mutants impaired in starch metabolism were screened by spraying cell patches with iodine crystal vapours. In contrast to dark-blue staining of wild type cell patches, cell mutants harboring a distinct iodine staining were picked on sister Petri dishes and transferred to liquid ASNIII medium. The segregation of genome copies has been checked for each individual mutant by an additional round of iodine screening on 48 colonies. Over 20000 cell patches screened, 88 mutants impaired in the starch metabolism pathway were isolated. Among them, 14 mutants harboring a yellow iodine phenotype replaced starch granule synthesis by that of glycogen.

3. Transmission electron microscopy observation

Wild-type and mutant strains were cultivated in 50 mL liquid ASO medium and harvested by centrifugation (5 min at 4000 g at 4°C) after two weeks of growth. The pellets were suspended and incubated during 1 h in fixing buffer containing 2.5% glutaraldehyde and

0.1 M cacodylate buffer at pH 7.2. The fixed cells were centrifuged at 16000 g during 15 min and the supernatants were discarded. The pellets were washed 3 times with 0.1 M cacodylate buffer at pH 7.2. The samples were then dehydrated by incubating 20 min in increasing percentages of ethanol (25%, 50%, 70%, 95%, and 100% three times). The impregnation step was conducted in EPON resin / ethanol mixtures at different ratios (1:2, 1:1, 2:1) for 60 min, in pure EPON for 2 x 30 min, and finally in pure EPON overnight at 4°C. Pellets were then transferred in a capsule containing fresh EPON and incubated during 48 h at 60°C for polymerization. Ultrathin sections were cut using a Leica UC6 ultramicrotome, then treated with 1% periodic acid during 30 min and washed 6 times in water. The sections were plunged in 20% acetic acid containing 1% thiosemicarbazide for 1 h and washing series were performed as follows: 20%, 10%, 5%, and 2% acetic acid during 5 min each. The sections were washed 6 times in water, and stained with 1% silver proteinate during 30 min in the dark, and finally washed 6 times in water.

As described in the following, water-soluble polysaccharides (WSP) were purified from wild-type and mutant strains subjected to nitrogen starvation. Rabbit liver glycogen was also used as a reference. Droplets of diluted WSP suspension were deposited onto glow-discharged carbon-coated copper grids. The liquid in excess was blotted with filter paper and a droplet of 2% uranyl acetate was added prior to drying. The stain in excess was blotted and the specimens allowed to dry.

Both sections and negatively stained preparations were observed using a Philips CM200 microscope operating at 80 kV. The images were recorded on Kodak SO163 films.

4. Quantification of water soluble and insoluble (starch-like) materials produced in the wild type and mutant strains.

Wild type and mutant strains were grown during 12 days in liquid ASO medium and harvested at the middle of the day by centrifugation at 3600 g during 15 minutes at 4°C. After washing the cells three times with extraction buffer (50 mM HEPES pH 8, Triton, 5 mM DTT, 1 mM EDTA, 0.025% triton), the cell suspension (10 mL) was disrupted through a French Press. Water Soluble Polysaccharide (WSP) and insoluble polysaccharide (starch-like granules) were separated by spinning the lysate at 16000g for 15 minutes at 4°C. The supernatant WSP and pelleted starch-like granules were quantified by the amyloglucosidase assay (R Biopharm Starch/amidon (R-Biopharm). The starch pellet was washed three times with cold sterile water and then solubilized in 100% DMSO at 90°C during 10 min before amyloglucosidase digestion. Results are expressed in mg of polysaccharide/milligram of total

protein. The total protein concentration is determined in the supernatant using the Bradford method (Bio-Rad).

5. Structural analysis of soluble polysaccharide

Water Soluble polysaccharide were further purified by using strong anion exchange chromatography (Roth DOWEX 50:8) followed by size exclusion chromatography pre-equilibrated in 10% DMSO (Toyopearl TSK HW 50) as described in (Colleoni *et al.,* 1999). Polysaccharide was quantified in each fraction (1mL) by the phenol-sulfuric acid method (Fox and Robyt, 1991). Polysaccharide fractions excluded of the column were pooled and then subjected to another size exclusion chromatography through a TSK HW 55S (Toyopearl) column. Polysaccharide detected in the exclusion volume was totally debranched by *Pseudomonas* sp. isoamylase of Megazyme, in Sodium Acetate 55mM pH 3.5. The chain length distribution of glucan chains was then analyzed by HPAEC-PAD as described in (Colleoni *et al.,* 1999).

6. Structural analysis of insoluble polysaccharide

Insoluble polysaccharide extracted and solubilized for quantitative analysis were precipitated in 70% of ethanol, after centrifugation at 6000g during 10 minutes the supernatant was removed and the pellet was dried at room temperature, then the pellet was resuspended in 10mM NaOH, and subjected to CL-2B. Glucans eluted in each fractions (of 300µl) were detected by their interaction with iodine, 80 µL of fraction was incubated with 20 µL of iodine solution (1% KI and 0.1% I_2) and analyzed by spectrophotometer in order to determine the maximum OD (optical density) and the wavelength at maximum OD (λmax). The amylopectin found in the exclusion volume was then collected and totally debranched as described for soluble polysaccharide and subjected to HPAEC-PAD.

7. Purification and identification of cation dependent debranching enzyme

After 12 days of growth in ASNIII medium, cells were harvested at the middle of the day by centrifugation (3000 g at 4°C during 15 min.) and washed three times with a cold Tris-acetate buffer (25 mM Tris-acetate, pH 7.5, 10 mM DTT). Cells were disrupted by sonication or by a French press at 1250 psi. The lysate was spinned at 16000g during 15 minutes at 4°C. The supernatant (20 mL) was loaded on preparative anion-exchange chromatography (AEC) column (HitrapQ sepharose FF, 5ml column volume, GE Healthcare) pre-equilibrated in buffer A (25 mM Tris-acetate, pH 7.5, 10 mM dithiothreitol). The proteins were eluted at 4

$ml.min^{-1}$ using a stepwise gradient of 40 %, 75 % and 100% of 1 M NaCl. Cation dependent debranching enzyme activity was monitored for each fraction (2mL) by zymogram analysis. The proteins were separated by non-denaturing polyacrylamide gel electrophoresis (PAGE) containing 0.15% of β-limit dextrin (Megazyme). After electrophoresis, gels were incubated overnight at room temperature in buffer A with 10 mM of $MgCl_2$. The cation dependent activity was then visualized as white or blue activity bands in the presence of BLD or starch polysaccharides, respectively, in the gel after iodine staining. Fractions containing cation dependent activity were pooled and ammonium sulfate was added in order to reach a final concentration of 1 M. After filtration, the pooled-fraction was loaded on a hydrophobic exchange chromatography column (Hitrap Butyl sepharose FF, 5 mL column volume, GE Healthcare) pre-equilibrated in buffer A with 1 M $(NH_4)_2SO_4$. The proteins were eluted at 4 $ml.min^{-1}$ with a linear gradient of 25 mL until 100% and the elution at 100% continued for 30 mL with buffer A. Debranching enzyme activity of each fraction (2 mL) were monitored by zymogram analysis as described above. Fractions containing debranching enzyme were further purified by AEC column (MonoQ 5/50 GL, 1mL column volume, GE Healthcare) pre-equilibrated in buffer A (25 mM Tris-acetate, pH 7.5, 10 mM dithiothreitol). The proteins were eluted at 1 $ml \cdot min^{-1}$ with a gradient of buffer A containing 1 M NaCl : 0 % to 40 % in 5 mL, 40 % to 60 % in 20mL, 60 % to 80 % in 10 mL, 80 % to 100 % in 5 mL. The enzyme activities of each fraction (1mL) were monitored by zymogram analysis. Fractions containing the activity were further concentrated with Amicon Ultra-0.5ml Centrifugal filters for protein purification and concentration (Merck Millipore). Concentrated activity was kept at -80°C in 20% glycerol.

8. Biochemical characterization of cation dependent activity towards different polysaccharides.

The purified cation dependent activity (0.526 μmol of reducing $end.min^{-1}.mL^{-1}$) was incubated in 120 mM Tris, 30 mM imidazole, 30 mM acetic acid pH 7.5 in presence of 0.5 % of polysaccharides (p/v) (amylopectin, glycogen, beta-limit dextrin, phosphorylase limit dextrin) and 10 mM $MgCl_2$. The increase of reducing ends was measured by adding dinitrosalicylic acid (v/v). After 10 min at 99°C, the optical density is measured at 540 nm and compared to a standard curve (0 to 100 μg) measured with 1 $mg.mL^{-1}$ of glucose. Results are expressed as μg of reducing-end equivalent glucose. Glucan chains released during the incubation were analyzed by HPAEC-PAD as described below. The cation dependent activity

was compared to isoamylase of *Pseudomonas* sp. diluted to 1/100 incubated with 0.5 % of polysaccharides and sodium acetate 55 mM at pH 3.5.

9. Identification of the 80 kD polypeptide by nano LC-MS-MS

The cation dependent activity was further purified by an additional size exclusion chromatography pre-equilibrated with 10 mM Tris/acetate buffer pH 7.5, 10 mM dithiotreitol, 150 mM NaCl (Ge-Healthcare superdex 200; D = 0.8 cM; H = 30 cM). 30 µL of each fraction (300 µL) was analyzed using zymogram procedures. Fractions containing the cation dependent activity were then precipitated with 20% TCA (60 µL TCA 20% for 1 mL of sample with 25 µL of sodium deoxycholate 2%), and kept on ice during 10-15 min. After centrifugation 5 min at 16100g, pellets were resuspended in 10 µL of buffer A (0.1M Na_2CO_3; 0.1M DTT). The samples were boiled at 95°C during 5 min in presence of 5 µL of buffer B (5 % SDS, 30 % saccharose, 0,1 % Blue of bromophenol). Proteins were separated on polyacrylamide gels of 7,5 % with 0,1 % SDS. SDS PAGE gels were then stained during one night using Roti Blue (Roth) and washed with 25% methanol. The polypeptide was cut off and placed in 60% acetonitrile in bicarbonate ammonium before nano LC MS/MS analysis as described in (Gurcel *et al.*, 2008).

10. Gene cloning and sequencing

Debranching enzyme genes (glgX2, glgX1), branching enzyme genes (be3) and starch/glycogen synthase genes (gbss) were amplified from genomic DNA of wild type and mutant strains. Primers were designed in order to hybridize outside the ORF:
glgX2F (Forward) GATCATTGTCGACTACCTCAAGACTCA glgX2R (Reverse); TCTTAACCATTTATGTTGGTTAAAGTTTCA , glgX1F GGATAAAGTTTTAATGATAGGTTATTACAA glgX1R ACCACGAACCCCGAACTTAT; gbssF TCCTCATGAATTGGTGACATAGTATGTT gbssR CAGATACAGGTGAAAATCGTAACGC; be3F AGTGAATAGCCAAAAATCAACGAT , be3R TGACCATCCCATTTGGCTCCTA. The PCR experiments were conducted as follows: 95°C for 5 min; 30 cycles of denaturation at 98°C for 30 s, annealing 30 s at 59.6°C for glgX2, GBSS and BE3 and at 59.1°C for glgX1, and extension 2 min30s at 72°C, and a final elongation step at 72°C for 5 min. The PCR products were cloned into pCR-BluntII-TOPO vector (Invitrogen) and transferred into the chemical competent *E.coli* TOP10 Mach1™-T^R, and plated on LB agar with Kanamycin and X gal. Purified plasmids were sequenced by GATC Biotech Company according to Sanger methods. Each gene was sequenced on both

strands using additional primers when required. The presence of a mutation was identified by alignment with the wild type gene using the BLASTn program.

Gene	Primer	Sequence	T°m
Amylopullulanase	apu13 Cloning For	GCATCCATATTAAAATGATCGCTAAGG	56.6
	apu13 Cloning Rev	TTCGTCCTTCGTCCCTCGTCCTT	
	apu13 Sequencing 1	GGTAATCGAGCTTTACCGAAAT	
	apu13 Sequencing 2	TCAAGATCTTCTCCTGCTACCGTTA	
	apu13 Sequencing 3	CAGCTAACCCAGTTAAATTTACTTGC	
	apu13 Sequencing 4	GATTGAATTTCGGTAAAGCTCGAT	
	Apu57 Cloning For	GGAGATCTTCTGCCCCTCTTTCA	58.6
	Apu57 Cloning Rev	AAAATAGCCAATAAACAGAAGCGA	
	Apu57 Sequencing 1	ACGAACTGCGAACCCAGAACTCC	
	Apu57 Sequencing 2	ACCCAATCCCCTGCCTTCGCAT	
	Apu57 Sequencing 3	GGCATATCAAAGGTACTCGTCATGG	
	Apu57 Sequencing 4	AGGTAAAGACAGGGGCTGTCAA	
Indirect debranching enzyme	idbe Cloning For	TCAATTATAAGTAGGGGTAGGAGA	
	Idbe Cloning Rev	ATGCTGAAAAAATCGTCCAAA	
	idbe Sequencing 1	CACAACTCTGATAACAGCCTCAA	
	idbe Sequencing 2	GTACACCTGACTCACCACCGTAA	
	idbe Sequencing 3	GATTGGCATATCAAAGGTACTCG	
	idbe Sequencing 4	ACCGCCTAGATAATGAGATTGGAT	
Branching enzyme	glgB3 Cloning Rev	TGACCATCCCATTTGGCTCCTA	61°C
	glgB3 Cloning For	AGTGAATAGCCAAAAATCAACGAT	
	glgB3 Sequencing 1	GTGGTTTAGGCAACTACGACGGTACA	
	glgB3 Sequencing 2	CGTTACGTATCATGGAACTAATTT	
	glgB3 Sequencing 3	CCAGGTGCGCCGATGATTTT	
	glgB3 Sequencing 4	TCGCCCCTGAATGATGATATAAAT	
	glgB2 Cloning For	TGCGAAAAGATGTCTTGCTATGCT	60°C
	glgB2 Cloning Rev	TCCACTCAGTCAAGTGTGAATCAA	
	glgB2 Sequencing 1	TTTTGACGGCTCATGGGGTTA	
	glgB2 Sequencing 2	ATTGTGATAGTTCGGGCGCTCAT	
	glgB2 Sequencing 3	AGCGCCCGAACTATCACAAT	
	glgB2 Sequencing 4	ACCATAGCGGGAAGTTGGTG	
	glgB1 Cloning For	TGAAACAGTCAGACAAGTTTTCCGT	
	glgB1 Cloning Rev	TCCCGAACTGAGGTTAAGTATTGA	
	glgB1 Sequencing 1	CACCTCGGTTCTTGGCTACA	
	glgB1 Sequencing 2	TGCATCCAGCCCATATTCCA	
	glgB1 Sequencing 3	TCGCAGAAGAATCAACCGCTTGGT	
	glgB1 Sequencing 4	TGTAGCCAAGAACCGAGGTGT	
	glgB57 Cloning For	CGCCATTAGACTCTGCTAGG	
	glgB57 Cloning Rev	TGGAGTTGTGGTTTCTTCTG TCA	
	glgB57 Sequencing 1	TGCGGTTATTACGGTGGCTT	
	glgB57 Sequencing 2	TAGCTAAACCCGCAGCATCG	
	glgB57 Sequencing 3	AACCAAATCCCTCTGGGAGCAA	
	glgB57 Sequencing 4	GAGCTTGCTTGTCGGCTAGT	

Debranching enzyme	glgX1 Cloning For	ATGTTAATGGGAGATGAATCTATGA	
	glgX1 Cloning Rev	TAATTAGTGGTTTTTAGTACTACTAACG	
	glgX1 Sequencing 1	AGCAGTGGTAGTCGACGAATCT	
	glgX1 Sequencing 2	ACGTCGTTGCTGATTGCGTA	
	glgX1 Sequencing 3	GTAACTTCTGGCGTGGCGAAA	
	glgX1 Sequencing 4	AAGCCGGCAAATGTTCCTCGTA	
	glgX2 Cloning For	ATGAACCATAAAACGTTACCTG	
	glgX2 Cloning Rev	CTATTTTGCCATTAATAAAATGCAAC	
	glgX2 Sequencing 1	TCGGATTCCTCTTGCTTCAACCGT	
	glgX2 Sequencing 2	TCATCGGTGTAGCTTGAGCGAGT	
	glgX2 Sequencing 3	CCACGATGGGGTGCAAAGAAACT	
	glgX2 Sequencing 4	ATGAAGCTAATGGTGAAGAGAACCGA	
Starch synthase	GBSS Cloning For	TCCTCATGAATTGGTGACATAGTATGTT	
	GBSS Cloning Rev	CAGATACAGGTGAAAATCGTAACGC	
	GBSS Sequencing 1	AAAACGCCCCTGGAAAGCAATA	
	GBSS Sequencing 2	GCAATAGTCAGCACCAGCCGTGATA	
	GBSS Sequencing 3	TTTATCACGGCTGGTGCTGACT	
	GBSS Sequencing 4	ATGTTAAACATTTGCTTTGTCTCTACGGAA	
	glgA1 Cloning For	AATTGTTAACAAAGTGGGGAGAGAAA	60°C
	glgA1 Cloning Rev	GATAACGATCGTGATAACTTTTAAGGTA	
	glgA1 Sequencing 1	GCTCTGCAACATATGACTTGGTGTCCTGA	
	glgA1 Sequencing 2	CCATAATAACGAGCCGCAAATTCTGCTA	
	glgA2 Cloning For	AAGTGAATCAACTGAATGAAAACGA	59.8°C
	glgA2 Cloning Rev	CGGGTATCTGTAATCTAGAACTCCTCA	
	glgA2 Sequencing 1	ATGATTGGCAAACTGGATTGATCCCTGTA	
	glgA2 Sequencing 2	ACCACCAACACGAACTACAGGTACT	
	glgA2 Sequencing 3	TTAAACGGAATTGATTACAATACTTGGGA	
	glgA2 Sequencing 4	GATTGGAATAGACAATACCACCTTTCA	
Phosphorylase	Phosphorylase Cloning For	GCTTACTAAGACGAAATTGATTATGAGCGAT	62°C
	Phosphorylase Cloning Rev	TTCACCGTAACGCTATCAGCCGTAAT	
	Phosphorylase Sequencing 1	TTAGGTAACGGTGGTTTGGGAA	
	Phosphorylase Sequencing 2	CCAAAGATAAAGGTGCGGGGATA	
	Phosphorylase Sequencing 3	CTTCTTCTTAGAAGATATTCGTTCATGGTA	
	Phosphorylase Sequencing 4	GAGTTTGTTGTGTAATCGACCACGCTT	
	Phosphorylase Sequencing 5	AACTCTTGACGGTGCCAATATTGA	
	Phosphorylase Sequencing 6	CTAAACCTGGATCTGGCTCTTCTT	
MalQ	MalQ Cloning For	CGCAGGTCCTGTGATTGTCT	62°C
	MalQ Cloning Rev	CCATCTACCGGTAACGCTGTAA	
	MalQ Sequencing 1	CTTCAGCGACAATCGGCAACT	
	MalQ Sequencing 2	CCACGATCGGTTGGTTTGATA	
	MalQ Sequencing 3	TGGTTGGATGACTACGCTCTGT	
	MalQ Sequencing 4	TGCAAATACCCCTTTGATGCT	

11. Phylogenetic tree of GlgX

Homologs of GlgX sequence were identified in GenBank, or other sources using BLASTp and aligned with MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/). The alignment was

manually refined using SeAl (http://tree.bio.ed.ac.uk/software/seal/) and blocks of missing data in some taxa or regions of low identity were manually removed. This reduced alignment was analyzed under maximum likelihood (ML). The best-fitting amino acid substitution model was selected according to the Akaike informational criterion with ProtTest using the default values (Abascal *et al.,* 2005). The LG (Le and Gascuel 2008) model with heterogeneous gamma rate distribution across sites (+Γ) was selected by ProtTest for this protein data set. The LG model parameter values were used under RAxML v.7.2.8 (Stamatakis 2006) for the ML tree searches. The stability of monophyletic groups was assessed using RAxML with 1000 bootstrap replicates.

12. Phylogenetic tree of other proteins

The sequences of homologous proteins or part of proteins were identified in NCBI, or other sources: Dragonblast or DB Lab database (red algae and Glaucophytes), using BLASTp and aligned with MUSCLE (Edgar 2004). The alignment was manually refined using Jalview (Waterhouse *et al.,* 2009) and blocks of missing data in some taxa or regions with low identity were manually removed. This reduced alignment was analyzed under maximum likelihood (ML). The best-fitting amino acid substitution model was selected according to the corrected Akaike informational criterion using the default values of ProtTest program (Abascal *et al.,* 2005). The latter defined whether both gamma rate distribution across sites and invariable site option have to be selected for each data set. The selected model and the parameter values were used under RAxML v.7.2.8 (Stamatakis 2006) for the ML tree searches. The stability of monophyletic groups was assessed using RAxML with 100 bootstrap replicates.

13. Loop analysis

Sequences of GlgX and isoamylase are aligned with Muscle (Edgar 2004), and the putative characteristic loop between the region II and III is shown.

14. Analysis of glucan released by cation dependent enzyme

The purified cation dependent enzyme (0.526 µmol of reducing end.min^{-1}.mL^{-1}), as described above, was incubated in 120 mM Tris, 30 mM imidazole, 30 mM acetic acid pH 7.5 in presence of 0.5 % of polysaccharides (p/v) (amylopectin or glycogen) and 10 mM $MgCl_2$ 30°C. Time-course experiment was carried out at 10min to overnight and liberated oligosaccharides were analyzed by HPAEC-PAD. Results are expressed as nanocoulomb (nC) representing the quantity of reducing end of each chain length liberated at one moment of the

kinetic. The cation dependent activity was compared to isoamylase of *Pseudomonas* sp. diluted to 1/100 incubated with 0.5 % of polysaccharides and sodium acetate 55 mM at pH 3.5.

15. Gene set of starch metabolism in *Oxytricha trifallax*

Proteome of *Oxytricha triffalax* is accessible in the NCBI's databank (Swart *et al.,* 2013). Proteins were selected with sequences of biochemically characterized enzymes using Blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Each enzyme with a e-value upper than (10^{-5}) was submitted to Interproscan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) and enzymes known to be active onto glycogen and amylopectin were retained.

CONVERGENT EVOLUTION OF POLYSACCHARIDE DEBRANCHING DEFINES A COMMON MECHANISM FOR STARCH ACCUMULATION IN CYANOBACTERIA AND PLANTS

1. Introduction

Bacteria, Archaea and Eukaryotes often store glucose in the form of glycogen. This storage material consist of small hydrosoluble particles composed of α-1,4 linked and α-1,6 branched glucan chains (Shearer and Graham, 2002). Because current structure models envision that each α-1,4 linked chain supports on average two novel branched chains, mathematical modeling predicts that chain density will increase with particle size up to a maximal possible diameter of 42 nm is reached. Such glycogen particles contain up to 55000 glucose residues with over 36% resting in the outer particle chains (Melendez *et al.,* 1999). These are thus readily accessible to cell metabolism without the need of polysaccharide debranching. Hence in glycogen, glucose residues remain rapidly available to cellular enzymes as if they were in the soluble phase but remain much less active osmotically, corresponding to the same amount of free glucose. Archaeplastida (also called Kingdom Plantae) consisting of the Chloroplastida (green algae and all land plants) the Rhodophyceae (red algae) and the Glaucophyta (glaucophytes) store starch granules of unlimited size and no glycogen (Ball *et al.,* 2011). Starch always contains a glycogen-like polymer named amylopectin occasionally mixed with a very moderately branched amylose polysaccharide. Amylopectin aggregates into semi-crystalline granules of unlimited diameter. This organization results from a concentration, in certain regions, of the distribution of branches allowing the formation of double helical structures that align and crystallize into two different allomorphs (the so-called A and B) or a mixture of both (Buléon *et al.,* 1998).

This aggregation directly impacts the properties of most food sources including their digestibility and is therefore of paramount importance in human health and the history of civilizations which all depend on starch storing crops. It also impacts the industrial functional properties in all non-food uses of the polymer.

Starch is osmotically inert allowing for the accumulation of very large quantities of glucose (60 to 90% of the dry weight) in the storage organs of plants. Until recently the distribution of starch seemed restricted to photosynthetic eukaryotes including several secondary endosymbiosis lineages derived from Archaeplastida such as the cryptophytes (Deschamps *et al.,* 2006), the dinoflagellates (Dauvillée *et al.,* 2010) and some apicomplexa parasites (Coppin *et al.,* 2005). However recent studies have revealed the existence of starch-like structures in unicellular diazotrophic cyanobacteria belonging to Chroococcale order (Nakamura *et al.,* 2005; Deschamps *et al.,* 2008; Suzuki *et al.,* 2013). The presence of "anomalous glycogen" particles had already been identified previously in this clade while it is

only very recently that this material was recognized as starch-like and the term "semi-amylopectin" was coined to describe the major amylopectin-like fraction within these granules, intermediate between glycogen and the true-starch of Chloroplastida (Schneegurt *et al.,* 1994; Suzuki *et al.,* 2013). 4 out of the 6 reported starch accumulating cyanobacteria strains accumulate only this polysaccharide fraction. However two different strains also synthesize amylose by using an enzyme phylogenetically related to the archaeplastidal Granule Bound Starch Synthase (GBSS), an enzyme known to be selectively responsible for the synthesis of this fraction in plants (Delrue *et al.,* 1992; Deschamps *et al.,* 2008).

Clues as to the nature of the biochemical mechanism distinguishing starch from glycogen synthesis came from the study of glycogen accumulating mutants of Chlamydomonas and cereals that proved defective for the same GH13-type, subfamily 11 (Glycosyl Hydrolase family 13 according to the CAZy classification) of debranching enzyme (James *et al.,* 1995; Mouille *et al.,* 1996; Kubo *et al.,* 1999). In the green algae the substitution of starch by glycogen was complete, thereby hinting that the absence of debranching enzyme prevented amylopectin synthesis altogether. This enzyme thus catalyzed an essential previously unrecognized step distinguishing starch from glycogen synthesis. From these observations several groups proposed that debranching enzyme was selectively responsible for the removing of misplaced branches preventing polysaccharide aggregation in an otherwise hydrosoluble precursor (Ball *et al.,* 1996). This model known as the preamylopectin trimming model was confirmed through mutant analysis in all plant systems examined to date including *Arabidopsis thaliana* (Wattebled *et al.,* 2005). The debranching enzymes missing in the glycogen accumulating plants are known as "isoamylases" and belong to a family of direct debranching enzymes that apart from Archaeplastida are only distributed in bacteria. In bacteria, such enzymes generally named GlgX (by reference to the *E. coli* glycogen metabolism locus coding it) are known to be involved in glycogen catabolism by debranching only those external chains that have been first recessed by the catabolic enzyme glycogen phosphorylase. The *E. coli* enzyme is thus known to be very selective for chains or 3-4 Glc residues left over by glycogen/starch phosphorylase action on external chains and displays very little residual activity on longer chains (Jeanningros *et al.,* 1976; Dauvillée *et al.,* 2005). This restricted enzyme selectivity prevents futile cycles during polysaccharide synthesis since such activities are unable to debranch directly the products of branching enzyme activity which branches a minimum of 6 glucose residues on an acceptor chain. In line with their supposed catabolic function, the *E. coli* glgX defective mutants overproduce glycogen with

short external chains (Dauvillée *et al.*, 2005). Interestingly analogous results have also been obtained in glycogen-accumulating cyanobacteria (Suzuki *et al.*, 2007).

We now report the existence of both glycogen and starch in a recently axenized marine *Cyanobacterium* (strain CLg1) related isolate. We show through the isolation of an allelic series of 14 distinct mutants that alteration, decrease or disappearance of a particular GH13 subfamily 11 GlgX-type of debranching enzyme correlates with the disappearance of the starch fraction and a large increase in the accumulation of glycogen. Remarkably the wild-type enzyme defective in the mutants has evolved an isoamylase-type of activity very similar to that of present in plants. This activity evolved, as was the case for the plant enzyme from pre-existing GlgX-like proteins. However, phylogenetic analysis proves that the phylogenetic origin of this gene is independent from the enzyme recruited by the Archaeplastida. These results point to the independent acquisition of starch metabolism in cyanobacteria and plants and suggest that polysaccharide debranching by an isoamylase-like enzyme may define a mandatory universal mechanism for the synthesis and aggregation of starch polymers from enzymes of a preexisting glycogen metabolism network.

2. Results

2.1. *Cyanobacterium* sp. CLg1 accumulates both glycogen and starch from a gene-rich suite of enzymes of bacterial glycogen metabolism.

Unlike Archaeplastida which only contain starch like material, the wild type axenized strain CLg1 always accumulates two distinct α-1,4-linked and α-1,6 branched polysaccharides fractions in the form of both a major (85 %) insoluble and a minor (15%) yet significant water-soluble fraction (Table 1).

Table 1 : Polysaccharide composition in wild type and class A mutant strains

	Starch granules		WSP[a]		% of each polysaccharide	
Strains	mg /mg of protein	%	mg/mg of protein	%	starch	WSP
WT	2.80±0.36	100	0.50±0.03	100	84.7	15.3
174H3	0.08±0.02	3	1.89±0.57	375	4.3	95.7
107D8	0.04±0.01	1.5	0.87±0.23	173	4.6	95.4
91G1	0.09±0.04	3.3	1.57±0.4	312	5.5	94.5
153H12	0.22±0.20	7.7	1.47±0.38	292	12.8	87.2
123B3	0.15±0.05	5.5	1.38±0.14	273	10.1	89.9
99A7	0.09±0.03	3.3	2.69±1.10	533	3.3	96.7
21'C6	0.07±0.02	2.6	2.04±0.42	403	3.4	96.6
118H9	0.05±0.03	1.7	1.64±0.83	324	2.9	97.1
91D6	0.07±0.03	2.45	1.99±0.35	395	3.3	96.7
175E12	0.12±0.06	4.2	2.19±0.72	435	5	95.0
174B6	0.17±0.1	6	1.64±0.51	326	9.3	90.7
134A4	0.06±0.03	2.1	1.80±0.70	354	3.2	96.8
81G3	0.07±0.02	2.6	2.37±0.1	469	3	97.0
80D5	0.07±0.04	2.5	2.13±0.2	422	3.2	96.8

[a] : Water Soluble Polysaccharide. The measurements are means ± SD of three independent extractions.

We confirm that the structure of the insoluble polysaccharide fraction (Figure 1A; Figure 1B) was composed of starch-like granules containing both semi-crystalline amylopectin and amylose with chain-length distribution consistent with the presence of starch

as previously described (Deschamps *et al.*, 2008). Interestingly, in addition to this major starch-like fraction, a small hydrosoluble fraction representing 15 % of the total glucan amount under nitrogen starvation conditions was also observed.

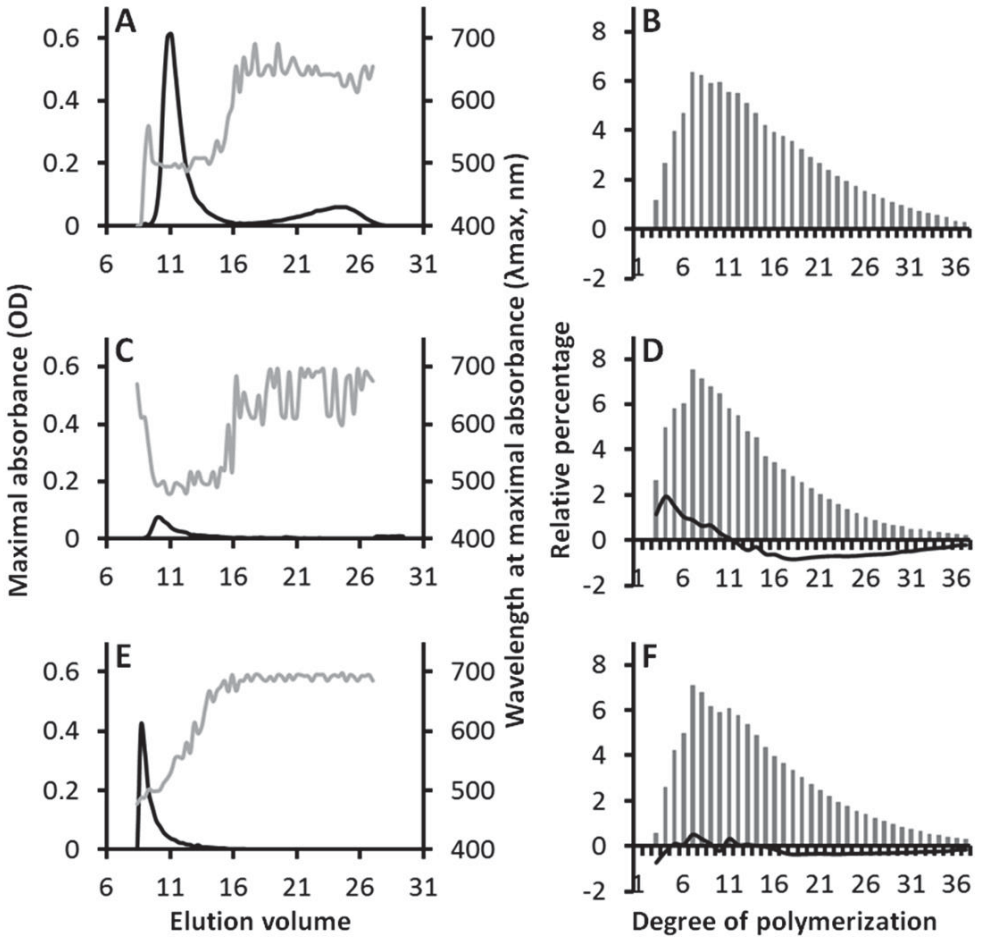


Figure 1. Structural analyses of starch-like granules accumulated by the wild type strain and class A mutants. Insoluble-water polysaccharides purified from wild type (A), a severe mutant 99A7 (C) and an intermediate mutant 153H12 (E) were subjected to size exclusion chromatography analysis (CL2B sepharose). The wavelength (λ_{max}, grey lines) at the maximal absorbance (black lines) of the iodine-polysaccharide complex was determined for each fraction. As previously described, starch-like granules of wild type are composed of semi-amylopectin (fractions 9 to 16) and amylose-like polysaccharides (fractions 17 to 27). Both starch-like granules of class A mutants are composed exclusively of abnormal semi-amylopectin fraction (λ_{max} values ranging between 490-500 nm instead of 507 nm for the wild type semi-amylopectin). Fractions containing semi-amylopectin material were pooled and subjected to chain length distribution (CLD) analysis. After complete digestion with commercial isoamylase, chains of glucose were separated according to their degree of polymerization (DP) by HPAEC-PAD. The relative abundance for each DP (grey bars) was determined for the wild type (B), 99A7 mutant (D) and 153H12 mutant (F) from three independent extractions. Subtractive analyses (% of each DP in the mutant's semi-amylopectin minus % of each DP in the wild type's semi-amylopectin), depicted as black lines in panels D and F, reveal an increase of short chains (DP 3 to 10) and a decrease in long chain content (DP 12 to 35) in the severe mutant 99A7 (in panel D) and no significant difference in the intermediate 153H12 mutant (in panel F).

The chain length distribution (Figure 2B) and the negative-staining TEM observations (Figure 2A), suggest that this material is a highly branched soluble polysaccharide capable to exclude the uranyl acetate molecule in a fashion similar to rabbit liver glycogen particles (Figure 2G).In addition, TEM images of wild-type CLg1 strain ultrathin sections, in which polysaccharides were stained by the PATAg method, confirm the presence not only of the very obvious large-size starch granules but also of the smaller size glycogen-like particles (Figure 3A). To get a better idea of the nature of the biochemical pathways explaining the presence of these two fractions and to better characterize the nature of the CLg1 strain we examined the recently reported sequence of CLg1 strain genome (Suzuki *et al.,* 2013). 16 rRNA and NifH phylogenetical analysis revealed a close relationship to the genus *Cyanobacterium* (Falcon *et al.,* 2004; Suzuki *et al.,* 2013). Strain although containing the genes required for nitrogen fixation remains unable to grow without reduced nitrogen under laboratory conditions.

Surprisingly, the network evidenced is a rather complex set of bacterial and archean distributed enzymes of glycogen metabolism containing among others three glycogen/starch synthases, four glycogen/starch branching enzymes, two GlgX-like DBE sequences, two amylopullulanases, and one iDBE-like α-1,6 glucosidase domain (Colleoni and Suzuki, 2012). With two exceptions, no traces of enzymes of eukaryotic affiliation involved in glycogen or starch metabolism were found in this genome. The important exceptions consist of GBSSI and ADP-glucose pyrophosphorylase, which were probably donated to the Archaeplastida through Endosymbiotic Gene Transfer (EGT) after plastid endosymbiosis. The affiliation of one of the soluble glycogen/starch synthases to the SSIII-IV group of enzymes of Archaeplastida is indirect and this enzyme groups with a number of other proteobacteria and chlamydia-sequences. In this case either the proteobacteria or more probably the chlamydia can be considered as the direct source of the archaeplastidal enzymes (Ball *et al.,* 2013).

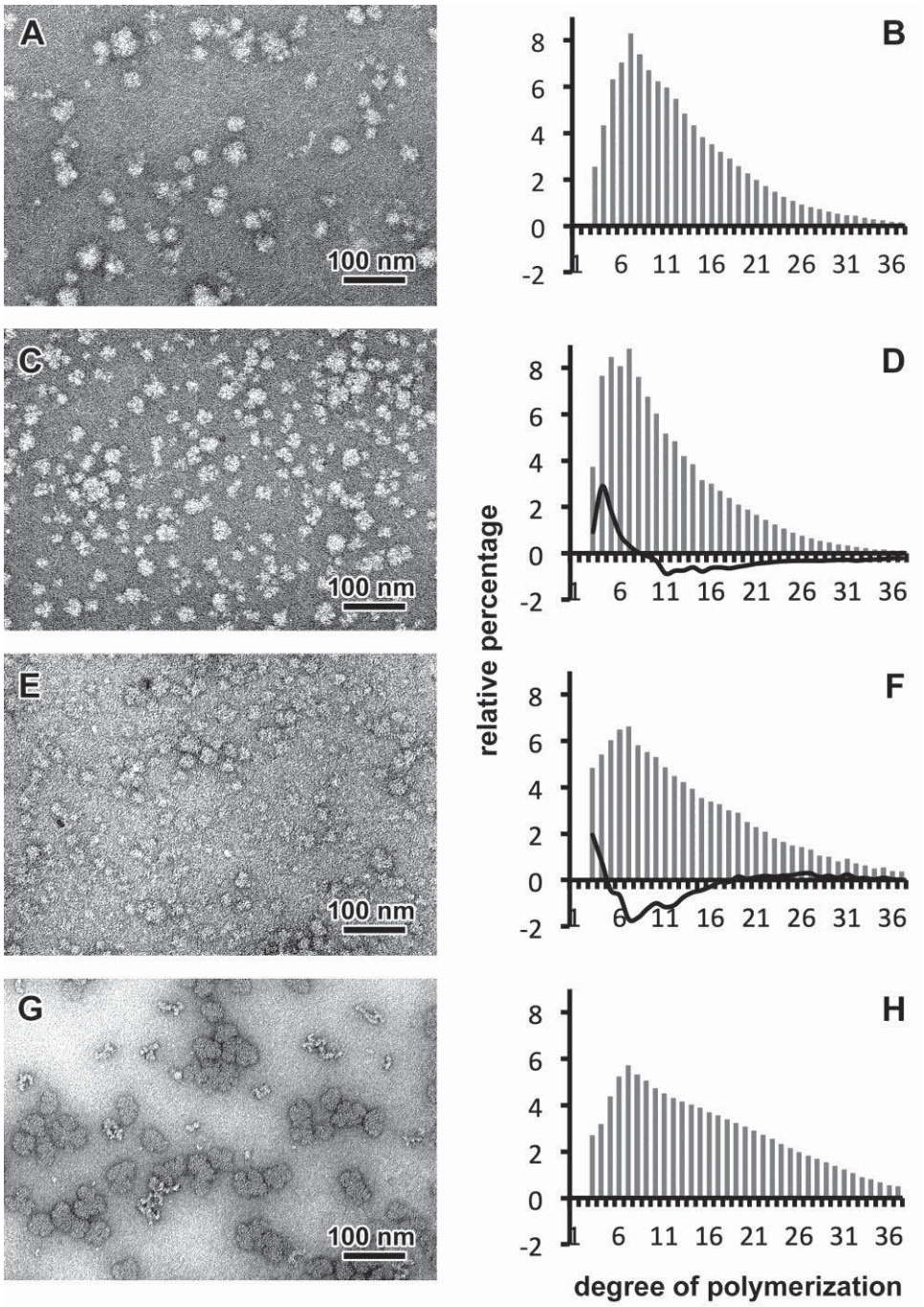
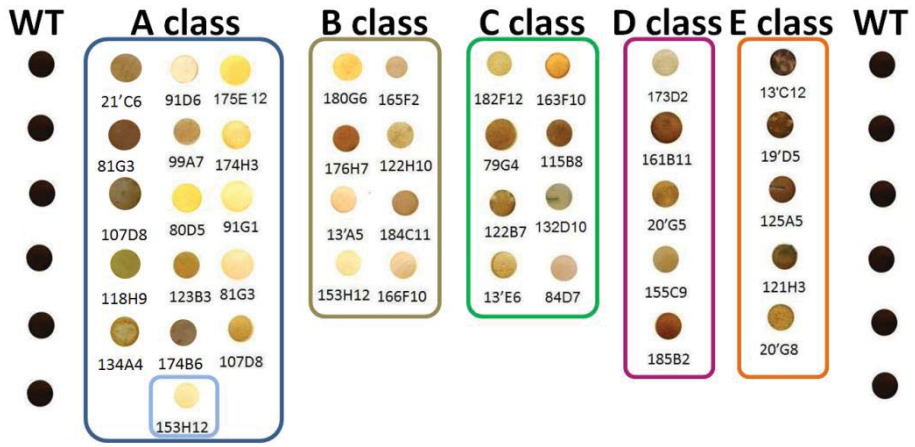


Figure 2. Characterization of water-soluble polysaccharide (WSP) accumulated in the wild type strain and class A mutants. Negative staining following TEM observations suggest that WSP of the wild type (A), the 99A7 mutant (C), the 153H12 mutant (E) are highly branched polysaccharides with a diameter below 50 nm similar to glycogen particles of rabbit liver (G). After purification and complete digestion with a commercial isoamylase, chains of glucose were separated according to their degree of polymerization (DP) by HPAEC-PAD. From three independent extractions, the relative abundance for each DP (grey bars) was determined for the wild type strain (B), the severe mutant 99A7 (D), the intermediate 153H12 mutant (F) and glycogen from rabbit liver (H). Subtractive analyses (% of each DP of mutant minus % of each DP in wild type), depicted as black lines in panels D and F, reveal an increase of short chains (DP 3 to 7) and a defect in long chain content (DP 12 to 35) in the WSP accumulated in the mutant 99A7 (in panel D) and in the intermediate

mutant 153H12 (in panel F), WSP appears to contain less chains with a DP ranging between 4 to 16.

2.2. Selection and characterization of the polysaccharide structure from glycogen accumulating mutants of *Cyanobacterium* sp. CLg1



Supplemental figure 1: Cells patches of the five classes of mutant stained with iodine vapors. Mutants are grouped according to the percentage of residual water-soluble polysaccharides and starch-like granules measured in the mutant strains in comparison to the wild type strain (WT). Class A mutants (starch-like granule < 100% ; WSP >100 %). The mutant strain 153H12 harbors a yellow phenotype despite the presence of 10 % of starch-like granules (bleu square). The intermediate phenotype class B mutants (starch-like granule < 100% ; WSP <100 %); class C (starch-like granule < 100% ; WSP = 100 %); class D mutants (starch-like granule = 100% ; WSP <100 %); Class E mutants (starch-like granule >100% ; WSP <100 %).

Because of the very slow growth rate of the CLg1 strain and the absence of established transgenesis procedures for this organism, we chose a forward genetic approach (UV mutagenesis) to dissect the genetic determinants of starch metabolism in cyanobacteria (for details see material and methods). Five phenotypic classes of mutants were found after three years of screening, segregation and purification (supplemental Figure 1). We benefited from the presence of amylose which yielded a very strong and sensitive iodine stain for the screening of 2.10^4 cell patches. Of relevance to this study was the finding of a class of mutants (class A), which lacked interaction with iodine and which yielded low starch and abundant water-soluble polysaccharides. Figure 3 shows that 14 mutants of this class accumulated abundant small size granules resembling the minor glycogen fraction of the wild-type reference. Table 1 shows that 11 out of 14 mutants of this class witness a 94 to 97% decrease of insoluble granules and a replacement by a significant amount of hydrosoluble polysaccharide. However, three additional mutant strains displayed a phenotype intermediate between standard class A mutants and the wild-type. The latter accumulated up to 9 to 12 %

of the wild-type amount of starch and over-accumulated glycogen (table I). This will be referred to as the incompletely defective class A mutants. They also gave clear mutant iodine stain phenotype of cell patches (Supplemental figure 1). The substitution of starch by glycogen restricted the total amount of stored glucans in all mutants from two to three fold. Purification of the water-soluble fraction followed by enzymatic debranching and analysis of the chain length (CL) distribution of the resulting glucans yielded a CL that mimicked those of glycogen and clearly differed from that of amylopectin-like polymers purified from the wild-type CLg1 reference (Fig. 2). Although similar to the CL distribution of the minor glycogen fraction from the wild-type reference the mutant water-soluble polysaccharide was selectively enriched in very small chains (DP 3 to 6) when compared to the wild type. A very minor amount (2-6% of the wild-type mount) of high molecular weight material could still be purified from the glycogen accumulating mutants. This material resembled the polysaccharide accumulating in the single incompletely defective class A mutant 153H12 (12%). Interestingly the amylopectin –like polysaccharide was equally enriched in very small glucans when compared to the wild type.

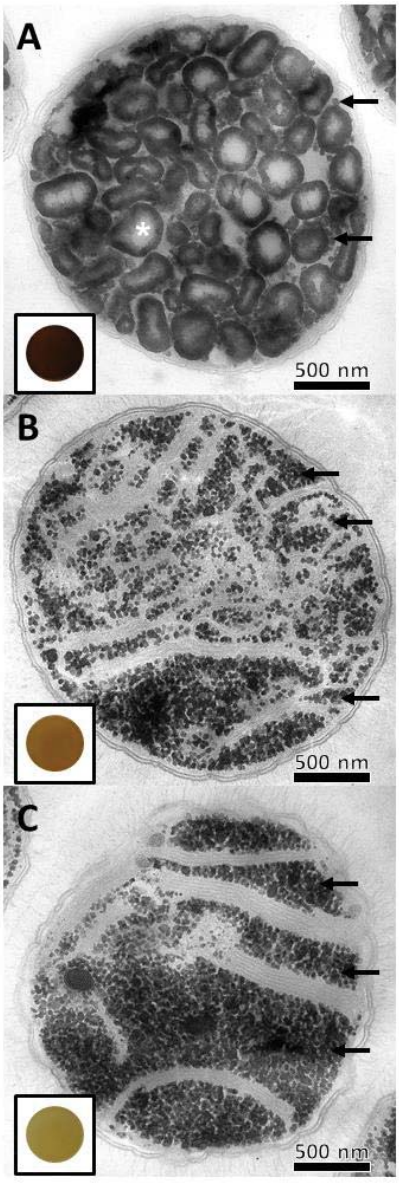
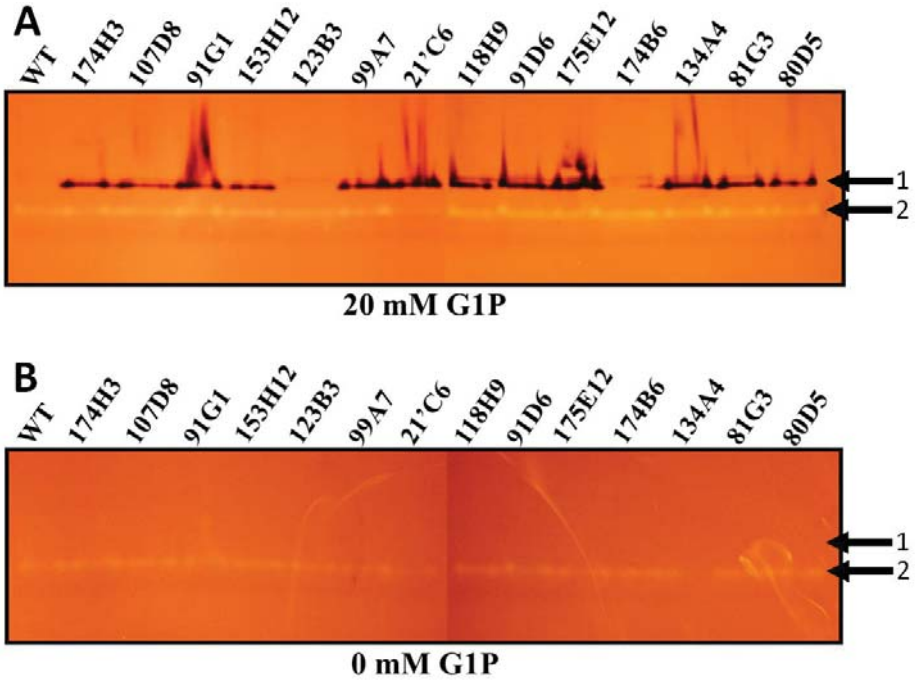


Figure 3. TEM observations of wild type and class A mutant strains. Polysaccharide contents of ultrathin sections (60 nm) of the wild type CLg1 strain (A), 99A7 (B) and 175E12 (C) mutant strains were observed after PATAg staining. Both starch like granules (white star) and glycogen particles (dots pointed by black arrows) are witnessed in the wild type strain. Starch granules are missing and substituted by a large amount of glycogen particles in class A mutants. The dark-blue iodine stain from a cell patch of the wild type strain (framed in panel A), is displayed. In contrast, the absence of starch granules in class A mutants is correlated by a yellow-orange stain of cell patches after spaying iodine vapours (framed in panels B and C).

2.3. The glycogen accumulating cyanobacteria are defective for a cation requiring debranching enzyme activity.



Supplemental figure 2: **Zymogram analysis of phosphorylase activity from wild type and class A mutants.** Total protein of WT and class A mutant strains were equilibrated and separated by native PAGE, followed by transfer of proteins to a native PAGE containing 0.6% (w/v) of glycogen. The native gels were then incubated with (A) or without (B) 20 mM Glucose-1-Phosphate (G-1-P). After overnight of incubation, gels are stained with iodine solution. Phosphorylase activity (arrow #1) is induced in all class A mutants, except for 174B6, 123B3 mutants and WT strain. A hydrolytic activity (arrow#2) is present in all crude extracts.

To get a better understanding of the underlying biochemical cause of the glycogen accumulating phenotype, we undertook a series of zymogram and enzyme assays for all possible enzymes of bacterial glycogen/starch metabolism. We were unable to find any significant difference in our crude extract assays. A notable increase of one starch phosphorylase isoform was noted through zymograms of the mutants (Supplemental Figure 2). However this increase was noted in all mutant types that displayed a decrease in total polysaccharide amount, and not only in Class A mutants. It thus seems to define a universal secondary effect of altered starch metabolism not only in the *Cyanobacterium* sp. CLg1 but also in other cyanobacteria species (Fu and Xu, 2006). Interestingly, the presence of white staining glycogen degrading bands that selectively disappeared from the class A mutants was

evidenced from the glycogen/starch synthase gels but was not mirrored by the starch amylopectin or glycogen containing gels designed to emphasize starch modifying enzymes (e.g. amylases, branching enzymes, and α-1,4 glucanotransferase) (Figure 4).

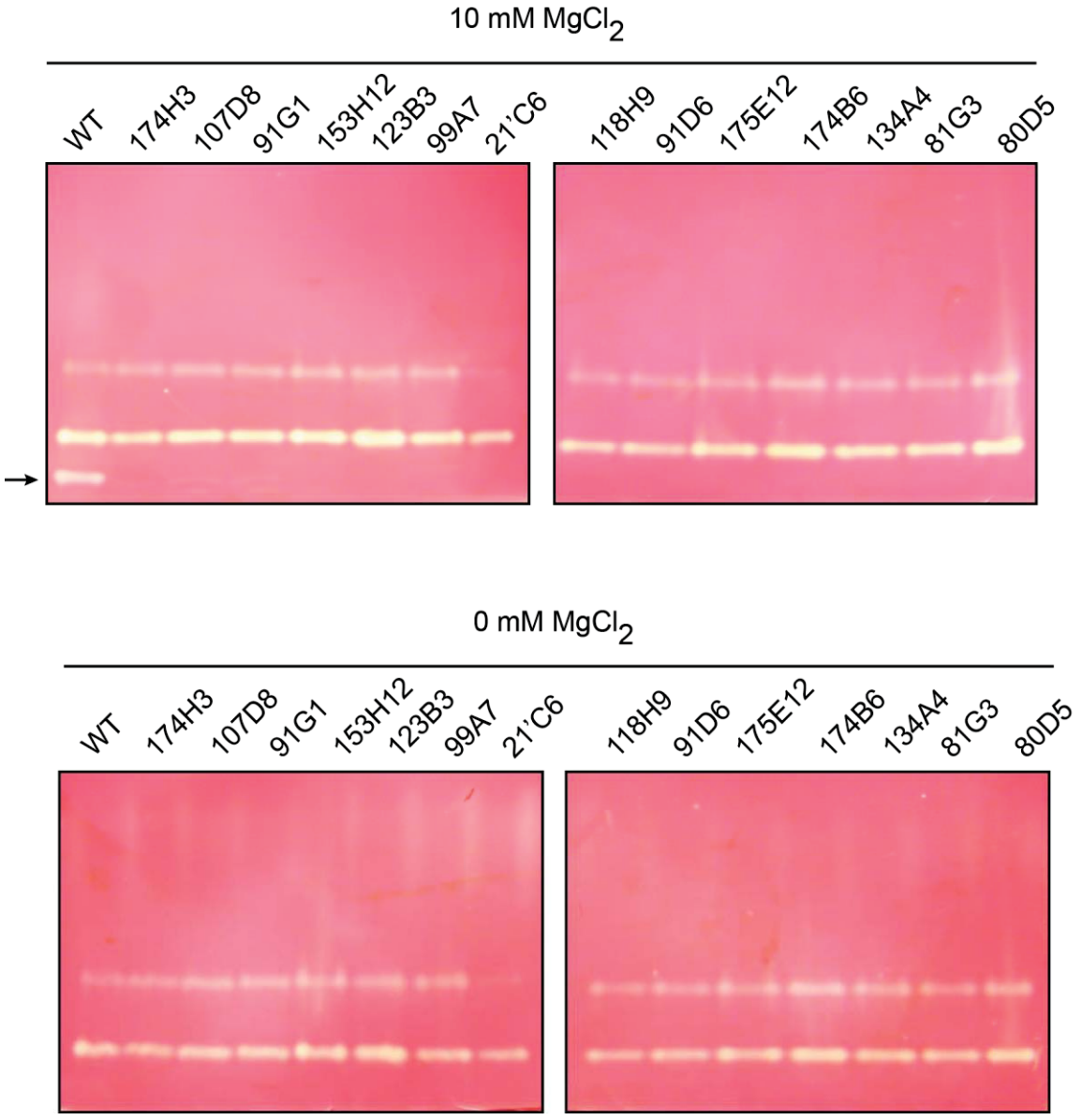
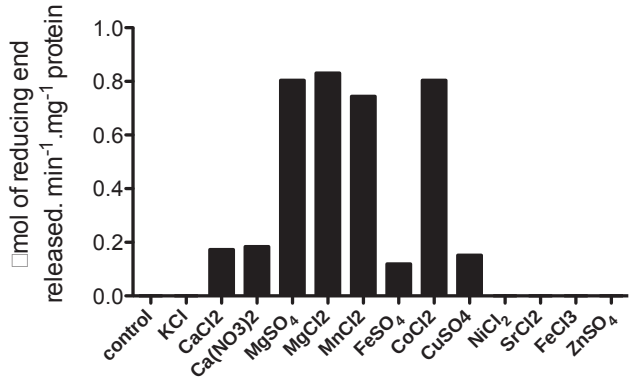


Figure 4. **Zymogram analysis of starch metabolizing enzymes from wild type and class A mutants.** Total protein of WT and class A mutant strains (21'C6, 80D5, 81G3, 91D6, 91G1, 99A7, 107D8, 118H9, 123B3, 134A4, 153H12, 174B6, 174H3, 175E12) were separated by native PAGE, followed by transfer of proteins to a native PAGE containing 0.6% (p/v) of amylopectin. The native gels were then incubated with (A) or without (B) 10 mM $MgCl_2$. A cation dependent activity enzyme is witnessed after iodine staining in the wild type's crude extract and disappears in all class A mutants (black arrow).

Because the procedures differed between the hydrolase and synthase zymograms with respect to pH and incubation buffers, we were able to test and narrow down the differences

responsible for the identification of this activity selectively in the glycogen/ starch synthase activity gels. We found that the missing glucan hydrolase required high levels of cation (supplemental Figure 3) that were supplied in the glycogen/ starch synthase zymograms (i.e Mg^{2+}) but not in our other procedures.



Supplemental figure 3: **Effect of cations on the debranching enzyme activity.** Cation free debranching enzyme was incubated in absence (control) or in the presence of 5 mM of K^+, Ca^{2+}, Mg^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, Sr^{2+}, Fe^{3+}, Zn^{2+} and 0,2% of amylopectin at 30°C during 30 minutes. Reactions were stopped by adding a volume of sodium 3,5-dinitrosalicylate (DNS) and boiled 10 minutes at 99°C. After cooling down, the absorbance is measured at 540 nm. The specific activity (µmol of reducing end.min^{-1}.mg^{-1} protein) was determined in presence of each cation.

Figure 3 shows that when starch or amylopectin containing gels were supplied with 10 mM $MgSO_4$ a blue staining glucan hydrolase band that disappeared in all class A mutants was evidenced. Interestingly the incompletely defective class A mutant display a slower migrating fainter activity. Because such a stain was indicative of the presence of a starch debranching enzyme, we purified the activity from wild-type cyanobacteria to near homogeneity. We then subjected amylopectin to the action of the purified enzyme and examined the proton NMR spectra of the substrate before and after incubation with the glucan hydrolase. Results displayed in Figure 5 demonstrate that amylopectin was completely debranched and that the amount of branch hydrolysis corresponded quantitatively to the appearance of reducing end resonance signals.

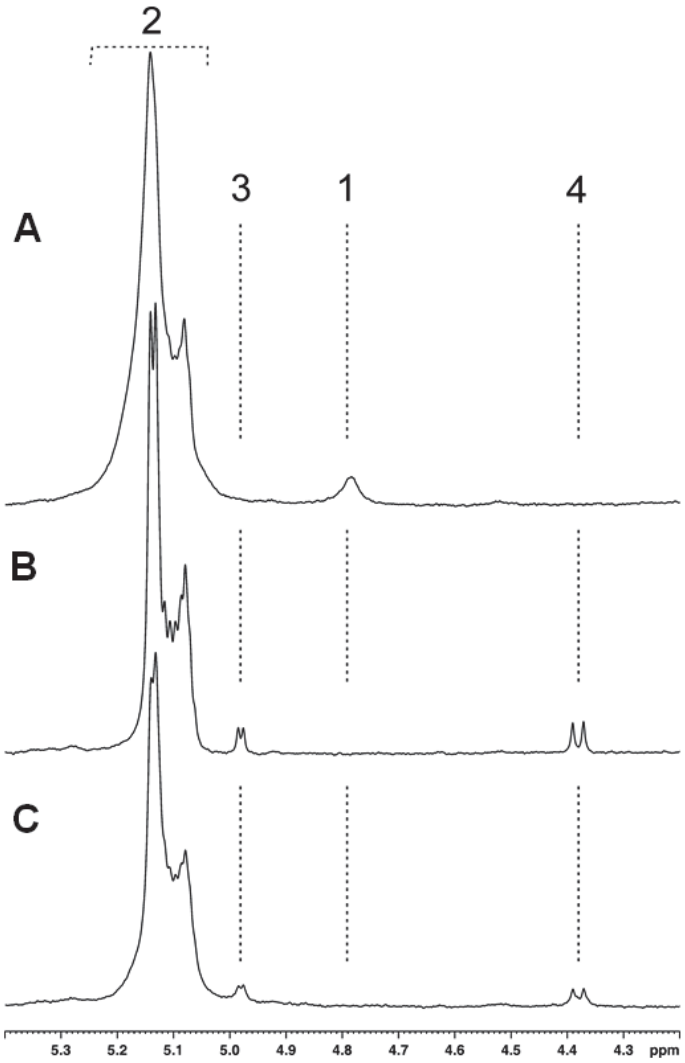
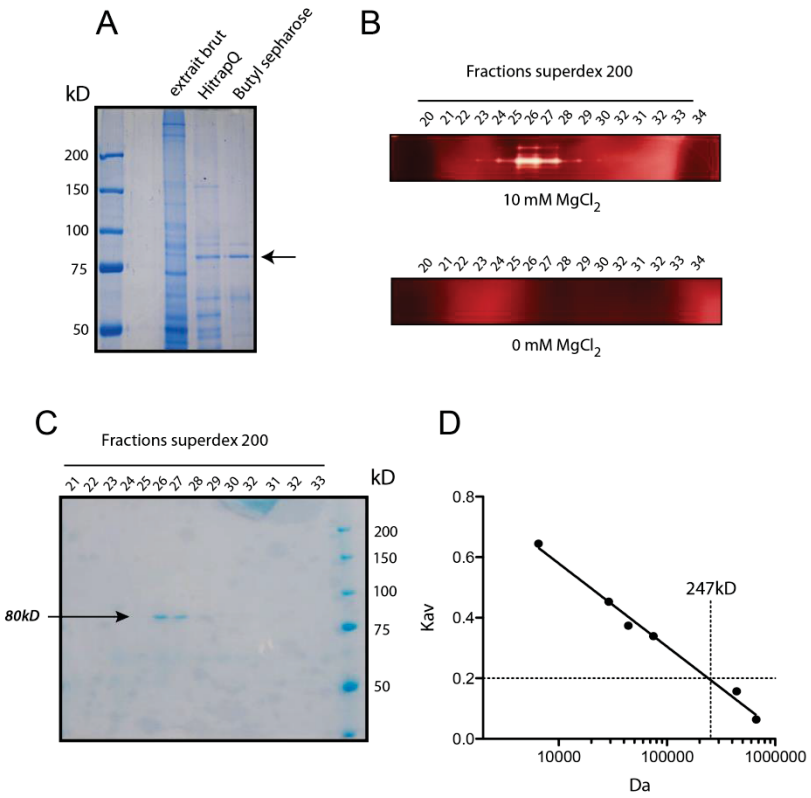


Figure 5. Part of the ^{1}H-NMR spectra of amylopectin in dimethyl-sulfoxyde. NMR analysis was carried out on amylopectin (A) and amylopectin samples incubated overnight with a commercial isoamylase (B) or with the cation dependent enzyme activity purified from wild type strain (C). Peak #2 (5.2 ppm to 5.08 ppm) and peak #1 (4.79 ppm) represent the signals of ^{1}H engaged in α-1,4 and α-1,6 linkages, respectively. Both incubation experiments with amylopectin (B and C) result in the release of reducing ends and the apparition of α- (peak #3) and β-anomeric protons (peak #4) at 4.98 and 4.38 ppm, respectively. The presence of peaks #3 and #4 are correlated with a disappearance of the ^{1}H signal engaged in α-1,6 linkages (peak #1). The NMR spectrometry analysis suggests that cation dependent activity hydrolyzes specifically α-1,6 linkages.

These results allow us to identify the purified enzyme as a direct debranching enzyme. In addition we were able to size the activity during our purification procedure as a polypeptide of 82 kD in denaturing conditions and with an apparent molecular weight of 247 kD in non-denaturing conditions (supplemental Figure 4), which suggests that the native enzyme eluted as a homotrimeric or as a homodimeric activity.



Supplemental figure 4: Purification of cation dependent activity. A. Cation dependent activity was purified at homogeneity using an anion exchange chromatography (HitrapQ), hydrophobic column (butyl sepharose) and gel permeation chromatography (superdex 200). **B**. Debranching enzyme activity was monitored by using zymogram analysis performed in presence or in absence of 10 mM of $MgCl_2$ during the purification steps. **C.** SDS-PAGE analysis performed on superdex 200 fractions. A polypeptide of weight of 80 kD is visualized in fractions # 26 and #27 after Coomassie blue staining. **D.** Superdex 200 column (GE-Healthcare) pre-equilibrated with 150mM NaCl, tris/acetate buffer pH 7.5, 10 mM DTT was calibrated with standard proteins (669; 440; 75; 29 and 6,5 kD) and dextran blue. The determination of a partition coefficient (Kav) of 0.2 suggests an apparent molecular weight of the cation dependent activity estimated at 247 kD.

2.4. Glycogen-accumulating mutants carry mutations in a GlgX-like debranching enzyme activity.

The presence of pure enzyme preparations enabled us to identify a dozen of oligopeptide sequences that matched the protein sequences deduced from the genome (supplemental Figure 5).

M L M G D E S M N P Q V W P G N P Y H L G A K W D G Q G T N F A L Y S E N A T T V E L C F F D
R Q N Q E T R I P L T E V Q N Y V W H A Y L P G I M P G Q R Y G Y R V D G V Y D P E E G H R F N
V N K L L I D P Y A K A L D G E I G F G E P I F G Y V W E D E D E D I S F S E L D S A H L V P K A V
V V D E S F D W E G D K P L D I P E H E T I I Y E L H V K G F T **K L H P D I P E D L R G** T F A G L
A H P S T I V Y L K S L G I T A I E L M P I H H F L S Q P G H L V E K D L T N Y W G Y D S I S Y L A
P F S G Y C A T K T P Q D Q V K E F K A M V K A L H Q E G I E V I M D V V Y N H T G E G N H F G
P T L S L R G I D N A T Y Y R V V E E D P R **Y Y M D F T G C G N S L N** L K H P Q V M K L I M D S
L R Y W V L E M H V D G F R F D L A A A L A R E L L E V D C L A T F F D I I H Q D P V L S N I K L I
A E P W D I G E D G Y H V G K F P V L W S E W N G R Y R D T V R N F W R G E K **S I L A E F A Y**
R V T G S S D L Y Q D N G R T P S A S I N F I T A H D G F T L N D L V S Y N H K H N D A N G E E N
K D G E Q F N H S W N C G E E G D S K D P E V L S L R N Q Q R R N F L V T L M L S Q G V P M L T
A G D E I G R T Q K G N N N A F C Q D N E I S W V D W S L E H K N A E L L K F V R D L I D L R H
Q H P V F R R R K W F Q G Q D I H G S G V S D I G W F N P D G F V T A E T Q W N L G F A K A I A
L F L N G Q E I P L K D Q Q G Q R V V D N S F L L F F N A H Y E A I E F V I P E S L G K Q D W I M
V I D T T Q S R L L E S G K R Y R H D V S I K V E A R S L V V L K T T N Stop

Supplemental figure 5: Identification of a cation dependent activity by Nano-LC-MS-MS. NCBI blast search identifies seven trypsic peptides common in GlgX proteins (underlined peptides). In addition, three trypsic peptides match specifically with CLg1-GlgX2 (bold underlined peptides) and not with CLg1-GlgX1. These data suggest that the cation dependent activity is probably an homotrimeric or homodimeric complex of GlgX2 (see text).

We were thus able to identified peptides that corresponded to *GlgX2*, one of the two GlgX-like genes present on the *Cyanobacterium* sp.CLg1 genome sequence.Highly pure enzyme preparations enabled us to size the protein and ascertain the possible presence of distinct enzyme subunits. No bands corresponding to the expected size of GlgX1 and no trypsic fragments matching this enzyme sequence could be retrieved. We therefore conclude that this enzyme is a homomultimer (dimer or trimer). We then sequenced all starch metabolism genes from the 11 class A as well as three incompletely defective class A mutants and systematically found mutations only in the GlgX2 sequence, altering the amino-acid sequence of the protein, while no mutations could be found in any of the other starch metabolic genes. These mutations are summarized in Figure 6.

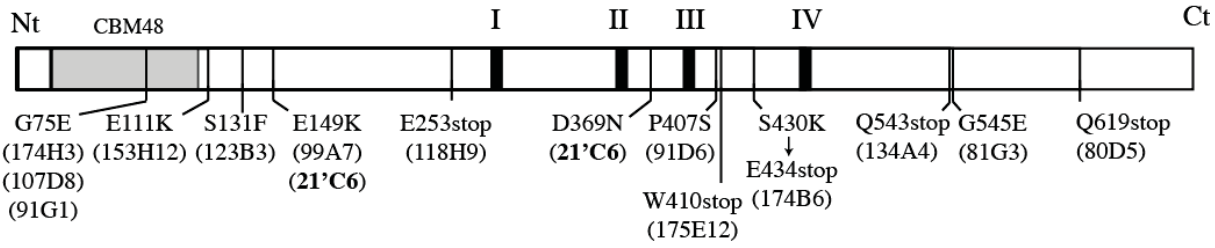


Figure 6. Various allelic mutations in the *Cyanobacterium* sp. CLg1 glgX2 gene. Forward and reverse primers were designed in the untranslated region of the glgX2 gene. PCR reactions were carried out on genomic DNA for each class A mutants. PCR products were cloned and DNA was sequenced on both strands. Punctual mutations (vertical black lines) found in each mutant (name in brackets) are shown on the glgX2 segment. A mutant strain, 21'C6 harbors two point mutations (bold name). Regions I, II, III and IV represent highly conserved sequences in the α-amylase family which include acid residues involved in the catalytic site (Suzuki *et al.*, 2007). The Carbohydrate Binding Domain 48, CBM 48, (grey box) is observed at the N terminus (Nt) (Janecek *et al.*, 2011).

Among these mutations we found four nonsense mutations resulting in three premature stop codons, one of which was preceded by a frameshift; the other ten mutants defined different missense alleles. Three of these missense alleles were responsible for the incompletely defective phenotype of 153H12, 123B3 and 176B6 strains. Because we hereby define an allelic series of 11 (+3 incompletely defective) independent mutants none of which carried additional mutations in starch metabolism genes (supplemental Table1) we can safely conclude that all defective phenotypes recorded in relation to polysaccharide accumulation result directly from the GlgX2 defect and not from other causes obviating the need for formal complementation of the mutant gene through transgenesis with the wild-type sequence. Successful transgenesis has indeed never been reported in this subgroup of cyanobacteria and all our attempts were unsuccessful.

activity	gene	Strains				
		80D5	99A7	118H9	153H12	175 E12
Debranching enzymes	*glgX2*	-	-	-	-	-
	iDBE	-	-	-	-	-
	apu GH13	-	N.D.	-	-	-
	apu GH57	-	-	-	N.D.	-
Branching Enzymes	*glgB1*	-	N.D.	-	-	-
	glgB2	N.D.	-	-	-	-
	glgB3	-	-	-	-	-
	glgB GH57	N.D.	-	-	-	-
Starch synthases	*glgA1*	-	-	-	-	-
	glgA2	-	-	-	N.D.	-
	gbss	-	-	-	-	-
Phosphorylase	*pho*	N.D.	N.D.	-	N.D.	-
a-1,4 glucano transferase	*malQ*	-	-	-	-	-

Supplemental table 1: Molecular cloning and sequencing of genes involved in starch metabolism pathway in five class A mutants. Primers used sequenced the others gene coding branching enzyme, of the GH13 family (*glgB1*, and *glgB2*) and of the GH 57 family (*glgB57*), and the others gene coding debranching enzyme: the indirect debranching enzyme (*iDBE*) and the both amylopullanase of the GH 13 family (*apu13*) and the GH57 family (*apu57*). This table shows no others mutation in the different gene of the class A mutant.

2.5. The glgX-like debranching enzyme defines an isoamylase type of activity

In bacteria and cyanobacteria GlgX-type enzymes display a marked substrate preference for glycogen whose outer chains have been recessed with starch phosphorylase and little activity with either amylopectin or glycogen. The pure GlgX2 enzyme displayed little activity with such substrates or even with glycogen (Figure 7). However unlike glgX and in fashion reminiscent to the plant isoamylase (Hussain et al. 2003) it displayed a marked preference for amylopectin which it debranched to completion (Figure 7). Hence the GlgX2 enzyme of *Cyanobacterium* resembles more the potato isoamylase substrate specificity than that of the reference *Pseudomonas* sp. enzyme. It certainly deviated from the bacterial GlgX type of activity classically involved in glycogen catabolism.

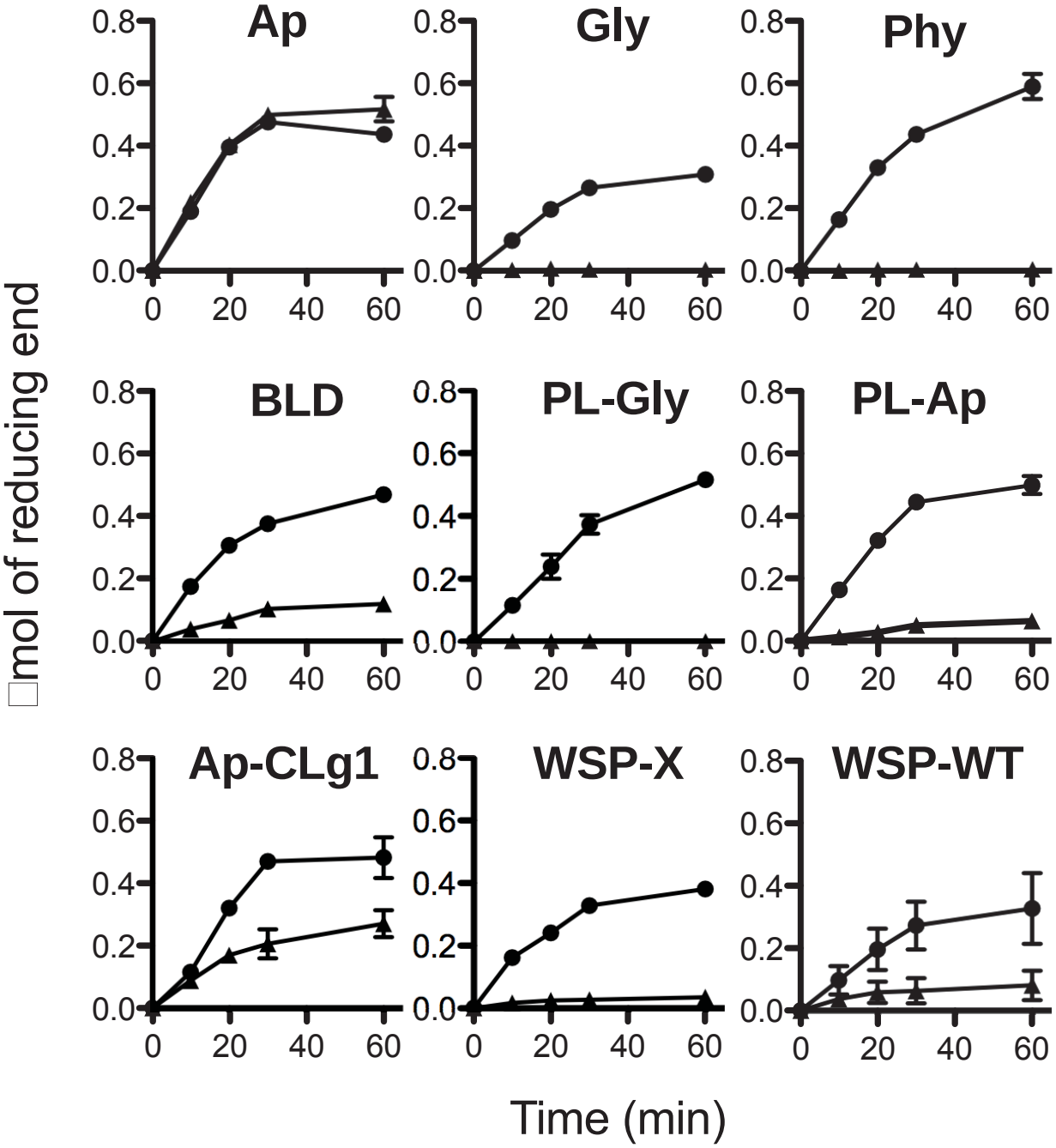


Figure 7. Specificity of the cation dependent debranching activity towards different polysaccharides. Both purified debranching enzyme activity (black triangle) and commercial isoamylase produced by *Pseudomonas sp*. (megazyme) (black circle) were incubated with 0.5% of amylopectin (Ap), glycogen (Gly), phytoglycogen (Phy), beta-limit dextrin of amylopectin (BLD), phosphorylase-limit dextrin of glycogen (PL-gly), phosphorylase-limit dextrin of amylopectin (PL-Ap), semi-amylopectin of the wild type CLg1 strain (Ap-CLg1), water soluble polysaccharide of class A mutant (WSP-X) and water polysaccharide of wild type CLg1 strain (WSP-WT). The release of reducing ends was determined at 10; 20; 30 and 60 min of incubation by using the DNS method. In order to compare both debranching enzyme activities, the specific activity of commercial isoamylase was adjusted at 0.5 µmol of reducing ends. min^{-1}. mL^{-1} using amylopectin as substrate.

2.6. GlgX2 displays a hydrolytic activity towards specific chains onto glycogen-type substrat.

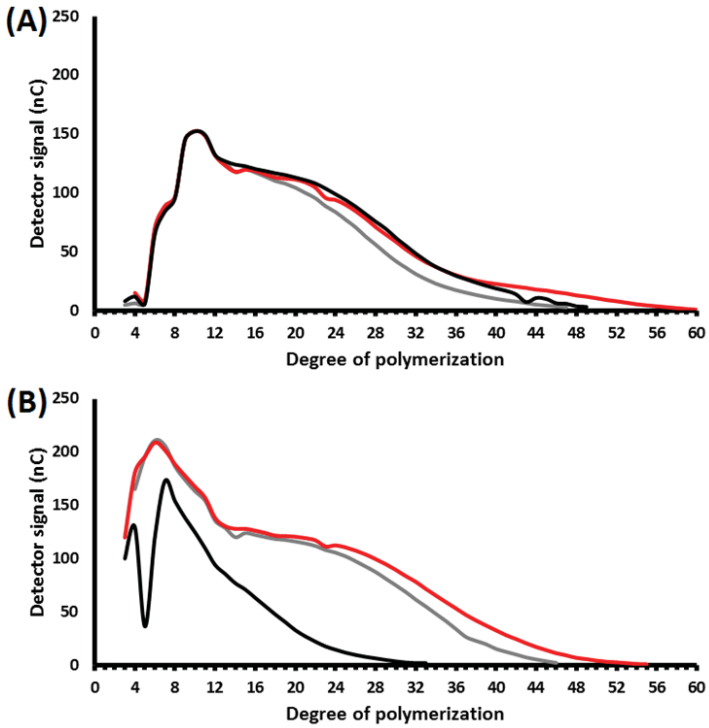


Figure 8: Linear glucans profiles released by different debranching enzymes. The commercial isoamylase (grey curve) and the enzyme GlgX2 (black line) were incubated with amylopectin (A) and glycogen (B) overnight. Linear glucans were separated by HPAEC-PAD according to their polymerization degree.The polysaccharide, incubated with GlgX2, was further incubated with commercial isoamylase in order to check whether the debranching reaction was complete (red curve). With amylopectin, a similar profile was observed with both enzymes; showing a total debranching mechanism by the cation-dependent enzyme, while the cation-dependent activity debranch partially the glycogen. Interestingly, in the case of glycogen the long chains of glucan (DP>12) and short glucan chains (DP=5) are not or fully debranched, suggesting a good arrangement of these chains to carry the precursor of amylopectin.

The isoamylase of Archaeplastida is considered to be active on a soluble glycogen-type polysaccharide (Ball *et al.*, 1996 Dauvillée *et al.*, 2001), however, recent studies on the ISA1/ ISA3 isoamylases of *S. tuberculosum* (Hussain *et al.*, 2003) and GlgX2 enzyme (Figure 7) suggest that these enzymes are active preferentially onamylopectin-like substrates, compared to glycogen-like polysaccharides. This marked substrat preference reflects probably the recognition of long chains and the organization of α,1-6 by enzymes involved in polysaccharides crystallization. In order to get some information about the debranching enzymes specificity, the linear glucan chains released after 12 hours of incubation either with GlgX2 (black curve) or with commercial isoamylase (grey curve) were obtained with amylopectin (Figure 8A) or with glycogen (Figure 8B). Both enzymes display a similar

efficacity with amylopectin as a substrate (Figure 8A). In addition, chain length distribution does not change when amylopectin, incubated first with GlgX2, was further debranched in presence of commercial isoamylase (red curve). The perfect match of both profiles suggests that GlgX2 acts unspecifically on α-1,6 linkages of amylopectin in a similar way thancommercial isoamylase. However, GlgX2 acts poorly onto a-1,6 linkages of glycogen and specifically onto short branched glucan made of five residus of glucose (Figure 8). This difference towards polysaccharide glycogen-like, in particular for the chain length of five glucoses (DP=5) could indicate unexpected debranching activity of the GlgX2 enzyme towards short DP, generally suspected as preventing the crystallisation of amylopectin polysaccharide.

2.7. GlgX2 defines the enzyme responsible for starch accumulation in *Cyanobacterium* only.

The mutants defined in this work establish that the *Cyanobacterium* sp. CLg1 aggregates starch from a hydrosoluble precursor in a fashion very similar to starch in green algae and land plants. Since in both cases an enzyme from the same CAZy GH 13 subfamily 11 has been recruited to splice out those branches that prevent polysaccharide aggregation during synthesis, it is of interest to examine the phylogeny of these enzyme sequences and correlate the presence of selective enzymes forms to those of starch-like polymers. The phylogeny displayed in Figure 7 confirms that the source of the archaeplastidal debranching enzymes cannot be traced to cyanobacteria. The strong monophyletic grouping of the archaeplastida enzymes with those of chlamydiales pathogens establish the latter as the source of direct debranching enzyme, a bacteria-specific activity, in photosynthetic eukaryotes (Ball *et al.,* 2013). In addition the presence of the particular GlgX2-type enzyme does not correlate with the presence of starch in cyanobacteria. In Figure 9, all other documented starch accumulating cyanobacteria lack this particular form of the GlgX enzyme while the closest relatives to GlgX2 are reported to accumulate glycogen. We can therefore conclude that while acquisition of starch in Chloroplastida can still be suggested to be monophyletic, transition of glycogen to starch metabolism in cyanobacteria cannot and defines distinctive polyphyletic events.

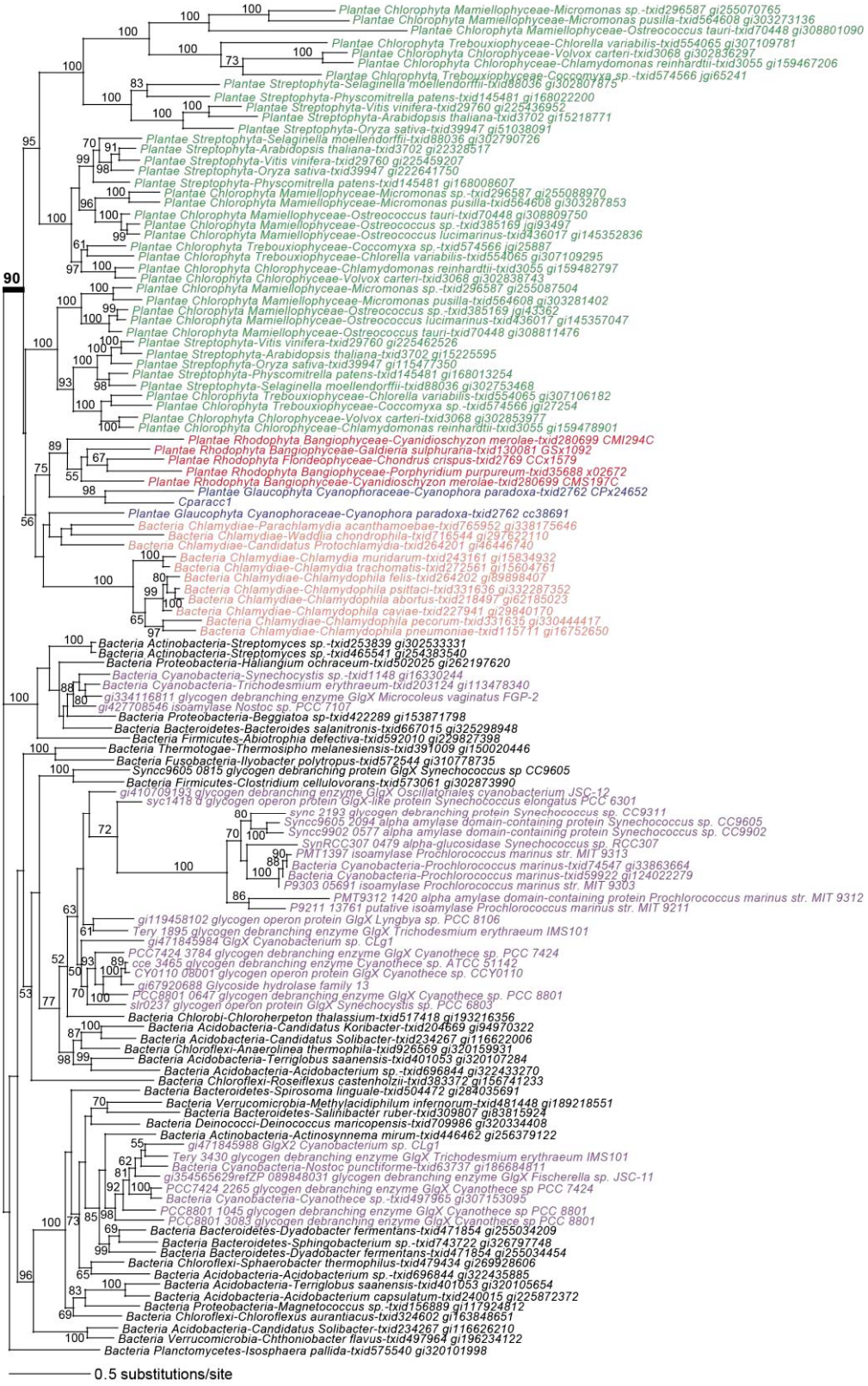


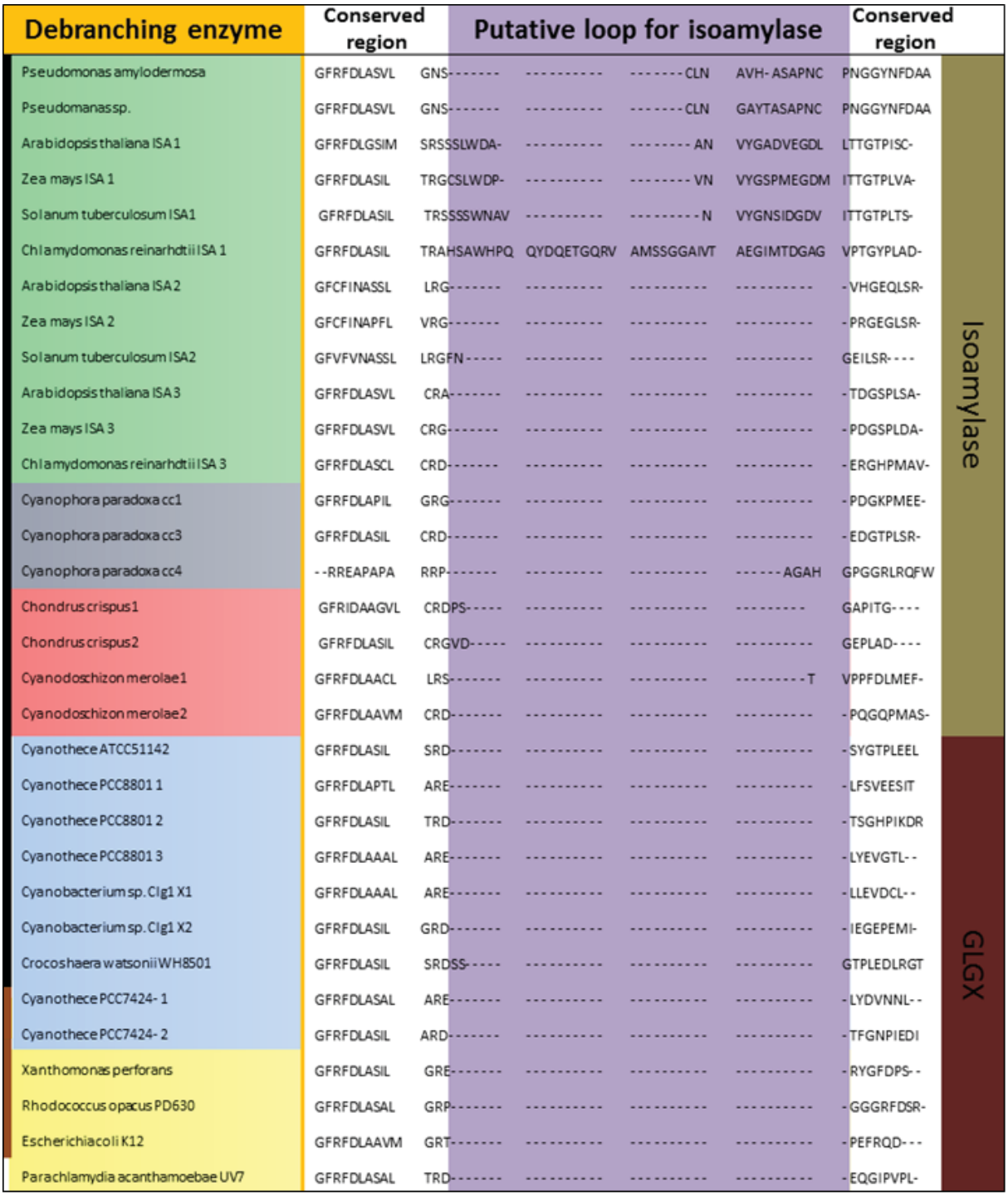
Figure 9. **Phylogenetic analysis of cyanobacterial, bacterial, and archaeplastidal glycogen debranching enzymes.** Starch debranching enzymes in Chloroplastida play an important role in polysaccharide synthesis and in starch degradation. The maximum likelihood phylogeny of these enzymes shows that the three isoforms of isoamylase in Chloroplastida (green text) and the debranching enzyme found in Glaucophyta (blue text), and Rhodophyta (red text) do not share the same phylogenetic origin as the cyanobacterial GlgX

(magenta text)), particularly for *Cyanobacterium* sp. CLg1. Rather, GlgX enzymes in Archaeplastida are sister to Chlamydiae (orange text; bootstrap support value = 90%), strongly suggesting a chlamydial origin of the gene in this eukaryotic supergroup (Ball *et al.*, 2013)

3. Discussion

***Cyanobacterium* sp. CLg1 accumulates both glycogen and cyanophycean starch.** The significant accumulation of both glycogen and starch-like material in wild-type axenized *Cyanobacterium* sp. CLg1 is a distinctive feature of these organisms. . Archaeplastida, with the exception of the red alga *Porphyridium sordidum,* (Shimonaga *et al.*, 2008), accumulates only starch; eventually traces of WSP and oligosaccharides are sometimes detectable. In this study however, under conditions of maximal polysaccharide synthesis 10% of the total storage polysaccharide pool always accumulates in the form of glycogen. We deem this amount to be physiologically significant. The presence of both types of storage polysaccharides may in this case be required to optimize cyanobacterial physiology by making a significant pool of glucose available through a more dynamic and readily accessible form of storage. Indeed in glycogen the glucose of the outer chains may be easier to mobilize following fluctuations of physiology both in light and darkness. The crystalline starch-like granules in turn will offer a larger pool of osmotically inert carbohydrate stores for delayed use by nitrogenase or cell division at night (Schneegurt *et al.*, 1994). It will be of interest to check for the presence of an analogous fraction in other starch storing cyanobacteria. We confirm the presence of starch-like granules in *Cyanobacterium* sp. CLg1 composed of both an amylopectin-like high mass fraction and a smaller amylose fraction. The chain-length (CL) distribution of the high mass polysaccharide complies to the definition given for semi-amylopectin as it contains fewer of those chains exceeding DP 40 (Nakamura *et al.*, 2005). Because the granules also contain a significant amount of amylose (5%), we propose to call this material cyanobacterial starch. Convergent evolution in cyanobacteria and Archaeplastida suggests that polysaccharide debranching may define a universal requirement for evolution of aggregated semi-crystalline polysaccharides from glycogen metabolism**.** The experiments reported here establish that *Cyanobacterium* sp. CLg1 use an homodimeric (or possibly trimeric) debranching enzyme to aggregate polysaccharides into insoluble granules**.** In Archaeplastida an analogous heteromultimeric isoamylase (possibly hexameric) composed of two distinct types of subunits achieves a similar function (Utsumi *et al.*, 2011). In the absence of the non-catalytic isa2 subunit, depending on the plant species under study an homodimer

isa1 may suffice to achieve starch synthesis (Kubo *et al.,* 2010). In *Chlamydomonas reinhardtii* disruption of the isa2 structural gene *STA8* yields an isa1 homodimer that yields the appearance of both starch and glycogen in a fashion reminiscent of *Cyanobacterium* sp.CLg1 (Dauvillée *et al.,* 2000; Dauvillée *et al.,* 2001a; Dauvillée *et al.,* 2001b). In both, cyanobacteria and plants, a glycohydrolase of a similar CAZy GH13 subfamily (subfamily 11 also known as TreX-GlgX) has been recruited. GlgX has been documented in proteobacteria and cyanobacteria to be selectively involved in glycogen degradation through both genetic and biochemical characterizations. In *E. coli*, GlgX was proven to be restrictive with respect to its substrate preference by debranching only or preferably those chains that have been recessed by glycogen phosphorylase during glycogen degradation (Dauvillée *et al.,* 2005). This restrictive specificity was shown through the determination of the 3D structure of the *E. coli* enzyme to be due to the shorter size of the inferred substrate binding groove. The potato isoamylase was demonstrated to display a preference that accommodates amylopectin with comparatively little activity on glycogen or on either glycogen or amylopectin predigested by glycogen phosphorylase (Hussain *et al.,* 2003). The cyanobacterial enzyme GlgX2 is here demonstrated to display the very same substrate preference as that reported for the potato enzyme (Hussain *et al.,* 2003). In addition, the theoritical loop described for isoamylase of plants and supposed to be responsible for the substrat specificity is missing in all GlgX including GlgX2 (Figure 10).

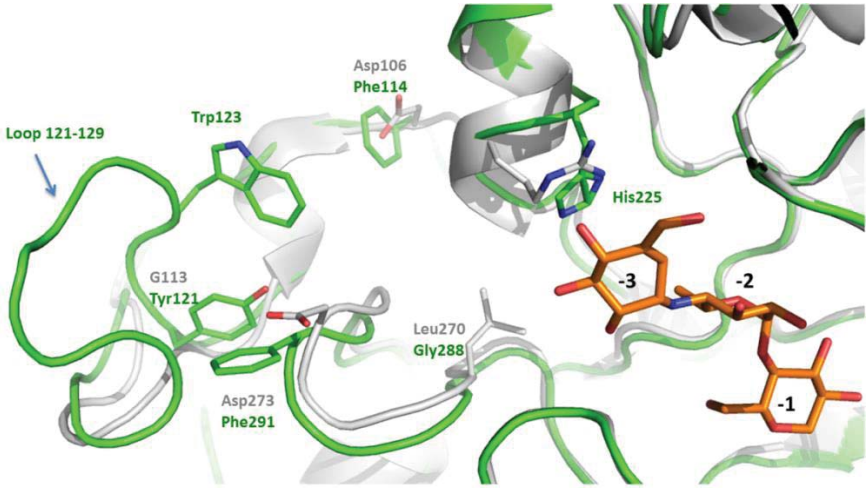


Debranching enzyme	Conserved region		Putative loop for isoamylase				Conserved region	
Pseudomonas amylodermosa	GFRFDLASVL	GNS-------	----------	----------	--------CLN	AVH-ASAPNC	PNGGYNFDAA	Isoamylase
Pseudomanas sp.	GFRFDLASVL	GNS-------	----------	----------	--------CLN	GAYTASAPNC	PNGGYNFDAA	
Arabidopsis thaliana ISA1	GFRFDLGSIM	SRSSSLWDA-	----------	----------	---------AN	VYGADVEGDL	LTTGTPISC-	
Zea mays ISA 1	GFRFDLASIL	TRGCSLWDP-	----------	----------	---------VN	VYGSPMEGDM	ITTGTPLVA-	
Solanum tuberculosum ISA1	GFRFDLASIL	TRSSSSWNAV	----------	----------	---------N	VYGNSIDGDV	ITTGTPLTS-	
Chlamydomonas reinarhdtii ISA 1	GFRFDLASIL	TRAHSAWHPQ	QYDQETGQRV	AMSSGGAIVT	AEGIMTDGAG	VPTGYPLAD-		
Arabidopsis thaliana ISA2	GFCFINASSL	LRG-------	----------	----------	----------	----------	-VHGEQLSR-	
Zea mays ISA 2	GFCFINAPFL	VRG-------	----------	----------	----------	----------	-PRGEGLSR-	
Solanum tuberculosum ISA2	GFVFVNASSL	LRGFN-----	----------	----------	----------	----------	GEILSR----	
Arabidopsis thaliana ISA3	GFRFDLASVL	CRA-------	----------	----------	----------	----------	-TDGSPLSA-	
Zea mays ISA 3	GFRFDLASVL	CRG-------	----------	----------	----------	----------	-PDGSPLDA-	
Chlamydomonas reinarhdtii ISA 3	GFRFDLASCL	CRD-------	----------	----------	----------	----------	-ERGHPMAV-	
Cyanophora paradoxa cc1	GFRFDLAPIL	GRG-------	----------	----------	----------	----------	-PDGKPMEE-	
Cyanophora paradoxa cc3	GFRFDLASIL	CRD-------	----------	----------	----------	----------	-EDGTPLSR-	
Cyanophora paradoxa cc4	--RREAPAPA	RRP-------	----------	----------	----------	------AGAH	GPGGRLRQFW	
Chondrus crispus1	GFRIDAAGVL	CRDPS-----	----------	----------	----------	----------	GAPITG----	
Chondrus crispus2	GFRFDLASIL	CRGVD-----	----------	----------	----------	----------	GEPLAD----	
Cyanodoschizon merolae1	GFRFDLAACL	LRS-------	----------	----------	----------	---------T	VPPFDLMEF-	
Cyanodoschizon merolae2	GFRFDLAAVM	CRD-------	----------	----------	----------	----------	-PQGQPMAS-	
Cyanothece ATCC51142	GFRFDLASIL	SRD-------	----------	----------	----------	----------	-SYGTPLEEL	GLGX
Cyanothece PCC8801 1	GFRFDLAPTL	ARE-------	----------	----------	----------	----------	-LFSVEESIT	
Cyanothece PCC8801 2	GFRFDLASIL	TRD-------	----------	----------	----------	----------	-TSGHPIKDR	
Cyanothece PCC8801 3	GFRFDLAAAL	ARE-------	----------	----------	----------	----------	-LYEVGTL--	
Cyanobacterium sp. Clg1 X1	GFRFDLAAAL	ARE-------	----------	----------	----------	----------	-LLEVDCL--	
Cyanobacterium sp. Clg1 X2	GFRFDLASIL	GRD-------	----------	----------	----------	----------	-IEGEPEMI-	
Crocoshaera watsonii WH8501	GFRFDLASIL	SRDSS-----	----------	----------	----------	----------	GTPLEDLRGT	
Cyanothece PCC7424- 1	GFRFDLASAL	ARE-------	----------	----------	----------	----------	-LYDVNNL--	
Cyanothece PCC7424- 2	GFRFDLASIL	ARD-------	----------	----------	----------	----------	-TFGNPIEDI	
Xanthomonas perforans	GFRFDLASIL	GRE-------	----------	----------	----------	----------	-RYGFDPS--	
Rhodococcus opacus PD630	GFRFDLASAL	GRP-------	----------	----------	----------	----------	-GGGRFDSR-	
Escherichia coli K12	GFRFDLAAVM	GRT-------	----------	----------	----------	----------	-PEFRQD---	
Parachlamydia acanthamoebae UV7	GFRFDLASAL	TRD-------	----------	----------	----------	----------	-EQGIPVPL-	

Figure 10: Alignment of plants and prokaryotes debranching enzymes protein sequences. Black and orange lines on the left indicate organisms accumulating starch or glycogen, respectively. The white line indicates that the polysaccharide stored is most likely a glycogen-like, but is not clearly described. Isoamylases or GlgX enzymes are shown on the right. This figure shows the presence of a loop between two conserved regions in all Chloroplastida and Pseudomonas isoamylase sequences. So far, this loop was considered to be responsible for the substrat specificity of these enzymes. However, the recent access of isoamylase sequences of Rhodophyta andr Glaucophyta reveals that the loop is missing in theses isoamylases and probably the amino acid insertion occurs independently in

chloroplastida lineage after separation from other Archaeplastida lineages (Rhodophyta and Glaucophyta).

To understand the basis for the change in enzyme specificity we have modeled the enzyme structure on the structure of the TreX enzyme and compared this to the GlgX structure reported for the *E. coli* enzyme (Woo *et al.*, 2008; Song *et al.*, 2010). One can speculate that CLg1-GlgX2 might have a more extended substrate binding groove compared to the *E. coli* GlgX, and can be attributed to the presence of large aromatic residues, which are absent in the *E. coli* GlgX (supplemental Figure 6).



Supplemental figure 6: Molecular modeling of CLg1-GlgX2. In Green: modeled structure of CLg1-GlgX (based on Swiss-model automated threading protocol, to Trex enzyme). Grey: Crystal structure of *E. Coli* GlgX. Orange: Ligand bound in the active site of Trex (-3, -2, -1 denote glucose subsites.). Represented in sticks are the residues which differ between CLg1-GlgX and the *E. coli* Glgx which may account for the higher activity of CLg1-GlgX for longer substrates. Aromatic residues in CLg1-GlgX (modeled in green), not found in *E.coli* GlgX (grey) are proposed to contribute towards an extended substrate binding site.

However these structure models need to be confirmed and detailed through X-Ray diffraction of the corresponding proteins. The correct substrate specificity may not be the only determinant required for isoamylase function in polysaccharide aggregation. The homo or heteromultimeric organization seems equally relevant. *Cyanobacterium* sp. CLg1-GlgX2 is part of the same GH13-subfamily 11 subclade as the *E. coli* GlgX2 which unlike TreX or isa1-2 from plants was proven to be monomeric. It would be of interest to investigate the quaternary structure of other members of this debranching enzyme clade and check the correlation with either glycogen or starch accumulation. Indeed the switch to an isoamylase specificity together with one particular type of multimeric organization might define the two

major determinants determining starch versus glycogen accumulation. Unfortunately structure modeling cannot bring convincing answers and more detailed biochemical and structural characterization will be required to ascertain this issue. In any case the work detailed in this paper proves that at least in *Cyanobacterium* sp. CLg1 a direct debranching enzyme has been recruited and evolved an isoamylase specificity together with a multimeric organization that enabled these cyanobacteria to switch from glycogen synthesis to starch metabolism. Because both, the chlamydial GlgX recruited by the Archaeplastida ancestors and the GlgX enzyme from cyanobacteria went through a similar convergent path to generate enzymes with similar properties, we propose that polysaccharide debranching could define a universal mechanism that enables conversion of glycogen metabolism into starch in all living cells. Such an evolution might have been favored each time a particular clade would have required a larger size of osmotically inert storage polysaccharide pool.

DISCUSSION

1. The function of direct debranching enzyme in starch metabolism within eukaryotes

When we initially tackled the mechanism of starch biosynthesis in cyanobacteria we assumed that the enzyme that was known to be required for amylopectin crystallization in Chloroplastida had been transmitted by the cyanobiont through plastid endosymbiosis by Endosymbiotic Gene Transfer (EGT) (for review Elias and Arechibald 2009) a process by which genes from the evolving plastid were transmitted to the nucleus of the host and subsequently lost from the cyanobiont genome. The finding of starch-like polymers in cyanobacteria (Nakamura *et al.,* 2005) together with that of cyanobacterial GBSS implied that the donor of plastid endosymbiosis was related to extant diazotrophic unicellular cyanobacteria that synthesize such polymers (Deschamps *et al.,* 2008). This finding was also compatible with the most recent phylogenomic studies demonstrating that such cyanobacteria together with filamentous diazotrophic cyanobacteria defined the extant groups whose ancestors were the most likely donors for the plastid (Dagan *et al.,* 2013, Desuch *et al.,* 2008, Deschamps *et al.,* 2008, Nakamura *et al.,* 2005, Scheengurt *et al.,* 1994, Gupta et Matthews 2010). Hence a simple and therefore satisfying model could be proposed whereby an "isoamylase" like gene responsible for amylopectin crystallization had evolved in cyanobacteria and was transmitted to the Archaeplastida common ancestor thereby achieving starch rather than glycogen synthesis in the host cytosol. As is often the case such a simple and satisfying model proved wrong in face of new data. First it became increasingly clear and is now beyond doubt that isoamylase is not an enzyme of cyanobacterial affiliation but rather an enzyme of chlamydial origin (Moustafa *et al.,* 2008, Ball *et al.,* 2010 and Ball *et al.,* 2013). Second we have no reason to believe that chlamydia differ from other bacteria in the enzymatic substrate specificity of their GlgX enzyme (although this point remains to be proven through characterization of the recombinant enzyme from Chlamydiales). Hence it appears that the gene coding this enzyme was recruited for amylopectin synthesis in Archaeplastida through mutation of its substrate preferences (James *et al.,* 1995, Mouille *et al.,* 1996, Ball *et al.,* 1996, Wattebled *et al.,* 2005 and Zeeman *et al.,* 2010). In face of the phylogeny that we propose in this thesis we believe that acquisition of this gene by the common archaeplastida ancestor defined possibly a unique event. GlgX from chlamydia if analogous in function to the *E.coli* enzyme is clearly redundant with iDBE of eukaryotes (indirect debranching enzyme) as both enzymes have been proven to be required for degradation of the short external chains left by phosphorylase-mediated glycogen degradation (Jeanningros *et al.,* 1976, Dauvillée *et al.,* 2005, Tabata and Hizuriki 1971, Teste *et al.,* 2000).

The phylogeny that we report in this thesis for iDBE (see annex1) is consistent with vertical inheritance in *Cyanophora paradoxa* of the enzyme that was present in the cytosolic glycogen metabolism enzyme network of the host (Ball *et al.,* 2010). We can speculate therefore that the EGT of the chlamydial gene was selected because of a novel function assumed by its protein product (Ball *et al.,* 2013). This function could possibly be defined by emergence of an isoamylase like enzyme. Biochemical evidence for a large size (multimeric) isoamylase has been published for *Cyanophora paradoxa* (Plancke *et al.,* 2008). The direct involvement of such enzymes in starch aggregation in glaucophytes and red algae remains however to be proven by functional analyses. This ancestral isoamylase enzyme may not necessarily be directly related to the green alga isa1 subunit that was demonstrated to be sufficient for amylopectin crystallization in maize endosperm (James *et al.,* 1996, Kubo *et al.,* 2010). This apparent contradiction stems from the fact that the green alga and plant pathway was reconstructed in plastids through the use of genes encoding the enzymes of the ancestral cytosolic starch metabolism (both events are thus not truly "independent"). Hence Archaeplastida may have evolved starch metabolism twice: once in the common ancestor and a second time in the chloroplast of the evolving green lineage during the step-wise reconstruction of the pathway in plastids (Deschamps *et al,.* 2008a and Deschamps *et al,.* 2008b). Analysis of the isoamylase-GlgX gene distribution (Table 2) shows that presently Archaeplastida are the only clades containing such enzymes within eukaryotes (Curtis *et al.,* 2012, Coppin *et al.,* 2005). Hence starch metabolism in cryptophytes and alveolates does not seem to require a direct debranching enzyme or at least not a GlgX derived enzyme. This observation seems at first glance to question the universality of the glucan trimming model as proposed by Ball *et al.,* in 1996.

Enzymes	CAZy family	Cyanobacteria accumulating starch	Chloroplastida	Rhodophyta	Glaucophyta	Cryptophtes	Ciliates	Apicomplexa accumulating starch
Cyclo Maltodextrinase	GH13-20	Y/N	Y/N	Y/N	N	Y	N	N
Pullulanase	GH13	Y	Y	Y	N	N	N	N
Isoamylases	GH13	Y/N	Y	Y	Y	N	N	N
Archeal DBE	?	N	N	Y/N	N	N	N	N
Indirect-DE	GH13	N	N	N	Y	N	Y	Y
Putative amylopullulanase	GH57	Y	N	N	N	N	N	N
Glucosyl hydrolase	GH57	Y	N	N	N	N	N	N

Table 2: Gene distribution of candidate enzyme sequences possibly selective for hydrolysis of the α-1,6 branch, according to the CAZy classification**. Y** indicates the presence of this gene in organism, **N** indicates absence of the gene, and **Y/N** indicates that this gene is not distributed in all organisms of this clade.

2. The function of direct debranching enzyme in starch metabolism within cyanobacteria

We have proven that in particular strain of the genus "*Cyanobacterium*" aggregation of starch granules depends on the presence of an "isoamylase" whose substrate and product preferences ressembles that found for the isa1 subunit of the green plants and algae (Hussain *et al.,* 2003). Because Archaeplastida and cyanobacteria have evolved their respective isoamylase from different GlgX-like proteins, this defines a most striking case of convergent evolution. From two very different glycogen metabolism networks (those of eukaryotes and cyanobacteria) nature has recruited a similar (yet phylogenetically unrelated) enzyme and has converged to yield starch through very similar physico-chemical mechanisms. This remarkable conclusion considerably strengthens our initial proposal that polysaccharide debranching is indeed mandatory to obtain aggregation of starch granules from branched polymers (Ball *et al.,* 1996). However as was the case for eukaryotes this does not seem to always require in cyanobacteria the presence of GlgX-like GH13 subfamily 11 type of enzyme (see table 2). Interestingly, in some cases the genomes of cyanobacterial starch accumulators do not contain any annotated candidate debranching enzyme gene (Suzuki and

Colleoni 2012). Once again the gene distribution seems to question the universality of the glucan trimming model while the observed convergent evolution on the other hand seems to confirm it in a quite spectacular fashion.

3. Resolving apparent contradictions: a problem of structural biology

One of the most exciting perspectives stemming from our work is defined by the comparative biochemistry of starch metabolism in cyanobacteria and Archaeplastida. A 3D structure has just been produced for the Chlamydomonas isa1 subunit which conveys a possible explanation for the role of this enzyme in amylopectin crystallization (Sim and Palcic personal communication). Establishing the 3D structure of the enzyme from Cyanobacterium offers a unique opportunity to define the most important common features in the 3D organization of such enzymes. From our and others biochemical characterizations (Dauvillée *et al.,* 2001b Kubo *et al.,* 2010), we predict two major features to be required for amylopectin crystallization. One of them is clearly related to substrate preference which seems to match the CL distribution of the polysaccharide produced *in vitro* by the corresponding branching enzymes, the second, we believe, is related to the organization in multimeric structure. The latter may reveal a particular spacing of the active sites of the enzyme facilitating the biogenesis of chain clusters whose spacing is compatible with polysaccharide crystallization.
Finally the apparent absence of debranching enzyme, for instance in cryptophyte algae (as well as in alveolates and specific cyanobacteria), begs the question of how starch is synthesized in these organisms. My laboratory intends to pursue its studies with *Guillardia theta*. It is highly suspected that in such biochemical networks other (possibly GH13) CAZy hydrolases replace the GlgX derived direct DBEs in its function (Table 1) (Annex 2) (Curtis *et al.,* 2012 and Deschamps *et al.,* 2006). Once again it will be interesting to prove that this is the case (or not) and to purify the corresponding enzyme and to investigate its function (if any) in polysaccharide crystallization. If required the 3D organization of such an enzyme will also have to be investigated. Only then will one be able to claim that polysaccharide debranching is indeed a universal and mandatory mechanism for amylopectin synthesis.

CONCLUSIONS AND PERSPECTIVES

Conclusions :

During my thesis ,I demonstrated that a cation dependent hydrolase is responsible for synthesis of amylopectin in *Cyanobacterium* sp. CLg1 by polysaccharide debranching, in a similar way to plant isoamylases.

Interestingly, this cation dependent hydrolase belongs to the family GH13 subfamily 11 like plants isoamylases, but is phylogenetically related to the classical GlgX type, which has been involved in the glycogen catabolism pathway of bacteria and belongs to the same CAZy family and subfamily. Our results show that starch metabolism evolved independently at least twice during evolution ; in Chloroplastida and Cyanobacteria. It is possible that starch accumulation appeared several times, since GlgX2 gene is not present in some cyanobacteria accumulating starch. However, if plant isoamylase and this cation dependant debranching enzyme are polyphyletic, they evolved independently to acquire the same substrate specificity. Thereby all together this data suggest that the debranching enzyme specificity, is a universal mechanism required for the conversion of glycogen to starch metabolism in living cells.

However the absence of dikinase-like enzyme in cyanobaceria, and the difference of structure between semi amylopectin and true starch, could indicate physiological differences in the function and availability of this polysaccharide and, probably, between the polysaccharide crystallization of cyanobacteria and Chloroplastida, which until now have not been investigated.

Perspectives:

This work based on a selection of mutants obtained by UV mutagenesis of a cyanobacterium strain, was focused on the characterization of one class of mutants, leaving to three major perspectives :

- This study is the first functional approach of enzymes responsibles for starch metabolism in an organism other than Chloroplastida, and if until now, all starch metabolism studies involved direct debranching enzyme to make crystallin starch a large scale functional approach have to be perform in several lineage, including others Archeaplastida, the secondary plastid endosymbiosis derivatives, and some

particularly other cyanobacteria without GH13 subfamily 11 referees. Another intriguing part of these enzymes is the convergent evolution of the substrate specificity which could be observed in all organisms, in addition enzyme presenting this specificity seems to possess a quaternary structures, which may suggest the importance of a multimeric structure in the acquisition of this enzyme specificity.

- Our work was focused exclusively on a group of mutants belonging to the Class A which are characterized by an increase of WSP and a strong decrease of starch granules synthesis). Another group belonging to the class A harbors an increase of WSP level and an intermediate quantity of starch synthesis. No mutation has been identified in these mutants, named X' mutants, in all genes involved in storage metabolism pathway including GlgX2 gene. However, zymogram analysis performed on crude extracts reveals a decrease of the cation dependant enzyme activity or GlgX2. A mutation in the promoter or in a gene involved in the regulation of GlgX2 could emphasize a regulation pathway. The latter could, in addition, indicate a control by the circadian clock, and help our understanding on the surprising activity towards amylopectin, which is theoritically the ultimate product of this enzyme. In any case the characterization of insoluble polysaccharide and soluble polysaccharide of these mutants -which may indicate a polysaccharide product intermediate between insoluble amylopectin and hydrosoluble glycogen- in order to help the synthesize of *in vitro* starch.
- As mentioned previously, our work was focused on class A mutants, but some others interesting *Cyanobacterium* sp. Clg1 mutants was found as well. Particullarly I have participated to the characterization of a mutant impaired in the synthase activity which leads to a huge decrease of insoluble polyaccharide and no significant modification in the amount of water soluble polysaccharide. The characterization of this mutant could help us to understand the physiological function of both soluble starch/glycogen starch in the biosynthesis of starch granules.

BIBLIOGRAPHY

Abascal, F., Zardoya, R., Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**, 2104-2105.

Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol. Sep-Oct;**52**(5):399-451.

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S,Shadwick RS, Schoch CL, Smirnov A, Spiegel FW. (2012) The revised classification of eukaryotes. J Eukaryot Microbiol. Sep;**59**(5):429-93.

Andersson, S. G. and C. G. Kurland (1999). Origins of mitochondria and hydrogenosomes. Curr Opin Microbiol **2**(5): 535-541.

Arias, M. C. *et al.,* (2012) Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants. Mobile genetic elements **2**, 81-8

Aoki, S., T. Kondo and M. Ishiura (2002). "A promoter-trap vector for clock-controlled genes in the cyanobacterium Synechocystis sp. PCC 6803." J Microbiol Methods **49**(3): 265-274.

Ball SG, Morell MK. (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu Rev Plant Biol.54:207-3

Ball S, Guan HP, James M, Myers A, Keeling P, Mouille G, Buléon A, Colonna P, Preiss J. (1996). From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell. 86(3):349-52

Ball, S. *et al.,* (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J. Exp. Bot. **62**, 1775–180

Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P. M., Gehre, L., Colleoni, C., Arias, M. C., Cenci, U., and Dauvillée, D. (2013). Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? Plant Cell **25**, 7-21

Ball, S., van de Wal, M. and Visser, R. (1998) Progress in understanding the biosynthesisof amylose. Trends in Plant Sci. **3**, 462-46

Ball, S., H. P. Guan, M. James, A. Myers, P. Keeling, G. Mouille, A. Buléon, P. Colonna and J. Preiss (1996). From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell **86**(3): 349-352.

Baum, D. (2013). The Origin of Primary Plastids: A Pas de Deux or a Ménage à Trois? Plant Cell **25**, 4-5

Baunsgaard, L. *et al.,* (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated alpha-glucans and is involved in starch degradation in Arabidopsis. Plant J. **41**, 595–605

Becker, B., Hoef-Emden, K. and Melkonian M. (2008) Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. BMC Evolutionary Biology **8**, 20.

Bertoft, E., A. Kallman, K. Koch, R. Andersson and P. Aman (2011). "The cluster structure of barley amylopectins of different genetic backgrounds." Int J Biol Macromol **49**(4): 441-453.

Bhattacharya, D., D. C. Price, H. S. Yoon, E. C. Yang, N. J. Poulton, R. A. Andersen and S. P. Das (2012). "Single cell genome analysis supports a link between phagotrophy and primary plastid endosymbiosis." Sci Rep **2**: 356.

Bhattacharya, D. *et al.,* (2013) Genome of the red alga Porphyridium cruentum has compact structure and shows extensive gene sharing. Nature Communications **4**:1941

Bischof, S. *et al.,* (2013) *Cecropia peltata* accumulates starch or soluble glycogen by differentially regulating starch biosynthetic genes. Plant Cell DOI 10.1105/tpc.113.10979

Buleon, A., Colonna, P., Planchot, V., and Ball, S. (1998). Starch granules: structure and biosynthesis. Int J Biol Macromol **23,** 85-112.

Buleon, A., D. J. Gallant, B. Bouchet, G. Mouille, C. D'Hulst, J. Kossmann and S. Ball (1997). "Starches from A to C. Chlamydomonas reinhardtii as a model microbial system to investigate the biosynthesis of the plant amylopectin crystal." Plant Physiol **115**(3): 949-957.

Burki, F., K. Shalchian-Tabrizi, M. Minge, A. Skjaeveland, S. I. Nikolaev, K. S. Jakobsen and J. Pawlowski (2007). "Phylogenomics reshuffles the eukaryotic supergroups." PLoS One **2**(8): e790.

Burki, F., N. Okamoto, J. F. Pombert and P. J. Keeling (2012). "The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins." Proc Biol Sci **279**(1736): 2246-2254.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **37** D233-8. doi:10.1093/nar/gkn663.

Carciofi, M., A. Blennow, S. L. Jensen, S. S. Shaik, A. Henriksen, A. Buleon, P. B. Holm and K. H. Hebelstrup (2012). "Concerted suppression of all starch branching enzyme genes in barley produces amylose-only starch granules." BMC Plant Biol **12**: 223.

Carmody, R. N. and R. W. Wrangham (2009). "The energetic significance of cooking." J Hum Evol **57**(4): 379-391.

Cavalier-Smith, T. (1981). "Eukaryote kingdoms: seven or nine?" Biosystems 14(3-4): 461-481.

Cavalier-Smith, T. (2004). "Only six kingdoms of life." Proc Biol Sci **271**(1545): 1251-1262.

Chan, E.M., Young, E.J., Ianzano, L., Munteanu, I., Zhao, X., Christopoulos, C.C., Avanzini, G., Elia, M., Ackerley, C.A., Jovic, N.J. *et al.,* (2003). Mutations in NHLRC cause progressive myoclonus epilepsy. Nat Genet 125-127.

Chanzy, H., J. L. Putaux, D. Dupeyre, R. Davies, M. Burghammer, S. Montanari and C. Riekel (2006). "Morphological and structural aspects of the giant starch granules from Phajus grandifolius." J Struct Biol **154**(1): 100-110.

Charvet J-P., Bart F., Diry J-P., Hinnewinkel J-C, Maillard J-C, Poulot M. (2008) Produire pour nourrir les hommes SEDES, Géographie, Broché, ISBN 10 : 2301000362, ISBN 13 : 9782301000361

Christiansen C, Abou Hachem M, Janecek S, Viksø-Nielsen A, Blennow A, Svensson B. (2009). The carbohydrate-binding module family 20--diversity, structure, and function. FEBS J. 276(18):5006-29

Collén J. *et al.,* (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. Proc. Natl. Acad. Sci. USA **110**, 5247-5252

Colleoni, C., and Suzuki, E. (2012). Storage polysaccharide metabolism in Cyanobacteria. Essential reviews in experimental biology: Starch: Origins, Structure and metabolism *(Essential Reviews in Experimental Biology)* (Tetlow, I, ed.) **volume 5,** chapter 5 pp. 217-253, SEB.

Colleoni, C., M. Linka, P. Deschamps, M. G. Handford, P. Dupree, A. P. Weber and S. G. Ball (2010). "Phylogenetic and biochemical evidence supports the recruitment of an ADP-glucose translocator for the export of photosynthate during plastid endosymbiosis." Mol Biol Evol **27**(12): 2691-2701.

Colleoni, C., Dauvillée, D., Mouille, G., Buléon, A., Gallant, D., Bouchet, B., Morell, M., Samuel, M., Delrue, B., d'Hulst, C., Bliard, C., Nuzillard, J.M., and Ball, S. (1999). Genetic and biochemical evidence for the involvement of alpha-1,4 glucanotransferases in amylopectin synthesis. Plant Physiol **120,** 993-1004.

Comparat-Moss *et al.* (2010) A putative phosphatase, LSF1, is required for normal starch turnover in Arabidopsis leaves. Plant Physiol. **152**, 685-697

Coppin, A., Varre, J. S., Lienard, L., Dauvillee, D., Guerardel, Y., Soyer-Gobillard, M. O., Buléon, A., Ball, S., and Tomavo, S. (2005) Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. J Mol Evol **60**, 257-267

Curtis BA *et al.*. 2012 Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature. Dec 6;**492**(7427):59-65.

Dagan, T., M. Roettger, K. Stucken, G. Landan, R. Koch, P. Major, S. B. Gould, V. V. Goremykin, R. Rippka, N. Tandeau de Marsac, M. Gugger, P. J. Lockhart, J. F. Allen, I. Brune, I. Maus, A. Puhler and W. F. Martin (2013). Genomes of Stigonematalean

cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol Evol **5**(1): 31-44.

Dauvillée D, Deschamps P, Ral JP, Plancke C, Putaux JL, Devassine J, Durand-Terrasson A, Devin A, Ball SG (2009) Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate *Crypthecodinium cohnii*. Proc. Natl. Acad. Sci. USA **106**: 21126-21130.

Dauvillée, D. *et al.* (1999) Novel starch-like polysaccharides are synthesized by a soluble form of granule-bound starch synthase in glycogen accumulating mutants of *Chlamydomonas reinhardtii.* Plant Physiol. **119**, 321-330

Dauvillée, D., Mestre, V.V., Colleoni, C., Slomianny, M., Mouille, G., Delrue, B., d'Hulst, C., Bliard, C., Nuzillard, J., and Ball, S. (2000). "The debranching enzyme complex missing in glycogen accumulating mutants of Chlamydomonas reinhardtii displays an isoamylase-type specificity." Plant Sci. **157,** 145-156.

Dauvillée, D., Colleoni, C., Mouille, G., Buleon, A., Gallant, D.J., Bouchet, B., Morell, M.K., d'Hulst, C., Myers, A.M., and Ball, S.G. (2001a). Two loci control phytoglycogen production in the monocellular green alga *Chlamydomonas reinhardtii*. Plant Physiol **125,** 1710-1722.

Dauvillée, D., Colleoni, C., Mouille, G., Morell, M.K., d'Hulst, C., Wattebled, F., Lienard, L., Delvalle, D., Ral, J.P., Myers, A.M., and Ball, S.G. (2001b). Biochemical characterization of wild-type and mutant isoamylases of Chlamydomonas reinhardtii supports a function of the multimeric enzyme organization in amylopectin maturation. Plant Physiol **125,** 1723-1731.

Dauvillée, D., Kinderf, I.S., Li, Z., Kosar-Hashemi, B., Samuel, M.S., Rampling, L., Ball, S., and Morell, M.K. (2005). Role of the *Escherichia coli* glgX gene in glycogen metabolism. Journal of bacteriology **187,** 1465-1473.

Dauvillée, D., Mestre, V.V., Colleoni, C., Slomianny, M., Mouille, G., Delrue, B., d'Hulst, C., Bliard, C., Nuzillard, J., and Ball, S. (2000). The debranching enzyme complex missing in glycogen accumulating mutants of *Chlamydomonas reinhardtii* displays an isoamylase-type specificity. Plant Sci **157,** 145-156.

Delatte, T. *et al.,* (2006) Evidence for distinct mechanisms of starch granule breakdown in plants. J. Biol. Chem. **285**, 12050-1205

Delrue, B., Fontaine, T., Routier, F., Decq, A., Wieruszeski, J.M., Van Den Koornhuyse, N., Maddelein, M.L., Fournet, B., and Ball, S. (1992). Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin. J Bacteriol **174,** 3612-3620.

de Loura, I. C., J. P. Dubacq and J. C. Thomas (1987). "The effects of nitrogen deficiency on pigments and lipids of cyanobacteria." Plant Physiol **83**(4): 838-843.

Deschamps, P., Haferkamp, I., Dauvillée, D., Haebel, S., Steup, M., Buléon, A., Putaux, J. L., d'Hulst, C., Gould, S., Maier, U., Neuhaus, E. and Ball S. (2006) The nature of the periplastidial pathway of amylose synthesis in the cryptophyte *Guillardia theta*. Eukaryot Cell **5**, 954-963.

Deschamps, P. *et al.,* (2008a) Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. Genetics **178**, 2373-2386

Deschamps, P. *et al.,* (2008b) The relocation of starch metabolism to chloroplasts: when, why and how. Trends in Plant Sci. **13**, 574-585

Deschamps, P., Colleoni, C., Nakamura, Y., Suzuki, E., Putaux, J.L., Buleon, A., Haebel, S., Ritte, G., Steup, M., Falcon, L.I., Moreira, D., Loffelhardt, W., Raj, J.N., Plancke, C., d'Hulst, C., Dauvillee, D., and Ball, S. (2008c). Metabolic symbiosis and the birth of the plant kingdom. Mol Biol Evol **25,** 536-548.

Deschamps P, Moreira D. (2012). Reevaluating the green contribution to diatom genomes. Genome Biol Evol.4(7):683-8.

Deusch, O., G. Landan, M. Roettger, N. Gruenheit, K. V. Kowallik, J. F. Allen, W. Martin and T. Dagan (2008). Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Mol Biol Evol **25**(4): 748-761.

Dorrell, R. G. and A. G. Smith (2011). Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. Eukaryot Cell. **10:** 856-868.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5):1792-7.

Elias, M. and J. M. Archibald (2009). Sizing up the genomic footprint of endosymbiosis. Bioessays **31**(12): 1273-1279.

Falcon, L.I., Lindvall, S., Bauer, K., Bergman, B., and Carpenter, E.J. (2004). Ultrastructure of unicellular N2 fixing cyanobacteria from the tropical north atlantic and subtropical north pacific oceans. J phycol **40,** 1074-1078.

Falcon, L. I., S. Magallon and A. Castillo (2010). "Dating the cyanobacterial ancestor of the chloroplast." ISME J **4**(6): 777-783.

Fordham-Skelton, A.P. *et al.,* (2002) A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. Plant J. **29**, 705-712

Fu, J., and Xu, X. (2006). The functional divergence of two *glgP* homologues in *Synechocystis* sp. PCC 6803. FEMS microbiology letters **260,** 201-209.

Gallant, D. J., B. Bouchet and P. M. Baldwin (1997). "Microscopy of starch: Evidence of a new level of granule organization." Carbohydrate Polymers **32**(3-4): 177-191.

Gentry, M.S. *et al.,* (2007) The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. J. Cell Biol. **178**, 477-484

Gentry, M.S. *et al.,* (2009) Lafora disease: Insights into neurodegeneration from plant metabolism. Trends Biochem. Sci. **34**, 628-635

Gentry, M.S. *et al.,* (2013) Laforin, a protein with many faces: glucan phosphatase, adapter protein, et alii. FEBS J. **280**, 526-534

Gibbs, S. P. (1981). The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann N Y Acad Sci **361**: 193-208.

Glaring, M.A. *et al.,* (2011) Starch-binding domains in the CBMfamily – low affinity domains from glucan, water dikinase and α-amylase involved in plastidial starch metabolism. FEBS J. **278**, 1175-1187

Golden, S.S. (1988). Mutagenesis of cyanobacteria by classical and gene-transfer-based methods. Methods Enzymol **167,** 714-727.

Goldsmith, E., S. Sprang and R. Fletterick (1982). "Structure of maltoheptaose by difference Fourier methods and a model for glycogen." J Mol Biol **156**(2): 411-427.

Gould SB, Waller RF, McFadden GI (2008). Plastid evolution. Annu Rev Plant Biol. **59**:491-517.

Grimaud F, Lancelon-Pin C, Rolland-Sabaté A, Roussel X, Laguerre S, Viksø-Nielsen A, Putaux JL, Guilois S, Buléon A, D'Hulst C, Potocki-Véronèse G (2013). In vitro synthesis of hyperbranched α-glucans using a biomimetic enzymatic toolbox. Biomacromolecules. Feb 11; **14**(2):438-47

Guan H, Li P, Imparl-Radosevich J, Preiss J, Keeling P. (1997) Comparing the properties of Escherichia coli branching enzyme and maize branching enzyme. Arch Biochem Biophys. **342**(1):92-8

Gupta R. S. and Mathews D. W. (2010). Signature proteins for the major clades of Cyanobacteria. BMC Evol Biol **10**: 24.

Gurcel, C., Vercoutter-Edouart, A.S., Fonbonne, C., Mortuaire, M., Salvador, A., Michalski, J.C., and Lemoine, J. (2008). Identification of new O-GlcNAc modified proteinsusing a click-chemistry-based tagging. Anal Bioanal Chem **390,** 2089-2097.

Hackett, J. D., H. S. Yoon, S. Li, A. Reyes-Prieto, S. E. Rummele and D. Bhattacharya (2007). "Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates." Mol Biol Evol **24**(8): 1702-1713.

Hanashiro, I., M. Tagawa, S. Shibahara, K. Iwata and Y. Takeda (2002). "Examination of molar-based distribution of A, B and C chains of amylopectin by fluorescent labeling with 2-aminopyridine." Carbohydr Res **337**(13): 1211-1215.

Hejazi *et al.,* (2009) The two plastidial starch-related dikinases sequentially phosphorylate glucosyl residues at the surface of both the A- and the B-allomorph of crystallized maltodextrins but the mode of action differs. Plant Physiol. **150**, 962-97

Hejazi M. *et al.,* (2008) Glucan, water dikinase phosphorylates crystalline maltodextrins and thereby initiates solubilization. Plant J. **55**: 323–33

Hejazi M. *et al.,* (2010) The laforin-like dual-specificity phosphatase SEX from Arabidopsis hydrolyzes both C6- and C3-phosphate esters introduced by starch-related dikinases and thereby affects phase transition of α-glucans. Plant Physiol. **152**, 711-721

Hejazi M. *et al.,* (2012a) Starch phosphorylation and dephosphorylation. In: Starch: Origins, Structure and Metabolism. Essential Reviews in Experimental Biology. **Vol. 5** (Tetlow, I, ed.), pp. 279-310, SEB

Hejazi M. *et al.,* (2012b) The plastidial glucan, water dikinase (GWD) catalyses multiple phosphotransfer reactions. FEBS J. **279**, 1953-1966

Hirabaru, C., A. Izumo, S. Fujiwara, Y. Tadokoro, T. Shimonaga, M. Konishi, M. Yoshida, N. Fujita, Y. Nakamura, M. Yoshida, T. Kuroiwa and M. Tsuzuki (2010). "The primitive rhodophyte Cyanidioschyzon merolae contains a semiamylopectin-type, but not an amylose-type, alpha-glucan." Plant Cell Physiol **51**(5): 682-693.

Hirose T. *et al.* (2013) Disruption of a rice gene for α-glucan water dikinase, OsGWD1, leads to hyperaccumulation of starch in leaves but exhibits limited effects on growth. *Frontiers in* Plant Sci. **4**, 147

Hizukuri S. *et al.,* (1970) Studies on starch phosphate. 1. Estimation of glucose 6-phosphate residues on starch and the presence of other bound phosphate(s). Starch/Stärke **45**, 417-42

Holt J.G. et al.,(1994). « Bergey's manual of determinative bacteriology : 9th edition » ISBN 978-0-683-00603-2

Honda, D., A. Yokota and J. Sugiyama (1999). "Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine Synechococcus strains." J Mol Evol **48**(6): 723-739.

Huang J. and Gogarten J. P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol **8**(6): R99.

Hussain, H., Mant, A., Seale, R., Zeeman, S., Hinchliffe, E., Edwards, A., Hylton, C., Bornemann, S., Smith, A.M., Martin, C., and Bustos, R. (2003). Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. Plant Cell **15,** 133-149.

Imberty, A. and S. Perez (1988). "A revisit to the 3-dimensional structure of B-type starch." Biopolymers **27**(8): 1205-1221.

Imberty A, Chanzy H, Pérez S, Buléon A, Tran V. (1988) The double-helical nature of the crystalline part of A-starch. J Mol Biol. May **20**;201(2):365-78

James M. G, Robertson, D. S. and Myers, A. M. (1995) Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell **7**, 417-429

James M.G., Robertson, D.S., and Myers, A.M. (1995). Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell **7,** 417-429.

Janecek S., Svensson, B., and MacGregor, E.A. (2011). Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme Microb Technol **49,** 429-440.

Jarvis, P. and J. Soll (2001). "Toc, Tic, and chloroplast protein import." Biochim Biophys Acta **1541**(1-2): 64-79.

Jeanningros R., Creuzet-Sigal, N., Frixon, C., and Cattaneo, J. (1976). Purification and properties of a debranching enzyme from *Escherichia coli*. Biochim Biophys Acta **438,** 186-199.

Johnson CH, Golden SS, Kondo T. Adaptive significance of circadian programs in cyanobacteria. Trends Microbiol. 1998 Oct;**6**(10):407-10.

Katsuya Y, Mezaki Y, Kubota M, Matsuura Y. (1998) Three-dimensional structure of Pseudomonas isoamylase at 2.2 A resolution. J Mol Biol. **281**(5):885-97

Keeling, P. J. (2004). "Diversity and evolutionary history of plastids and their hosts." Am J Bot **91**(10): 1481-1493.

Keeling, P. J. (2013). "The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution." Annu Rev Plant Biol.

Kerk D. *et al.,* (2006) A chloroplast-localized dual-specificity protein phosphatase in *Arabidopsis* contains a phylogenetically dispersed and ancient carbohydrate-binding domain, which binds the polysaccharide starch. Plant J. **46**, 400-411

Kneip C., Lockhart P., Voss C. and Maier U. G. (2007). Nitrogen fixation in eukaryotes--new models for symbiosis. BMC Evol Biol **7**: 55.

Kötting O. *et al.,* (2005) Identification of a novel enzyme required for starch metabolism in *Arabidopsis* leaves. The phosphoglucan, water dikinase. Plant Physiol. **137**, 242–255

Kötting O. *et al.,* (2009) STARCH-EXCESS is a laforin-like phosphoglucan phosphatase required for starch degradation in *Arabidopsis thaliana*. Plant Cell **21**, 334-342

Kubo A., Colleoni C., Dinges J.R., Lin Q., Lappe R.R., Rivenbark J.G., Meyer A.J., Ball S.G., James M.G., Hennen-Bierwagen T.A., and Myers A.M. (2010). Functions of heteromeric and homomeric isoamylase-type starch-debranching enzymes in developing maize endosperm. Plant Physiol **153,** 956-969.

Kubo A., Fujita N., Harada K., Matsuda T., Satoh H., and Nakamura Y. (1999) The starch-debranching enzymes isoamylase and pullulanase are both involved in amylopectin biosynthesis in rice endosperm. Plant Physiol **121,** 399–409.

Le S.Q., Gascuel O. (2008) LG An improved, general amino-acid replacement matrix. Mol. Biol. Evol. **25**, 1307-1320, doi:10.1093/molbev/msn067

Liu, Y., S. S. Golden, T. Kondo, M. Ishiura and C. H. Johnson (1995). "Bacterial luciferase as a reporter of circadian gene expression in cyanobacteria." J Bacteriol **177**(8): 2080-2086.

Lomako J, Lomako WM, Whelan WJ. (2004) Glycogenin: the primer for mammalian and yeast glycogen synthesis Biochim Biophys Acta. **1673**(1-2):45-55

Lorberth R. *et al.,* (1998) Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. Nat. Biotechnol. 16, 473–477

Lucia A, Ruiz JR, Santalla A, Nogales-Gadea G, Rubio JC, García-Consuegra I,Cabello A, Pérez M, Teijeira S, Vieitez I, Navarro C, Arenas J, Martin MA, AndreuAL. (2012) Genotypic and phenotypic features of McArdle disease: insights from the Spanish national registry. J Neurol Neurosurg Psychiatry. Mar;**83**(3):322-8.doi: 10.1136/jnnp-2011-301593.

Madsen, N. B. and C. F. Cori (1958). "The binding of glycogen and phosphorylase." J Biol Chem **233**(6): 1251-1256.

Marshall J.J. and Rickson F.R. (1973) Characterization of the α-D-glucan from the plastids of *Cecropia peltata* as a glycogen-type polysaccharide. Carbohydrate Res. **28**, 31-39

Martin, W. and M. Muller (1998). "The hydrogen hypothesis for the first eukaryote." Nature **392**(6671): 37-41.

Meekins D.A. *et al.* (2013) Structure of the Arabidopsis glucan phosphatase LIKE SEX FOUR2 reveals a unique mechanism for starch dephosphorylation. Plant Cell **25**, 2302-2314

Meléndez-Hevia E, Waddell T G and Shelton E D (1993). Optimization of molecular design in the evolution of metabolism: the glycogen molecule. Biochem. J. **295** (477–483).

Melendez R., Melendez-Hevia, E., and Canela, E.I. (1999). The fractal structure of glycogen: A clever solution to optimize cell metabolism. Biophys J **77,** 1327-1332.

Mereschkowsky, C. (1905). Uber Natur und Ursprung der Chromatophoren im Panzenreiche. Biol. Centralbl. **25**: 593±604

Minassian B.A., Lee J.R., Herbrick J.A., Huizenga J., Soder S., Mungall A.J., Dunham I., Gardner R., Fong C.Y., Carpenter S. *et al.,* (1998). Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nat Genet 171-174.

Mohr, W., M. P. Intermaggio and J. LaRoche (2010). "Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium Crocosphaera watsonii WH8501." Environ Microbiol **12**(2): 412-421.

Moreira, D., H. Le Guyader and H. Philippe (2000). "The origin of red algae and the evolution of chloroplasts." Nature **405**(6782): 69-72.

Mouille G., Maddelein M.L., Libessart N., Talaga P., Decq A., Delrue B., and Ball S. (1996). Preamylopectin Processing: A Mandatory Step for Starch Biosynthesis in Plants. Plant Cell **8,** 1353-1366.

Moustafa A., Reyes-Prieto A. and Bhattacharya D. (2008). Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One **3**(5): e2205.

Moustafa A., Beszteri B., Maier U. G., Bowler C., Valentin K. and Bhattacharya D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science **324**(5935): 1724-1726.

Myers A. M. *et al.,* 2000. Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol. **122**, 989-99

Nakamura Y. *et al.,* (1997) Correlation between activities of starch debranching enzymes and α-polyglucan structure in endosperms of sugary- mutants of rice. Plant J. **12**, 143-152

Nakamura Y., Takahashi J., Sakurai A., Inaba Y., Suzuki E., Nihei S., Fujiwara S., Tsuzuki M., Miyashita H., Ikemoto H., Kawachi M., Sekiguchi H., and Kurano N. (2005). Some Cyanobacteria synthesize semi-amylopectin type alpha-polyglucans instead of glycogen. Plant Cell Physiol **46,** 539-545.

Nakamura Y, Utsumi Y, Sawada T, Aihara S, Utsumi C, Yoshida M, Kitamura S. (2010) Characterization of the reactions of starch branching enzymes from rice endosperm Plant Cell Physiol. 51(5):776-94

Nakamura Y, Ono M, Utsumi C, Steup M. (2012) Functional interaction between plastidial starch phosphorylase and starch branching enzymes from rice during the synthesis of branched maltodextrins. Plant Cell Physiol.**53**(5):869-78

Nielsen T.H. *et al.,* (1994) Starch phosphorylation in potato tubers proceeds concurrently with de novo biosynthesis of starch. Plant Physiol. **105**, 111-118

Niittylä T. *et al.,* (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. J. Biol. Chem. **281**, 11815-1181

Nirmal Raj J (2010) Selection of mutants defective for starch biosynthesis in marine single-cell cyanobacteria. http://www.theses.fr/149348339

Nitschke F. *et al.,* (2013) Hyperphosphorylation of glucosyl C carbons and altered structure of glycogen in the neurodegenerative epilepsy Lafora disease. Cell Metabolism **17**, 756-764

Patron N. J. and Keeling P. K. (2005) Common evolutionary origin of starch biosynthetic enzymes in green and red algae. J. Phycol. **41**, 1131-1147

Patron, N. J., Y. Inagaki and P. J. Keeling (2007). "Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages." Curr Biol **17**(10): 887-891.

Plancke C. *et al.,* (2008) Pathway of cytosolic starch synthesis in the model glaucophyte Cyanophora paradoxa. Eukaryot. Cell **7**, 247-258

Plaxton, W. C. and J. Preiss (1987). "Purification and Properties of Nonproteolytic Degraded ADPglucose Pyrophosphorylase from Maize Endosperm." Plant Physiol **83**(1): 105-112.

Preiss, J., S. G. Yung and P. A. Baecker (1983). "Regulation of bacterial glycogen synthesis." Mol Cell Biochem **57**(1): 61-80.

Price, D. C., C. X. Chan, H. S. Yoon, E. C. Yang, H. Qiu, A. P. Weber, R. Schwacke, J. Gross, N. A. Blouin, C. Lane, A. Reyes-Prieto, D. G. Durnford, J. A. Neilson, B. F. Lang, G. Burger, J. M. Steiner, W. Loffelhardt, J. E. Meuser, M. C. Posewitz, S. Ball, M. C. Arias, B. Henrissat, P. M. Coutinho, S. A. Rensing, A. Symeonidi, H. Doddapaneni, B. R. Green, V. D. Rajah, J. Boore and D. Bhattacharya (2012). "*Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants." Science **335**(6070): 843-847.

Pulido R. and Hooft van Huijsduijnen R. (2008) Protein tyrosine phosphatises: dual-specificity phosphatises in health and disease. FEBS J. **275**, 848-862

Ral J.-P. *et al.,* (2012) Down-regulation of glucan, water-dikinase activity in wheat endosperm increases vegetative biomass and yield. Plant Biotechn. J**. 10**, 871-882

Rasmussen, B., I. R. Fletcher, J. J. Brocks and M. R. Kilburn (2008). "Reassessing the first appearance of eukaryotes and cyanobacteria." Nature **455**(7216): 1101-1104.

Rickson F.R. (1971) Glycogen plastids in Müllerian Body cells of *Cecropia peltata* – a higher green plant. Science **173**, 344-34

Reddy, K. J., J. B. Haskell, D. M. Sherman and L. A. Sherman (1993). "Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece." J Bacteriol **175**(5): 1284-1292.

Reyes-Prieto A, Moustafa A, Bhattacharya D (2008). "Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic". Curr Biol **18**:956–962.

Reyes-Prieto A, Moustafa A. (2012). Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. Sci Rep. **2**:955.

Rice, D. W. and J. D. Palmer (2006). "An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters." BMC Biol **4**: 31.

Rippka R., Deruelles J., Waterbury J.B., Herdman M., and Stanier R.Y. (1979). Generic Assignments, strain histories and properties of pure cultures of cyanobacteria. Journal of General Microbiology **111,** 1-61.

Ritte G. *et al.,* (2002) The starch-related R protein is an glucan, water dikinase. Proc. Natl. Acad. Sci. **99**, 7166-7173

Ritte G. *et al.,* (2004) Phosphorylation of transitory starch is increased during degradation. Plant Physiol. **135**, 2068-207

Ritte G. *et al.,* (2006) Phosphorylation of C and C positions of glucosyl residues in starch is catalyzed by two distinct dikinases. FEBS Lett. **580**, 4872-487

Roach P.J. *et al.,* (2012) Glycogen and its metabolism: some new developments and old themes. Biochem. J. **441**, 763-787.

Rodriguez-Ezpeleta, N., H. Brinkmann, S. C. Burey, B. Roure, G. Burger, W. Loffelhardt, H. J. Bohnert, H. Philippe and B. F. Lang (2005). "Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes." Curr Biol **15**(14): 1325-1330.

Roussel X, Lancelon-Pin C, Viksø-Nielsen A, Rolland-Sabaté A, Grimaud F, Potocki-Véronèse G, Buléon A, Putaux JL, D'Hulst C (2013). Characterization of substrate and product specificity of the purified recombinant glycogen branching enzyme of Rhodothermus obamensis. Biochim Biophys Acta. Jan;**1830**(1):2167-77

Rust, M. J., S. S. Golden and E. K. O'Shea (2011). "Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator." Science **331**(6014): 220-223.

Sakai M., Austin J., Witmer F., and Trueb L. (1970). Studies in myoclonus epilepsy (Lafora body form). II. Polyglucosans in the systemic deposits of myoclonus epilepsy and in corpora amylacea. Neurology 160-176.

Santelia et *al*. (2011) the phosphoglucan phosphatase like SEXFour2 dephosphorylates starch at the C3-position in *Arabidopsis*. Plant Cell **23**, 4096-4111

Schirrmeister BE, Antonelli A, Bagheri HC (2011). The origin of multicellularity in cyanobacteria. BMC Evol Biol. Feb 14;**11**:45.

Schirrmeister BE, de Vos JM, Antonelli A, Bagheri HC (2013). "Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event". Proc Natl Acad Sci U S A. Jan 29;**110**(5):1791-6.

Schneegurt MA, Sherman DM, Nayar S, Sherman LA (1994). Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium Cyanothece sp. strain ATCC 51142. J Bacteriol. Mar;**176**(6):1586-97

Schneegurt MA, Sherman DM, Sherman LA (1997). Composition of the carbohydrate granules of the cyanobacterium, Cyanothece sp. strain ATCC 51142. Arch Microbiol. Feb-Mar;**167**(2-3):89-98

Schönknecht G. *et al.* (2013) Gene transfer from bacteria and archea facilitated evolution of an extremophilic eukaryote. Science **239**, 1207-1210

Seetharaman, K. and E. Bertoft (2012a). "Perspectives on the history of research on starch Part I: On the linkages in starch." Starch - Stärke **64**(9): 677-682.

Seetharaman, K. and E. Bertoft (2012b). "Perspectives on the history of research on starch Part II: On the discovery of the constitution of diastase." Starch - Stärke **64**(10): 765-769.

Seetharaman, K. and E. Bertoft (2012c). "Perspectives on the history of research on starch. Part III : On the era of confusion due to terminology " Starch - Stärke **64**(11): 841-845.

Seetharaman, K. and E. Bertoft (2012d). "Perspectives on the history of research on starch Part IV: On the visualization of granule architecture." Starch - Stärke **64**(12): 929-934.

Seetharaman, K. and E. Bertoft (2013a). "Perspectives on the history of research on starch Part V: On the conceptualization of amylopectin structure." Starch - Stärke **65**(1-2): 1-7.

Seetharaman, K. and E. Bertoft (2013b). "Perspectives on the history of research on starch Part VI: Postscriptum." Starch - Stärke **65**(1-2): 107-111.

Seok, Y. J., M. Sondej, P. Badawi, M. S. Lewis, M. C. Briggs, H. Jaffe and A. Peterkofsky (1997). "High affinity binding and allosteric regulation of Escherichia coli glycogen phosphorylase by the histidine phosphocarrier protein, HPr." J Biol Chem **272**(42): 26511-26521.

Shearer J., and Graham T.E. (2002). New perspectives on the storage and organization of muscle glycogen. Can J Appl Physiol **27,** 179-203.

Shimonaga T., Konishi M., Oyama Y., Fujiwara S., Satoh A., Fujita N., Colleoni C., Buleon A., Putaux J.L., Ball S.G., Yokoyama A., Hara Y., Nakamura Y., and Tsuzuki M. (2008). Variation in storage alpha-glucans of the Porphyridiales (Rhodophyta). Plant Cell Physiol **49,** 103-116.

Silver DM, Silva LP, Issakidis-Bourguet E, Glaring MA, Schriemer DC, Moorhead GB. (2013) Insight into the redox regulation of the phosphoglucan phosphatase SEX4 involved in starch degradation. FEBS J. 280(2):538-48

Smith AM, Zeeman SC, Smith SM. (2005). Starch degradation. Annu Rev Plant Biol. 56:73-98

Song H.N., Jung T.Y., Park J.T., Park B.C., Myung P.K., Boos W., Woo E.J., and Park K.H**.** (2010). Structural rationale for the short branched substrate specificity of the glycogen debranching enzyme GlgX. Proteins **78,** 1847-1855.

Stamatakis A**.** (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed model. Bioinformatics **22**, 2688-2690, doi:10.1093/bioinformatics/btl446

Summons, R. E., L. L. Jahnke, J. M. Hope and G. A. Logan (1999). "2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis." Nature **400**(6744): 554-557.

Suzuki E., Onoda M., Colleoni C., Ball S., Fujita N., and Nakamura Y. (2013). Physicochemical Variation of Cyanobacterial Starch, the Insoluble alpha-glucans in Cyanobacteria. Plant Cell Physiol. **54**:465-473

Suzuki E., Umeda K., Nihei S., Moriya K., Ohkawa H., Fujiwara S., Tsuzuki M., and Nakamura Y. (2007). Role of the GlgX protein in glycogen metabolism of the cyanobacterium, *Synechococcus elongatus* PCC 7942. Biochim Biophys Acta **1770,** 763-773.

Suzuki E. and Suzuki R. (2013). "Variation of storage polysaccharides in phototrophic microorganisms. J. appl. Glycosci. **60**, 21-27

Szollosi, G. J., Boussau B., Abby S. S., Tannier E. and Daubin V. (2012). Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. Proc Natl Acad Sci U S A. **109**: 17513-17518.

Szydlowski N, Ragel P, Hennen-Bierwagen TA, Planchot V, Myers AM, Mérida A, d'Hulst C, Wattebled F. (2011). Integrated functions among multiple starch synthases determine both amylopectin chain length and branch linkage location in Arabidopsis leaf starch. J Exp Bot. 62(13):4547-59

Szydlowski N, Ragel P, Raynaud S, Lucas MM, Roldán I, Montero M, Muñoz FJ, Ovecka M, Bahaji A, Planchot V, Pozueta-Romero J, D'Hulst C, Mérida A. (2009) Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starchsynthases. Plant Cell. 21(8):2443-57

Tabata S. and Hizukuri S. (1971) Studies on starch phosphate. Part 2. Isolation of glucose 3-phosphate and maltose phosphate by acid hydrolysis of potato starch. Starch/Stärke **23**, 267-275

Tabata S, Hizukuri S. (1992) Properties of yeast debranching enzyme and its specificity toward branched cyclodextrins. Eur J Biochem. Jun1;**206**(2):345-8.

Tagliabracci V.S. *et al.,* (2011) Phosphate incorporation during glycogen synthesis and Lafora Disease. Cell Metabolism **13**, 274-285

Tagliabracci V.S., Girard J.M., Segvich D., Meyer C., Turnbull J., Zhao X., Minassian B.A., Depaoli-Roach A.A., and Roach P.J. (2008). Abnormal metabolism of glycogen phosphate as a cause for Lafora disease. J Biol Chem 33816-33825.

Takeda Y., Guan H.-P. and Preiss J. (1993). Branching of amylose by the branching isoenzymes of maize endosperm. Carbohydrate Research **240**(0): 253-263.

Teste M. A. *et al.,* (2000) The *Saccharomyces cerevisiae* YPR184w gene encodes the glycogen debranching enzyme. FEMS Microbiol. Lett. **193**, 105-115

Tiberia E, Turnbull J, Wang T, Ruggieri A, Zhao XC, Pencea N, Israelian J, Wang Y, Ackerley CA, Wang P, Liu Y, Minassian BA. (2012) Increased laforin and laforin binding to glycogen underlie Lafora body formation in malin-deficient Lafora disease. J Biol Chem **287**,25650-25659.

Tyra, H. M., M. Linka, A. P. Weber and D. Bhattacharya (2007). "Host origin of plastid solute transporters in the first photosynthetic eukaryotes." Genome Biol **8**(10): R212.

Ugalde JE, Parodi AJ, Ugalde RA. (2003) De novo synthesis of bacterial glycogen: Agrobacterium tumefaciens glycogen synthase is involved in glucan initiation Proc Natl Acad Sci U S A.100(19):10659-63and elongation.

Utsumi Y., Utsumi C., Sawada T., Fujita N., and Nakamura Y. (2011). Functional diversity of isoamylase oligomers: the ISA1 homo-oligomer is essential for amylopectin biosynthesis in rice endosperm. Plant Physiol **156,** 61–77.

Van den Eynde, H., R. De Baere, E. De Roeck, Y. Van de Peer, A. Vandenberghe, P. Willekens and R. De Wachter (1988). "The 5S ribosomal RNA sequences of a red algal rhodoplast and a gymnosperm chloroplast. Implications for the evolution of plastids and cyanobacteria." J Mol Evol **27**(2): 126-132.

Vriet C. *et al.,* (2010) A suite of *Lotus japonicus* starch mutants reveals both conservedand novel features of starch metabolism. Plant Physiol. **154**, 643-65

Waterhouse A.M., Procter J.B., Martin D.M.A, Clamp M. and Barton G. J. (2009).Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. Bioinformatics **25** (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

Wattebled F., Buléon A., Bouchet B., Ral J. P., Lienard L., Delvalle D., Binderup K., Dauvillee D., Ball S. and D'Hulst C. (2002). Granule-bound starch synthase I. A major enzyme involved in the biogenesis of B-crystallites in starch granules. Eur J Biochem **269**(15): 3810-3820.

Wattebled F., Dong Y., Dumez S., Delvalle D., Planchot V., Berbezy P., Vyas D., Colonna P., Chatterjee M., Ball S., and D'Hulst, C. (2005). Mutants of *Arabidopsis* lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiol **138,** 184-195.

Weber, A. P., M. Linka and D. Bhattacharya (2006). "Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor." Eukaryot Cell **5**(3): 609-612.

Welsh, E. A., M. Liberton, J. Stockel, T. Loh, T. Elvitigala, C. Wang, A. Wollam, R. S. Fulton, S. W. Clifton, J. M. Jacobs, R. Aurora, B. K. Ghosh, L. A. Sherman, R. D. Smith, R. K. Wilson and H. B. Pakrasi (2008). "The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle." Proc Natl Acad Sci U S A **105**(39): 15094-15099.

Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J. (2010) Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev. 34(6):952-85

Woo E.J., Lee S., Cha H., Park J.T., Yoon S.M., Song H.N., and Park K.H. (2008). Structural insight into the bifunctional mechanism of the glycogen-debranching enzyme TreX from the archaeon *Sulfolobus solfataricus*. J Biol Chem **283,** 28641-28648.

Worby C.A. *et al.,* (2006) Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates. J. Biol. Chem. **281**, 30412-3041

Yoo, S. H., C. Keppel, M. Spalding and J. L. Jane (2007). "Effects of growth condition on the structure of glycogen produced in cyanobacterium Synechocystis sp. PCC6803." Int J Biol Macromol **40**(5): 498-504.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto and D. Bhattacharya (2004). "A molecular timeline for the origin of photosynthetic eukaryotes." Mol Biol Evol **21**(5): 809-818.

Yu T. S. *et al.,* (2001) The *Arabidopsis* sex mutant is defective in the R protein, ageneral regulator of starch degradation in plants, and not in the chloroplast hexose transporter.Plant Cell **13**, 1907–1915

Zeeman S. C. *et al.* (1998) A mutant of *Arabidopsis* lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. Plant Cell **10**, 1699-1712

Zeeman, S. C., A. Tiessen, E. Pilling, K. L. Kato, A. M. Donald and A. M. Smith (2002). "Starch synthesis in Arabidopsis. Granule synthesis, composition, and structure." Plant Physiol **129**(2): 516-529.

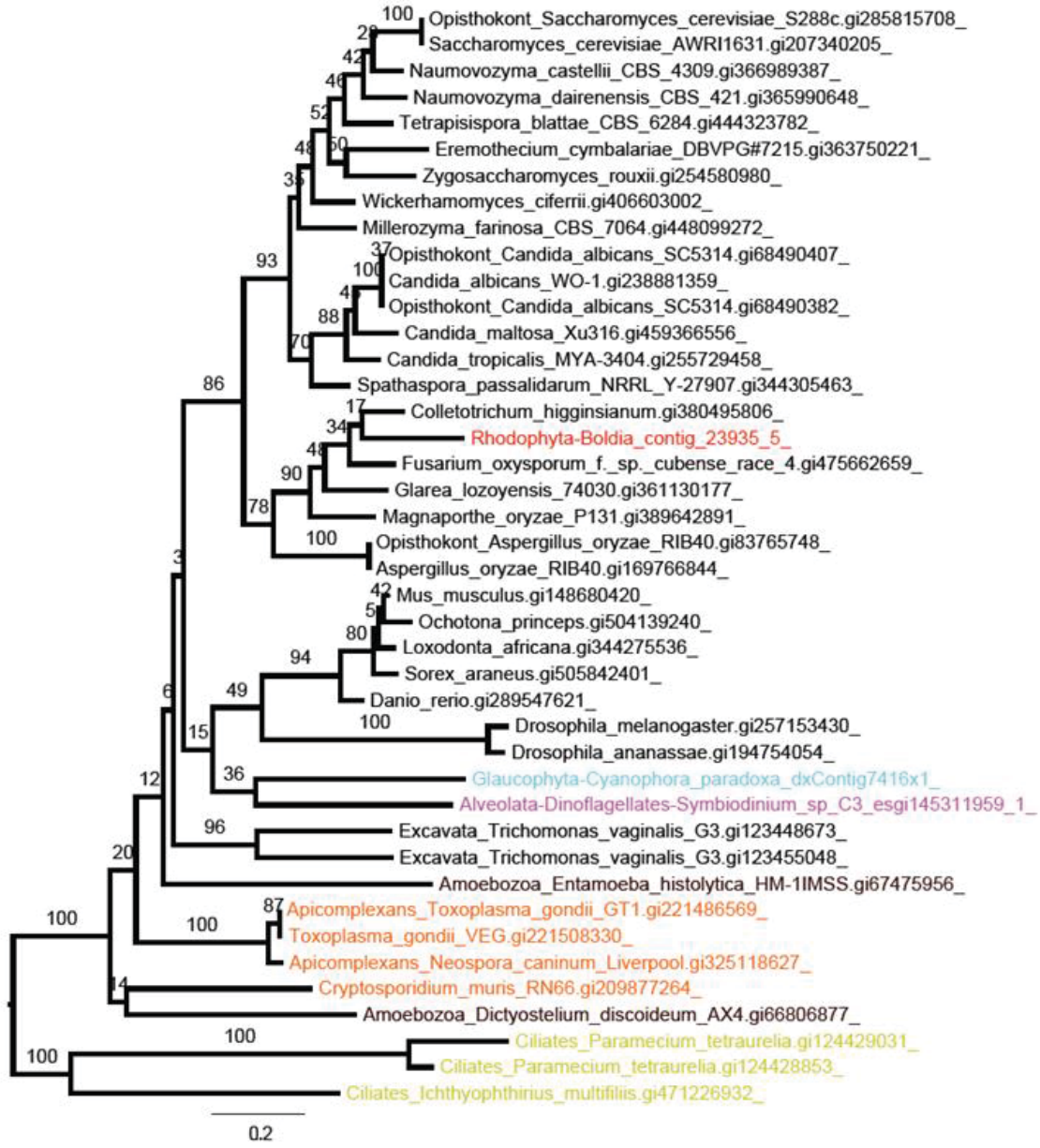
Zeeman SC, Kossmann J, Smith AM. Starch: its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol. 2010;61:209-34.

Zehr, J. P. (2011). "Nitrogen fixation by marine cyanobacteria." Trends Microbiol **19**(4): 162-173.

Zhang X, Szydlowski N, Delvallé D, D'Hulst C, James MG, Myers AM. (2008) Overlapping functions of the starch synthases SSII and SSIII in amylopectin biosynthesis in Arabidopsis. BMC Plant Biol. 8:96

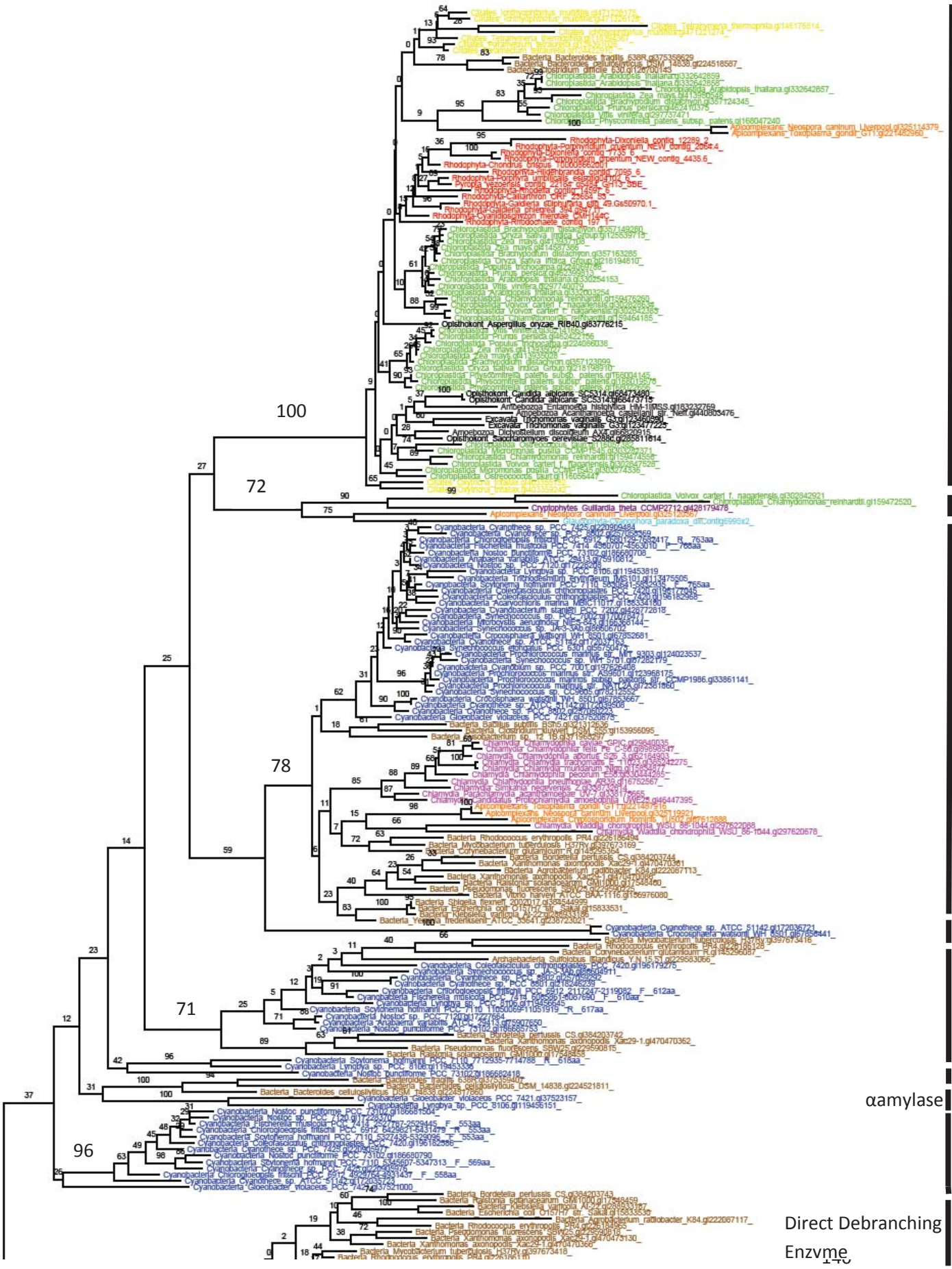
ANNEX

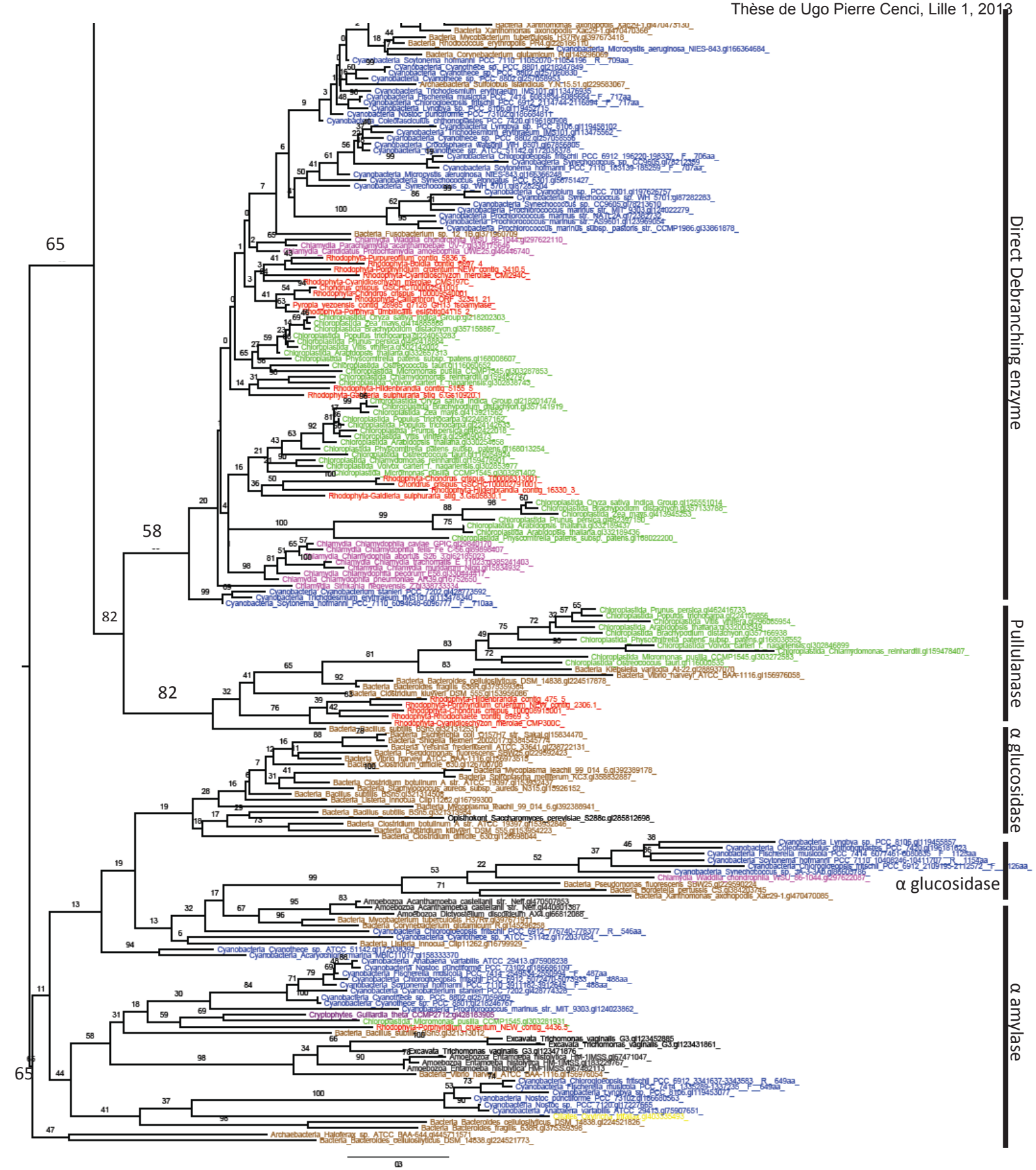
Annex 1



Annex 1: **Phylogenetic tree of putative iDBE enzymes.** Alignments and block selections were performed using the LG+G best-fit model of ProtTest program. Bootstrap values (100 replicates) were calculated using PhyML Putative iDBE sequences found in eucaryotes are indicated, in function of their origin: Rhodophyta in red, Glaucophyta in cyan, Dinoflagellates in purple, Apicomplexans in orange, Amoebozoa in brown, others species are in black (Opisthokonts, Excavata…).

Annex 2





Annex 2 : **Maximum-likelihood phylogenetic tree of archaebacteria, bacteria and eukaryotes enzymes belonging to GH13 family according to CAZy nomenclature.** Alignments and block selections were performed using the LG+G best-fit model of ProtTest 3.2 program. Bootstrap values (100 replicates) were calculated using PhyML, bootstrap values of interesting group are indicated.

Journal of Experimental Botany, Page 1 of 27
doi:10.1093/jxb/erq411

Journal of Experimental Botany
www.jxb.oxfordjournals.org

DARWIN REVIEW

The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis

Steven Ball*, Christophe Colleoni, Ugo Cenci, Jenifer Nirmal Raj and Catherine Tirtiaux

Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS-USTL, Bâtiment C9, Cité Scientifique, F-59655 Villeneuve d'Ascq, France

* To whom correspondence should be addressed: E-mail: steven.ball@univ-lille1.fr

Received 10 September 2010; Revised 18 November 2010; Accepted 23 November 2010

Abstract

Solid semi-crystalline starch and hydrosoluble glycogen define two distinct physical states of the same type of storage polysaccharide. Appearance of semi-crystalline storage polysaccharides appears linked to the requirement of unicellular diazotrophic cyanobacteria to fuel nitrogenase and protect it from oxygen through respiration of vast amounts of stored carbon. Starch metabolism itself resulted from the merging of the bacterial and eukaryote pathways of storage polysaccharide metabolism after endosymbiosis of the plastid. This generated the three Archaeplastida lineages: the green algae and land plants (Chloroplastida), the red algae (Rhodophyceae), and the glaucophytes (Glaucophyta). Reconstruction of starch metabolism in the common ancestor of Archaeplastida suggests that polysaccharide synthesis was ancestrally cytosolic. In addition, the synthesis of cytosolic starch from the ADP-glucose exported from the cyanobacterial symbiont possibly defined the original metabolic flux by which the cyanobiont provided photosynthate to its host. Additional evidence supporting this scenario include the monophyletic origin of the major carbon translocators of the inner membrane of eukaryote plastids which are sisters to nucleotide-sugar transporters of the eukaryote endomembrane system. It also includes the extent of enzyme subfunctionalization that came as a consequence of the rewiring of this pathway to the chloroplasts in the green algae. Recent evidence suggests that, at the time of endosymbiosis, obligate intracellular energy parasites related to extant *Chlamydia* have donated important genes to the ancestral starch metabolism network.

Key words: Archaeplastida, *Chlamydia*, cyanobacteria, endosymbiosis, evolution of photosynthesis, glycogen, plastids, starch.

Introduction

Sometime between 0.7–1.5 billion years ago (Cavalier-Smith, 2006; Yoon *et al.*, 2004) an ancestor of present-day cyanobacteria was internalized, probably through phagocytosis (Raven *et al.*, 2009) by a heterotrophic eukaryotic cell. That this was a unique event is suggested by the fact that both protein sequences derived from the cyanobiont (the cyanobacterial endosymbiont) and those from the eukaryotic host are monophyletic and thus can be traced back to a pair of unique ancestors (McFadden and van Dooren, 2004; Rodriguez-Ezpeleta *et al.*, 2005).

Although nothing is known about the nature of the ancient endosymbiotic link, it is reasonable to assume that the latter was based on the export of photosynthate from the cyanobiont to the host cytosol. Endosymbiosis of the plastid thus brought the ability to perform oxygenic photosynthesis into the eukaryotic world. As the cyanobiont slowly evolved to become a true organelle, the majority of cyanobacterial genes were lost as they were neither involved in oxygenic photosynthesis nor essential for maintenance and division of the symbiont. During this process, a complex

machinery of protein targeting from the cytosol to the evolving plastid appeared, thereby facilitating a process by which the remaining genes were transferred to the nucleus and their protein products synthesized on cytosolic ribosomes to be retargeted to the organelle. In addition, a number of other protein products and pathways were rewired to the evolving organelle which were not all necessarily present in the ancestral cyanobiont.

Three eukaryotic lineages emerged after or during this metabolic integration of the plastid (Fig. 1): the Chloroplastida (green algae and land-plants), the Rhodophyceae (red algae), and the Glaucophyta (glaucophytes). These three lineages generated through primary endosymbiosis contain the original 'old' plastids with two membranes and were therefore recently named 'Archaeplastida' (Adl *et al.*, 2005). Some single cell members or ancestors from these lineages were internalized, probably also through phagocytosis, by other heterotrophic eukaryotes, thereby generating a variety of secondary endosymbiosis lines with derived plastids (Keeling, 2009). These secondary plastids are always surrounded by more than two and most of the time by four membranes. This generated a number of other important photosynthetic eukaryotes such as the brown algae, diatoms, dinoflagellates, cryptophytes, and haptophytes.

In addition to photosynthesis, eukaryotes have gained a number of other important biochemical features not found in heterotrophic eukaryotes unrelated to Archaeplastida. Among these, is the ability to store starch, an insoluble and semi-crystalline form of storage polysaccharide, which, until quite recently, was only reported in Archaeplastida and some, but not all, of their secondary endosymbiosis derivatives. Plant biologists are familiar with a form of starch found in the chloroplast or amyloplast of land plants and green algae. However, this polysaccharide is only found in the cytosol of red algae, glaucophytes, dinoflagellates, and the non-photosynthetic sister lineages of the latter: the apicomplexa parasites. In the cryptophytes, starch is found in the periplastidial space a compartment corresponding to the cytosol of the archaeplastidal alga that was internalized through secondary endosymbiosis to generate, among others, the cryptophyte lineage. Cytosolic starch was historically first studied in Florideophycidae, a complex group of multicellular red algae (for a review see Viola *et al.*, 2001). The term floridean starch was therefore coined to describe this form of storage material (cytosolic or periplastidial starch will thus be referred to as 'floridean' starch in this review). Therefore plastidial starch remains the exception rather than the rule among the diversity of starch-storing lineages.

This review is centered on the evolution of the starch pathway. Developments and refinements in the evolution of starch metabolism in grasses have recently been reviewed (Comparat-Moss and Denyer, 2009). In this Darwin Review, the focus will be on the means by which storage polysaccharide metabolism from the cyanobiont and its eukaryotic host merged to generate the starch pathway. We will propose that this merging of pathways was central to the success of primary endosymbiosis as it established the first biochemical link between the two unrelated partners.

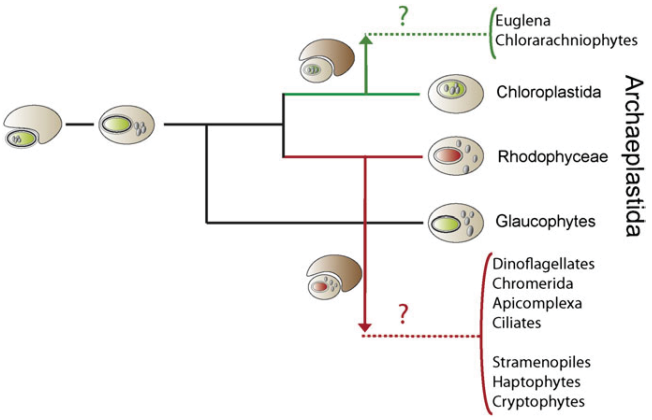


Fig. 1. Primary and secondary plastid endosymbiosis. Photoysynthetic eukaryotes are derived from a unique event involving phagoyctosis of a cyanobacterial ancestor by a heterotrophic eukaryotic host. The ancestral cyanobiont is depicted as a peptidoglycan-containing single cell cyanobacterium (in green) with both inner and outer membranes and no outer layer capsular polysaccharides as is presently the case for the *Paulinella* chromatophores or the glaucophyte cyanelles. The Archaeplastida define the extant photosynthetic eukaryotic lineages that have emerged from this unique ancestor. Among the Archaeplastida, the glaucophytes define single-cell freshwater algae containing a plastid called the cyanelle with phycobilisomes and other typical cyanobacterial-like features but displaying the same level of genome simplification and organization as other plastids. Starch is found in the cytosol of all glaucophytes. Red (Rhodophyceae) and green (Chloroplastida) algae contain plastids with no peptidoglycan called, respectively, rhodoplasts and chloroplasts. They can be distinguished by the structure and composition of their photosynthetic antennae which, in Chloroplastida, contain chlorophyll *b* while red algae still rely on bacterial phycobilins such as the red pigment phycoerythrin. Starch is found in the cytosol of red algae while it is found in plastids of all Chloroplastida including the land plants. It is presently thought that the cyanobacterial ancestor was a starch accumulator while the heterotrophic eukaryotic host partner synthesized glycogen in its cytosol. The ability to synthesize starch was transmitted to the archaeplastidal ancestor cytosol while it was lost by the cyanobiont. Upon evolution of the Chloroplastida, starch metabolism was rewired to the evolving chloroplasts. The Archaeplastida themselves became the substrate for 'secondary' endosymbiosis through phagocytosis by other heterotrophic eukaryotic lineages. The exact number of secondary endosymbiosis events is still debated (hence this is symbolized by dotted lines). Unlike primary endosymbiosis, the phagocytosis vacuole was either kept or fused with the ER, thus yielding, in most cases, four-membrane 'secondary' plastids. Starch was not universally transmitted to secondary endosymbiosis lines which mostly accumulate β-glucans. Starch, however, is still found in the cytosol of most dinoflagellates and some apicomplexa parasites and between the 2nd and 3rd membrane of secondary cryptophyte plastids.

Starch and glycogen define two different physical states of α-glucan storage polysaccharide metabolism

Living cells store carbohydrates in the form of a variety of polymers and oligomers. Among these, glycogen defines by far the most widespread form of storage as it is found in Archaea, Bacteria, and Eukaryotes. Glycogen is made of α-1,4 linked chains of glucose (α-1,4 glucans) that are branched together through α-1,6 linkages. The α-1,6 branches accounts for 7–10% of the linkages and are evenly distributed within the glycogen particle (for a review of glycogen structure see Shearer and Graham, 2002). Each chain, with the exception of the outer unbranched chains, supports two branches. This branching pattern allows for spherical growth of the particle generating tiers (a tier corresponds to the spherical space separating two consecutive branches from all chains located at similar distance from the center of the particle). This type of growth leads to an increase in the density of chains in each tier leading to a progressively more crowded structure towards the periphery (Fig. 2A).

Mathematical modelling predicts a maximal value for the particle size above which further growth is impossible as there would not be sufficient space for interaction of the chains with the catalytic sites of glycogen metabolism enzymes. This generates a particle consisting of 12 tiers corresponding to a 42 nm maximal diameter including 55 000 glucose residues. 36% of this total number rests in the outer (unbranched) shell and is thus readily accessible to glycogen catabolism without debranching (Shearer and Graham 2002). *In vivo*, glycogen particles are thus present in the form of these limit size granules (macroglycogen) and also smaller granules representing intermediate states of glycogen biosynthesis and degradation (proglycogen) (Shearer and Graham, 2002). Glycogen particles are entirely hydrosoluble and, therefore, define a state where the glucose is rendered less active osmotically yet readily accesible to rapid mobilization through the enzymes of glycogen catabolism as if it were in the soluble phase.

Starch defines a solid semi-crystalline state composed of a mixture of two different polysaccharides with the same basic chemical linkages as glycogen (for a review of starch structure see Buléon *et al.*, 1998). Amylopectin, the major polysaccharide fraction is indispensable for starch granule formation and contains 4–6% branches while the minor fraction amylose contains less than 1% α-1,6 linkages. Amylose requires a pre-existing amylopectin-containing granule for its formation (Dauvillée *et al.*, 1999). Mutants deprived of this fraction can be readily isolated in green plants and algae (for a review see Ball *et al.*, 1998). These mutants build

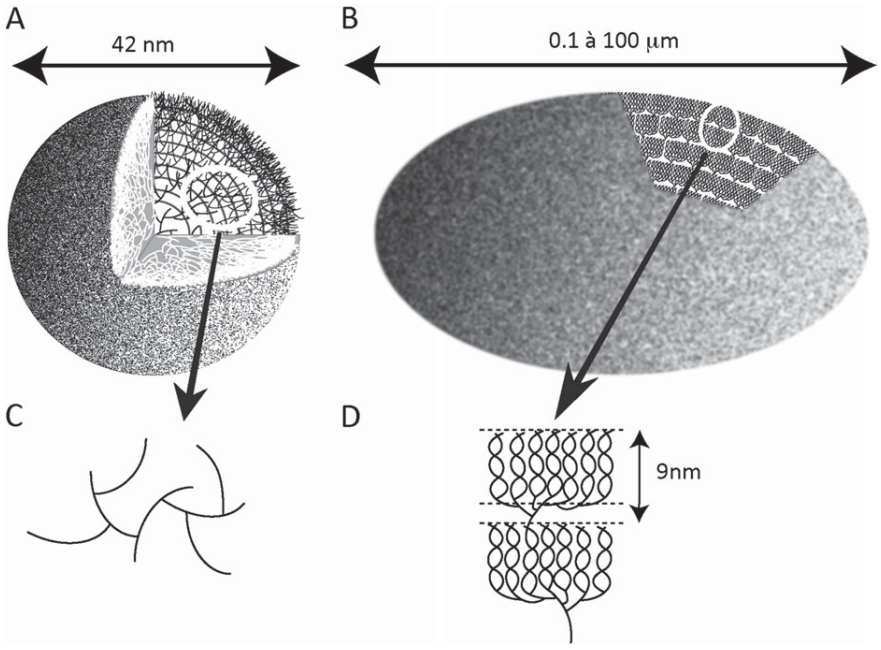


Fig. 2. Schematic representation of whole glycogen (A) and starch (B) granules. The lines represent α-1,4-linked glucan chains and the intersections of such lines symbolize the α-1,6 branches. (C, D) Enlarged views of the circled sections of the corresponding glycogen (C) and starch (D) granules. The distribution of branches exemplified in (C), with two α-1,6 linkages per glucan, leads to the exponential increase in the density of chains as one moves away from the centre of the particle. This leads to a predictable maximum of 42 nm for the glycogen granule displayed in (A). Indeed, further density increases will not accommodate the sizes of the glycogen metabolism enzymes active sites. (D) Two typical amylopectin clusters are displayed. The cluster structure is generated through the asymmetric distribution of the branches which are shown at the base of each of the two clusters. The small portion containing the branches is called the amorphous lamella of the unit cluster while the chains generated through the branches intertwine to form the double helical structures that define the unit crystalline lamella. The sum of one amorphous and one crystalline lamella amounts to 9 nm in all amylopectin clusters examined so far.

wild-type amounts of normally organized granules. On the other hand some floridean starch-accumulating lineages, such as florideophycideae red algae (Viola *et al.*, 2001) or apicomplexan parasites (Coppin *et al.*, 2005), lack amylose while sister lineages of the latter (such as the Porphyridales red algae; Shimogana *et al.*, 2007, 2008) or the dinoflagellates (Deschamps *et al.*, 2008*d*) typically include this polysaccharide fraction. Amylose, however, is always found in the granules synthesized within plastids by wild-type green algae and land plants (Ball *et al.*, 1998).

Amylopectin defines one of, if not the largest, biological polymer known and contains from 10^5–10^6 glucose residues (Buléon *et al.*, 1998). There is no theoretical upper limit to the size reached by individual amylopectin molecules. This is not due to the slightly lesser degree of overall branching of the molecule when compared to glycogen. Rather it is due to the way the branches distribute within the structure. As displayed in Fig. 2B, the branches are concentrated in sections of the amylopectin molecule leading to clusters of chains that allow for indefinite growth of the polysaccharide. Another major feature of the amylopectin cluster structure consists of the dense packing of chains generated at the root of the clusters where the density of branches locally reaches or exceeds that of glycogen. This dense packing of branches generates tightly packed glucan chains that are close enough to align and form parallel double helical structures. The helices within a single cluster and neighbouring clusters align and form sections of crystalline structures separated by sections of amorphous material (containing the branches) thereby generating the semi-crystalline nature of amylopectin and of the ensuing starch granule (Buléon *et al.*, 1998). Indeed the crystallized chains become insoluble and typically collapse into a macrogranular solid. This osmotically inert starch granule allows for the storage of unlimited amounts of glucose that become metabolically unavailable. Indeed the enzymes of starch synthesis and mobilization are unable to interact directly with the solid structure with the noticeable exception of granule-bound starch synthase the sole enzyme required for amylose synthesis. This enzyme is able to extend amylose chains by synthesizing α-1,4 glucosyl linkages progressively within the polysaccharide matrix (reviewed in Ball *et al.*, 1998). Because no other enzyme is significantly active within granules, this will lead to the formation of long unbranched polysaccharides.

On the other hand, in Archaeplastida, glucan-water dikinase initiates amylopectin degradation by phosphorylating selective glucose residues within the clusters, thereby disrupting the crystal and facilitating access and attack by hydrosoluble enzymes of starch catabolism (reviewed in Fettke *et al.*, 2009). The solid state of starch thereby generates glucose stores which are not as readily accessible as those of glycogen. Consequently, starch can be seen as a very efficient intracellular sink immobilizing vast amounts of carbon out of cellular metabolism. Mobilizing starch is thus anything but trivial. Indeed because starch defines the most important source of calories in the human diet, human populations have duplicated genes encoding salivary α-amylase as a function of their local diet (Shadan, 2007). Only a small fraction of damaged uncooked starch granules are mobilized during digestion. Because starch granules swell and melt at high temperatures, thereby loosening the crystal structure, cooking meals has vastly improved the amount of calories that humans can extract from such polysaccharides in their diet.

As previously mentioned, the distribution of starch polysaccharides seemed, until recently, to be limited to Archaeplastida and some of their secondary endosymbiosis derivatives. Therefore the large amounts of carbohydrates and energy available through photosynthesis do not, per se, explain the appearance of this form of storage material. Indeed most photosynthetic bacteria including cyanobacteria were reported to accumulate glycogen and not semi-crystalline starch.

Comparative biochemistry of glycogen metabolism in bacteria and opistokonts

As we will see, the enzymes of glycogen and starch metabolism are clearly related. In addition, in Archaeplastida, the pathways of starch biosynthesis and degradation define a mosaic of enzymes phylogenetically related either to bacterial (cyanobacterial and chlamydial) or eukaryotic glycogen metabolism (Coppin *et al.*, 2005; Patron and Keeling, 2005; Deschamps *et al.*, 2008*a*). The obvious explanation for this observation would be that both partners of plastid endosymbiosis had the ability to synthesize related storage polysaccharides before endosymbiosis. These certainly consisted of α-1,4-linked glucans branched through α-1,6 linkages. Glycogen metabolism defines well-studied and conserved pathways within gram-negative bacteria and opistokonts (fungi and animals) who define those eukaryotes that have by far been the most intensively studied. To understand the merging of these pathways that occurred after endosymbiosis, their common and distinctive features will be briefly outlined. Figure 3 summarizes the basic common pathway of storage polysaccharide synthesis in gram negative bacteria (and cyanobacteria) (for a review see Preiss, 1984) and opistokonts (for reviews see Roach, 2002; Wilson *et al.*, 2010).

Briefly, glucose is polymerized within these polysaccharides, thanks to its activation in the form of a nucleotide-sugar through the action of NDP-glucose pyrophosphorylase. All eukaryotes known (with the exception of Archaeplastida) synthesize glycogen from UDP-glucose while all gram-negative glycogen accumulating bacteria use ADP-glucose. ADP-glucose is a bacterial-specific metabolite not found in heterotrophic eukaryotes. Unlike UDP-glucose which is used by all living cells to synthesize a large number of different molecules, ADP-glucose is devoted to the synthesis of glycogen in bacteria (and also to the osmoprotectant glucosylglycerol in cyanobacteria) (Preiss, 1984; Miao *et al.*, 2003, 2006). Thus, the synthesis of ADP-glucose defines the first committed step of glycogen synthesis in bacteria while glucan elongation defines the first committed step of eukaryotic

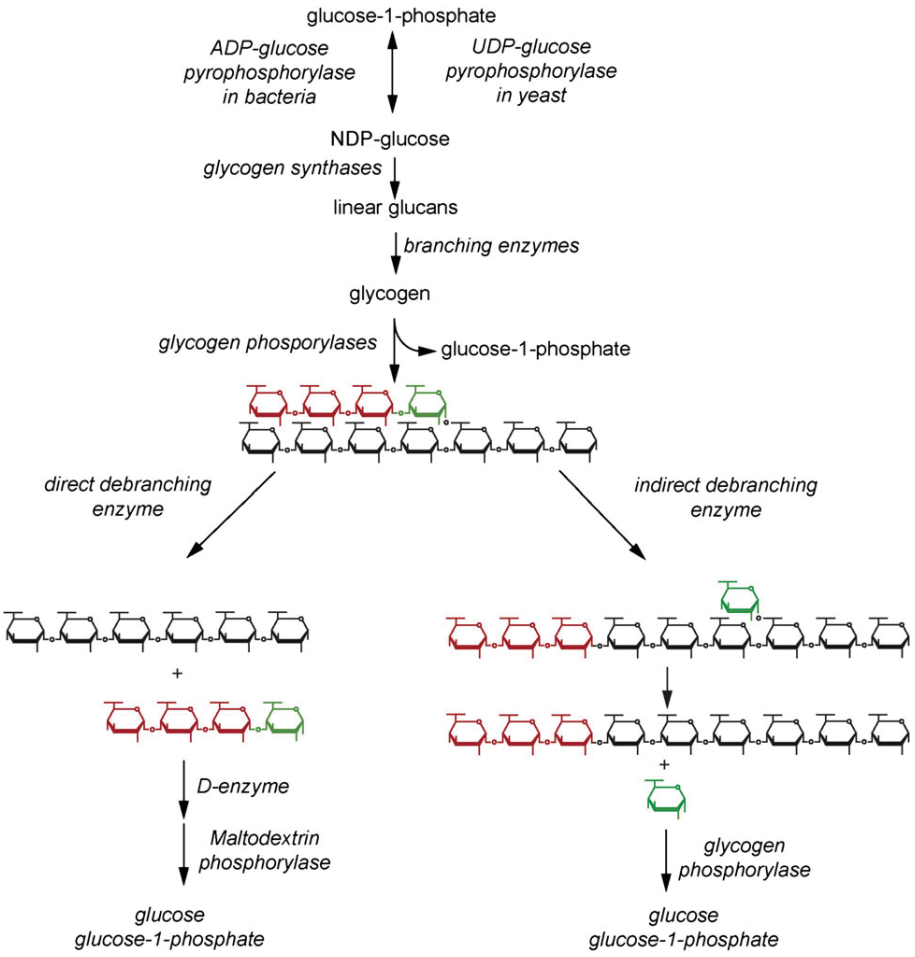


Fig. 3. Overview of glycogen metabolism in bacteria and eukaryotes. Both bacteria and eukaryotes synthesize glycogen from activated nucleotide-sugar substrates. The latter consist of ADP-glucose in bacteria and UDP-glucose in eukaryotes. The nucleotide sugar (NDP-glucose) is, in both cases, used for the transfer of glucose on the non-reducing end. Branching involves similar enzymes and reactions in both cases (see text). Glycogen breakdown is initiated in bacteria and eukaryotic pathways by glycogen phosphorylase which, in the presence of orthophosphate, generates glucose-1-P from the available non-reducing ends. In both cases, phosphorylase stops four glucose residues away from an α-1,6 branch generating the outer chain polysaccharide structure illustrated with the α-1,6 bound glucose in green and the novel chain generated by the branch symbolized by glucose residues in red. In bacteria (left column of the drawing), the direct debranching enzyme releases a maltotetraose (four Glc residues) and a glycogen particle with a longer outer chain (in this example, six Glc residues) that becomes a substrate for glycogen phosphorylase. In eukaryotes (right column of the drawing) indirect debranching enzyme hydrolyses the α-1,4 linkages next to the branch and transfers it on the neighbouring outer chain leading to further release of glucose-1-P by phosphorylase. The unmasked branch (in green) is then hydrolysed by a second active site within the same indirect DBE enzyme. Release of maltotetraose by the direct DBE of bacteria (left column) requires the presence of MOS metabolism enzymes such as D-enzyme and maltodextrin phosphorylase.

glycogen synthesis. The glucose from the glycosyl-nucleotide is then transferred to the non-reducing end of a growing α-1,4 linked chain through an elongation reaction catalysed by glycogen synthase. Branching proceeds differently through an hydrolytic cleavage of a pre-existing α-1,4-linked glucan synthesized through glycogen synthase and an intra or intermolecular transfer of a segment of chain in the α-1,6 position. The branched polymers are subjected to degradation through a combination of glycogen phosphorylase and debranching enzyme. Glycogen phosphorylase defines an enzyme which releases glucose-1-P from the non-reducing-end of glycogen in the presence of orthophosphate. This enzyme is unable to cleave the α-1,6 branch and is known to stop four glucose residues away from the branch (Dauvillée *et al.*, 2005; Alonso-Casajús *et al.*, 2006). Therefore the short four-glucose-residues-long external chains need to be further digested through the action of debranching enzymes. Debranching enzymes in eukaryotes and bacteria operate differently. In eukaryotes, indirect debranching enzyme defines a bifunctional enzyme containing both an α-1,4 glucanotransferase and an α-1,6 glucosidase catalytic site. The transferase will first hydrolyse the last α-1,4 linkage before the branch and thus transfer three glucose residues (maltotriose) to an outer neighbouring chain within the glycogen particle. Glycogen phosphorylase will further degrade this seven-glucoseresiduelong chain back to four while

the second catalytic site will hydrolyse the α-1,6 linkage from the residual unmasked glucose at the branch (for reviews see Roach, 2002; Wilson *et al.*, 2010). The net result will consist of complete degradation of glycogen to glucose-1-P and glucose.

Bacteria operate through a simpler debranching enzyme that directly cleaves the α-1,6 branch thereby producing a four-glucose-residue-long malto-oligosaccharide (maltotetraose) (Dauvillée *et al.*, 2005). These malto-oligosaccharides are then degraded through a combination of α-1,4 glucanotransferase and a maltodextrin phosphorylase distinct from the glycogen phosphorylase (reviewed in Boos and Schuman, 1998). Here again the transferase elongates an acceptor maltotetraose with a donor maltotriose enabling maltodextrin phosphorylase to degrade the chains further. Thus direct debranching in bacteria implies the coupling of glycogen and malto-oligosaccharide (MOS) metabolism (Boos and Schuman, 1998) while MOS metabolism is not needed and indeed is not found in opistokont genomes.

In addition to this phosphorolytic pathway of glycogen mobilization, there is good evidence for the presence of a hydrolytic pathway in opistokonts and circumstantial evidence for the presence of such a pathway in gram-negative bacteria. Fungi and animals indeed contain an enzyme able to hydrolyse both the α-1,4 and α-1,6 linkages responsible for the degradation of a significant pool of cellular glycogen (reviewed by François and Parrou, 2001; Roach, 2002; Wilson *et al.*, 2010). However, this enzyme is contained in the lysosome (or yeast vacuole) leading to a clear partitioning between the locations of both glycogen synthesis or phosphorolysis which occurs in the cytosol and glycogen hydrolysis which is confined to the lysosome (or yeast vacuole). In yeast, autophagy clearly further impacts the regulation of glycogen metabolism (Wang *et al.*, 2001). Undisputable functional proof of the importance of the glycogen hydrolysis pathway has been obtained, both in yeasts where it is triggered during sporulation or the late stationary phase and also in humans, where its absence is known to lead to Pompe's disease (glycogen storage disease type II) (reviewed by François and Parrou, 2001; Roach, 2002; Wilson *et al.*, 2010). In bacteria and cyanobacteria, α-amylase-like sequences are often found in the genomes, suggestive of the presence of such a pathway but mutant evidence is lacking (Wing-Ming *et al.*, 1994; Reyes-Sosa *et al.*, 2010).

It is striking to note that mutations abolishing analogous enzyme activities in model organisms, such as *E. coli* and yeast lead to similar or identical phenotypes establishing that all enzymes play analogous functions in the storage polysaccharide metabolism network. This even remains true for very different enzymes such as indirect or direct DBEs from bacteria and opistokonts (Teste *et al.*, 2000; Dauvillée *et al.*, 2005). Nevertheless, the use of distinct nucleotide sugars for glycogen polymerization will impact very differently the regulation of the prokaryotic and eukaryotic pathways. The synthesis of ADP-glucose by ADP-glucose pyrophosphorylase being the first committed step of bacterial glycogen synthesis, this enzyme will be subjected to tight allosteric regulations with effectors that vary according to metabolic specialization of the bacterial species (Preiss, 1984). Cyanobacterial ADP-glucose pyrophosphorylase, in particular, is known to be activated by 3-PGA and inhibited by orthophosphate. This regulation, in addition to the presence of ATP and glucose-1-P as substrates, further couples ADP-glucose synthesis to carbon fixation through the Calvin cycle and thus to photosynthesis, a regulation which was conserved in the case of plastidial starch synthesis in Chloroplastida (for a review see Ballicora *et al.*, 2003) In opistokonts, protein phosphorylation and dephosphorylation through protein kinases and phosphatases has been known for years to activate or inhibit glycogen synthase and glycogen phosphorylase by modifying their sensitivity to allosteric effectors. Historically, protein kinases and phosphatases were discovered by studying the physiology of glycogen metabolism in animals (Krebs, 1983).

The glycogen synthase of opistokonts is a complex enzyme belonging to a distinct class of glycosyltransferase (GT3 according to the CAZy classification) than that of the bacterial enzyme (GT5). The GT3 opistokont wild-type enzyme is unable to prime the reaction and requires a separate malto-oligosaccharide primer. The 'natural' primer for the fungal or animal enzyme is a small protein capable of autoglucosylation: glycogenin. Functional evidence for the importance of glycogenin in glycogen metabolism has been produced in yeast and animals (Roach, 2002; Wilson *et al.*, 2010). However, in bacteria, biochemical evidence suggests that the GT5 ADP-glucose requiring starch synthase is capable of autoglucosylation and therefore does not need the presence of another protein to prime glycogen synthesis (Ugalde *et al.*, 2003). In total, the glycogen pathways of bacteria and opistokonts consists of a network of 6–12 enzymes of related function.

Comparative biochemistry of glycogen metabolism in opistokonts, amoebozoa, and other heterotrophic eukaryotes

Over 99% of the studies performed on glycogen metabolism in eukaryotes concerns fungi and animals (for reviews see François and Parrou, 2001; Roach 2002; Shearer and Graham, 2002; Wilson *et al.*, 2010). Fungi, animals, and lesser known related lineages such as the choanoflagellates define a monophyletic lineage named the opistokonts. This, while of great importance to humans, defines only a small subset of the diversity that typifies eukaryotes (Fig. 4). Because the eukaryotic ancestor that hosted the cyanobiont is not presently thought to define an opistokont or an opistokont ancestor, it becomes important to investigate the nature of storage polysaccharide metabolism to ascertain that the model generated by available studies also applies to other lineages. Among the non-opistokont glycogen accumulating lineages, a number of genomes have recently appeared that are relevant to this question.

Amoebozoa define an important and diverse group of organisms thought to be located closer to the proposed root

of the eukaryotic tree of life (Richards and Cavalier-Smith, 2005) (Fig. 4). *Dictyostellium discoideum* defines an interesting model familiar to cell biologists and geneticists. The genome of this organism has been sequenced (Eichinger *et al.*, 2005). Among the surprising features displayed by this genome is the presence of a greater number of distinct protein domains than that found either in fungi and animals (Eichinger *et al.*, 2005). A logical explanation for this increase would be the conservation of the initially greater diversity of genes that typified the ancestors of eukaryotes. Glycogen metabolism also displays this increase in complexity. Indeed not only does *Dictyostellium* harbour the full suite of genes found in fungi and animals for glycogen metabolism, but, in addition, it includes a second type of glycogen synthase belonging to the GT5 CAZy family (Deschamps *et al.*, 2008*a*; Cantarel *et al.*, 2009). Interestingly, it also contains an α-1,4 glucan transferase named dpe2 and related amoeba, such as the pathogen *Entamoeba histolytica*, contain both dpe2 and β-amylase (Loftus *et al.*, 2005; Deschamps *et al.*, 2008*a*) as do all Archaeplastida where these enzymes were first described (see the next section for a description of the function of these enzymes). Dpe2 is found together with β-amylase in other eukaryotic lineages unrelated to amoebozoa including the parabasalid *Trichomonas vaginalis* (Carlton *et al.* 2007; Deschamps *et al.*, 2008*a*). As to the GT5 glycogen synthase, this enzyme is also found in place of the GT3 enzyme in ciliates, other amoebas, parabasalids, and also in Glaucophyta, Rhodophyceae (red algae), and lineages thought to derive from them by secondary endosymbiosis such as the apicomplexa parasites (Coppin *et al.*, 2005; Aury *et al.*, 2006; Eisen *et al.*, 2006; Carlton *et al.*, 2007; Deschamps *et al.*, 2008*a*). Dpe2 and β-amylase were first reported in green plants and believed therefore to define green-lineage-specific genes. However, the very wide distribution of these additional enzymes of glycogen metabolism among eukaryotic lineages separated by over a billion years of evolution (Song *et al.*, 2005) argues that their presence cannot be explained by lateral gene transfer from Chloroplastida. The most logical explanation would consist of the existence of a richer suite of genes of glycogen metabolism in the eukaryotic ancestors that was followed by different histories of selective gene losses in distinct eukaryotic lineages. For instance, opisthokonts would have lost β-amylase, dpe2, and the GT5 glycogen synthase while parabasalids and archamoebas would only have lost the GT3 enzyme. Ciliates would have lost β-amylase, dpe2, and the GT3 enzyme. Amoebas, in general, and mycetozoa, in particular, such as *Dictyostellium discoideum*, would have experienced fewer gene losses than other eukaryotes.

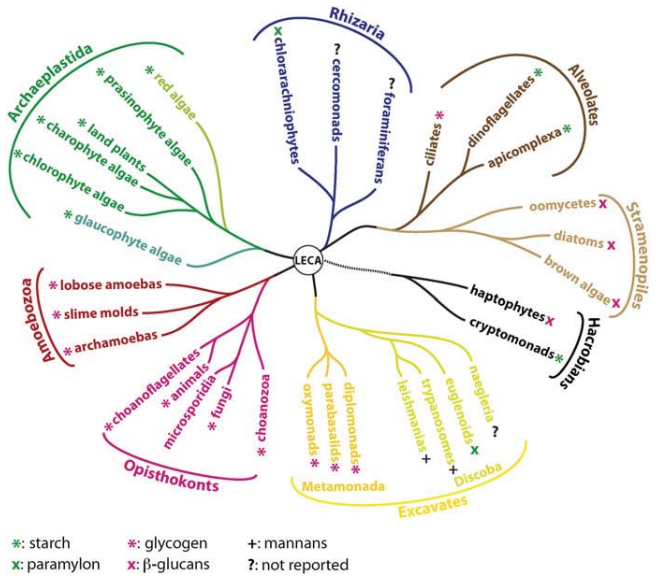


Fig. 4. Eukaryotic tree of life. The tree summarizing and simplifying our current understanding of eukaryotic phylogeny was inspired by those recently published by Baldauf (2003), Hampl *et al.* (2009), and Koonin (2010). The position of Hacrobians within the tree is problematic since they have been shown by Hampl *et al.* (2009) to disrupt the Archaeplastida monophyly. LECA illustrates the last eukaryotic common ancestor. When known, the nature of the storage polysaccharide present is reported; when unknown, the latter is displayed with a question mark (?). In short, eukaryotes can be subdivided into α and β glucan accumulators. In both cases, hydrosoluble oligo or polysaccharides and crystalline polysaccharides have been reported. Starch and paramylon define the solid insoluble crystalline α and β glucan storage polysaccharides, respectively, while maltoligosaccharides, glycogen, laminarin, chrysolaminarin (leucosin), and mycolaminarin define the hydrosoluble storage oligosaccharide and polysaccharide versions.

The detailed function of the GT5 UDP-glucose utilizing enzyme (although the suspected substrate specificity remains to be formally proven) in the glycogen metabolism network, as well as the cytosolic or lysosomal location of the putative β-amylase-dependent pathway of glycogen hydrolysis, remains to be ascertained. An interesting question concerning the GT5 UDP-glucose utilizing glycogen synthase consists of its dependence on glycogenin for priming, and its possible regulation through the well-known set of protein kinases and phosphatases that normally control the GT3 enzyme. Because of the maintenance of a richer suite of enzymes involved in glycogen metabolism in Amoebozoa *Entamoeba histolytica* has been chosen as our reference genome to exemplify the status of glycogen metabolism as it possibly existed in the eukaryotic partner of endosymbiosis before the latter engulfed the cyanobiont.

A brief overview of starch metabolism in Chloroplastida

Decades of research and a wealth of studies concerning starch metabolism in Chloroplastida have led to the identification of a very well-conserved pathway from the earliest diverging prasinophyte single cell alga such as *Ostreococcus* to the most complex multicellular terrestrial plants such as maize or rice (Ral *et al.*, 2004; Derelle *et al.*, 2006; Deschamps *et al.*, 2008*b*). Many reviews are accessible for the interested reader that concern our present detailed

understanding of starch biosynthesis and degradation (Myers *et al.*, 2000; Ball and Morell, 2003; Tomlinson and Denyer, 2003; Tetlow *et al.*, 2004*a*; Morell and Myers, 2005; Zeeman *et al.*, 2007; Fettke *et al.*, 2009). A general feature of the plastidial pathways of starch metabolism is defined by its astonishing apparent complexity. Over 40 genes (not including regulatory genes) seem involved in the building and mobilization of starch in plastids while fewer than 12 genes are comparatively directly involved in glycogen metabolism both in eukaryotes and bacteria. This apparent increase in complexity is largely due to the high number of isoforms that catalyse each of the steps that have been outlined in the preceding sections. For instance, a minimum of five starch synthases participate in polymer elongation, three branching enzymes are reported to introduce the α-1,6 linkage, four direct debranching enzymes are involved in different facets of starch metabolism etc. These enzymes play only partly redundant functions with one another and are often responsible for distinctive roles in the building or degradation of different substructures of starch. Because the starch granule defines a highly organized structure, it was believed by many that this was required to explain the underlying complexity of the granule architecture.

The chloroplastidal pathway relies on the sole use of ADP-glucose (Lin *et al.*, 1988; Zabawinski *et al.*, 2001). The enzymes of ADP-glucose production and those that elongate glucans with this substrate display a distinctive bacterial phylogeny which apparently correlates with the plastidial location of starch in the green lineage (Coppin *et al.*, 2005; Patron and Keeling, 2005, Deschamps *et al.*, 2008*a*). Most importantly, ADP-glucose pyrophosphorylase has conserved the major regulatory properties of the cyanobacterial enzyme and has thus remained throughout its history an enzyme which is tightly coupled to the Calvin cycle and photosynthesis (Ballicora *et al.*, 2003). The starch pathway resembles that of cyanobacterial glycogen metabolism with two major differences. The first difference pertains to the means by which plants achieve the asymmetric distribution of branches within the amylopectin clusters that explains the solid semi-crystalline state of starch and most of its physical properties. Mutant work in *Chlamydomonas*, cereals, and later in *Arabidopsis* (James *et al.*, 1995; Mouille *et al.*, 1996; Nakamura *et al.*, 1997; Zeeman *et al.*, 1998; Wattebled *et al.*, 2005) have strongly suggested that a form of direct debranching enzyme, named isoamylase, debranches the loosely spaced α-1,6 linkages only within the hydrosoluble precursor of amylopectin, thereby generating the tight spacing of branches required at the root of clusters for polymer crystallization (Ball *et al.*, 1996). In the absence of this activity, mutants of *Chlamydomonas* revert entirely to the synthesis of glycogen (Mouille *et al.*, 1996).

A second major difference consists in the way the starch granule is degraded. Chloroplastida enzymes of starch catabolism are unable directly to attack the solid granule. In order to mobilize starch, the latter must first be phosphorylated through an enzyme named glucan water dikinase (GWD) that carries both a starch binding and a dikinase domain (Ritte *et al.*, 2002; reviewed in Blennow *et al.*, 2002; Fettke *et al.*, 2009). The β-phosphate from ATP is thus bound to the C6 of a few glucose residues within the crystalline lamellae (Ritte *et al.*, 2002). The pre-phosphorylated lamellae are then further phosphorylated through PWD (phosphoglucan water dikinase) which cannot initiate starch phosphorylation but requires the prior action of GWD. PWD introduces phosphates at the C3 position. This phosphorylation is sufficient to loosen the tight crystal packing of glucans locally within the granule and to allow for degradation of amylopectin through the concerted action of β-amylases (Scheidig *et al.*, 2002) and a specialized form of direct debranching enzyme named isa3 (Edner *et al.*, 2007; reviewed in Fettke *et al.*, 2009). It is suspected but not proven that other enzymes of starch catabolism may be active at this stage (α-amylase and phosphorylases). The phosphate is then released through the action of sex4, a phosphatase which is functionally equivalent to laforin (see below for definition), but with a different organization of starch binding and phosphatase domains (Gentry *et al.*, 2007; Kotting *et al.*, 2009). β-amylase is an exo-hydrolase producing maltose processively from the non-reducing end of an amylopectin cluster. The maltose cannot be degraded by plastidial enzymes but will be exported by a specialized transporter named mex (maltose export) to the cytosol (Niittylä *et al.*, 2004). In the cytosol, the maltose will be metabolized thanks to the action of dpe2 (Fettke *et al.*, 2009). Dpe2 is an α-1,4 glucanotransferase that will cleave the α-1,4 linkage of maltose with concomitant transfer of one glucose residue to a required acceptor glucan. The acceptor glucan is believed to be a cytosolic heteroglycan whose outer chains consist of α-1,4 linked glucose residues. A cytosolic phosphorylase is thought to degrade these outer chains and thereby release glucose-1-P in the presence of orthophosphate (Fettke *et al.*, 2009).

All components of the starch degradation machinery with the noticeable exception of isa3 are either of eukaryotic phylogeny or of unknown phylogeny (such as α-amylase and pullulanase) (Deschamps *et al.* 2008*a*). Isa3 itself is not of clear cyanobacterial phylogeny and may be more related to Chlamydiae than cyanobacteria (see below). On the whole, the degradation pathway is completely unrelated to polysaccharide degradation in bacteria. In addition, there is no indication that storage polysaccharides are phosphorylated in cyanobacteria and no equivalent to GWD, PWD, β-amylase, and dpe2 can be observed in extant bacteria and cyanobacteria. In short, Chloroplastida display exceedingly complex pathways of starch synthesis and degradation that only very superficially resembles cyanobacterial storage polysaccharide metabolism. Phylogenetically, the pathways define a mosaic of enzymes of distinctive host and cyanobacterial origin (Coppin *et al.*, 2005; Patron and Keeling, 2005; Deschamps *et al.*, 2008*a*) (Table 1).

The very simple pathways of floridean starch synthesis and degradation

Much less is known about the pathway of starch synthesis and degradation in the two other Archaeplastida lineages:

the Rhodophyceae and Glaucophyta. In both instances, starch accumulates in the cytosol of these organisms. A growing body of biochemical and molecular evidence point to the existence of a UDP-glucose-based pathway both in Rhodophyceae (for review see Viola *et al.*, 2001) and Glaucophyta (Plancke *et al.*, 2008). A UDP-glucose pathway is also suspected to be at work in those lineages that are thought to be derived from red algae through secondary endosymbiosis, such as the dinoflagellates, apicomplexa parasites, and cryptophytes (Coppin *et al.*, 2005; Deschamps *et al.*, 2006; Deschamps *et al.*, 2008*d*). Rhodophyceae are very poor biochemical and genetic models and no starch accumulating red alga can fulfil the prerequisite to become an efficient system allowing for the functional dissection of starch metabolism. The only exception to this pessimistic view comes from the study of the secondary endosymbiont *Crypthecodinium cohnii* (Deschamps *et al.*, *2008d*; Dauvillée *et al.*, 2009). This homothallic heterotrophic dinoflagellate species does allow for the selection of mutants and crossing. Mutants of *Crypthecodinium* have very recently been reported that have decreased starch amounts and (or) a modified polysaccharide structure (Deschamps *et al.*, 2008*d*; Dauvillée *et al.*, 2009). Severely impaired mutants of *C. cohnii* were demonstrated to have a decreased and modified UDP-glucose requiring starch synthase (Dauvillée *et al.*, 2009). The defect in starch amount and the alteration in amylopectin structure cosegregated in crosses with the modification in enzyme activity. Because no other assayable enzyme of starch metabolism was affected in these mutants, we believe this brings functional proof that floridean starch, in this case, is indeed synthesized through the UDP-glucose substrate.

Four Rhodophyceae genomes have recently been sequenced including two unicellular cyanidiales and two complex multicellular species (Matsuzaki *et al.*, 2004). As with the Chloroplastida, the pathways are very well conserved throughout the lineage. The gene content is displayed in Table 1. The most striking feature of the red lineage pathway is the paucity of enzymes required to synthesize and mobilize starch. Fewer than 12 genes seem required to operate starch metabolism making it no more complex than glycogen metabolism. Yet Rhodophyceae do accumulate complex starch granules with all the major features found in Chloroplastida starch. Some red alga lineages such as the Porphyridiales also accumulate amylose at variance with the initial report that floridean starch lacked this fraction (Nakamura *et al.*, 2005; Shimonaga *et al.*, 2007). This very important result proves that a complex pathway is not required to explain the biogenesis of the starch granule architecture.

Another striking feature of the pathway is that, with the noticeable exception of the enzymes producing or using the nucleotide sugar substrate, all other steps of starch synthesis and degradation are analogous in Rhodophyceae and Chloroplastida. Indeed, phylogenetic trees show a common origin for all enzymes of starch metabolism in complete agreement with the monophyletic nature of Archaeplastida (Coppin *et al.*, 2005; Patron and Keeling, 2005; Deschamps *et al.*, 2008*a*; Plancke *et al.*, 2008). The only major difference is defined by the absence of ADP-glucose

Table 1. The number of isoforms found for each class of glycogen/starch metabolism enzymes

Using phylogenetics, it was possible to determine the origin of each isoform in the red and green lineages except for GWDs. Enzymes of cyanobacterial phylogeny are highlighted in blue. Enzymes of eukaryotic origin are highlighted in beige and those that were presumably transferred by Chlamydia are displayed in pink. Enzymes of uncertain origin are shaded in grey. The cyanobacteria, eukaryotes, and green plants display little variations in the corresponding sets of enzymes. *Crocosphaera watsonii*, *Entamoeba histolytica*, *Cyanidioschizon merolae*, and *Ostreococcus tauri* were chosen as paradigm genomes for, respectively, cyanobacteria, heterotrophic eukaryotes, red algae, and green plants. Reconstruction of starch metabolism in the common ancestor is emphasized within the black box and has been performed as explained in the text.

Activity	Cyanobacteria (*Crocosphaera watsonii*)	Eukaryotes (*Entamoeba histolytica*	Common ancestor	Green lineage (*Ostreococcus tauri*)	Red lineage (*Cyanidioschizon merolae*)
ADP-glucose pyrophosphorylase	1	0	1	2	0
Soluble starch synthase (ADPG) SSIII-SSIV	2	0	1	3	0
Soluble starch synthase (ADPG) SSI-SSII			1	2	
Soluble starch synthase (UDPG)	0	1	1	0	1
GBSS I	1	0	1	1	1
Branching enzyme	3	1	1	2	1
Isoamylase	1	0	2	3	2
Indirect debranching enzyme	0	1	0	0	0
Phosphorylase	2	2	2	2	1
Glucanotransferase	1	0	1	1	0
Transglucosidase	0	2	1	1	1
β-Amylase	0	4	1	2	1
Glucan water dikinase	0	?	1	3	1
Phosphoglucan water dikinase	0	?	1	2	1
Laforin or Sex4 type phosphatases	0	1	1	1	1

pyrophosphorylase and of the cyanobacterial type of GT5 ADP-glucose requiring starch synthase in Rhodophyceae and Glaucophyta. However, GBSS the enzyme of cyanobacterial phylogeny responsible for amylose synthesis within granules is present in Glaucophyta, Porphyridiales red algae, in cryptophytes and in dinoflagellates (Plancke *et al.*, 2008; Deschamps *et al.*, 2006, 2008*a, d*; Shimonaga *et al.*, 2007). In addition the floridean starch GBSS shows a marked preference for UDP-glucose while remaining capable of using ADP-glucose in Glaucophyta, cryptophytes, and Porphyridiales.

The soluble starch synthase used by the Rhodophyceae for amylopectin synthesis seems to be unique (no other candidate genes are found in these genomes) and to consist of the GT5 type of glycogen synthase found in many eukaryotic lineages distinct from the opistokonts (Deschamps *et al.*, 2008*a*). The same enzyme was found and its sequence cloned during the preliminary characterization of starch metabolism in the the glaucophyte *Cyanophora paradoxa* (Plancke *et al.*, 2008). Remarkably, this enzyme is thus able to fulfil all functions which, in Chloroplastida, seem to require four different soluble starch synthases. This enzyme was initially thought by Patron and Keeling (2005) to descend from the cyanobacterial GT5 enzymes. However, at the time of their study, these authors were unable to realize that, in fact, it represented one of the two major forms of glycogen synthase found in heterotrophic eukaryotes. All major steps of starch synthesis and degradation are represented by a single enzyme in the rhodophycean pathway (Table 1). The only interesting exception to this is defined by starch debranching enzyme (isoamylase) which is represented by two isoforms of bacterial phylogeny. Interestingly, in the glaucophyte *Cyanophora paradoxa*, isoamylase is known to be synthesized as a large size multimeric complex as in green plants and algae, suggesting that this enzyme may have a similar function to that proposed for Chloroplastida. The absence in the rhodophycean genome of dpe1 (D-enzyme), an enzyme required for the assimilitation of maltooligosaccharides longer than maltose, may suggest that, in this lineage, the other α-1,4 glucanotransferase (dpe2) possibly supplies an equivalent function in addition to its function in maltose assimilation.

The evolutionary origin of starch-like structures

The appearance of starch in Archaeplastida begs the question of the origin of this structure. Was a starch-like polymer synthesized before endosymbiosis by either the host or the cyanobacterium or did starch result accidently from the merging of related yet dissimilar pathways? The existence of such polymers in the eukaryotic ancestors seems highly unlikely. Indeed, this would suggest that, among the diversity of extant heterotrophic eukaryotes, one would expect several lineages unrelated to primary endosymbiosis to contain such polymers. However, each time a heterotrophic eukaryote was reported to contain starch-like polymers it turned out to define lineages which have lost photosynthesis either among Archaeplastida (the white algae such Polytomella, Polytoma, Prototheca, and Helicosoporidium (Hamana *et al.*, 2004; De Koning and Keeling, 2006; Pombert and Keeling, 2010) or among secondary endosymbiosis lines. The most striking case is defined by several apicomplexa parasite species such as *Toxoplasma gondii* which had been known for years to accumulate amylopectin granules (Coppin *et al.*, 2005). It was indeed subsequently found that apicomplexa harboured a cryptic plastid that resulted from the secondary endosymbiosis of an Archaeplastida ancestor. As to cyanobacteria, all species examined were reported to contain glycogen and no convincing report or claim of the presence of starch had appeared until very recently.

Because the enzyme responsible for generating the crystalline structure of starch displays a bacterial phylogeny and because GBSSI the only enzyme able to elongate glucans within the starch granule itself also displays a cyanobacterial origin, it remained possible that the cyanobiont's ancestor synthesized such polymers. Nakamura *et al.* (2005) were the first to report the existence of starch-like polymers organized into insoluble granules within one group of cyanobacteria which was named subgroup V according to the classification by Honda *et al.* (1999). Because, in their survey, they had not found bona fide large-size granules containing amylose they used the term semi-amylopectin to name this type of polymer.

Prior to this survey, the studies by Schneegurt *et al.* (1994, 1997) established that *Cyanothece* sp. strain ATCC 51142, another subgroup V cyanobacterium, synthesized a branched glucan which they thought represented a novel sort of glycogen molecule based on a measured branching ratio of 9%. They had, nevertheless, noted that the granule size exceeded the theoretical limits imposed on individual β-particles of glycogen and concluded that the granules contained several distinct glycogen molecules. Looking back on the data supporting this conclusion, we believe it is possible that chemical methylation would have yielded a slight overestimate of the branching ratio. Indeed a mere 20% overestimate would have been sufficient to turn an amylopectin-like candidate into a putative glycogen structure. In fact, *Cyanothece* sp. strain ATCC 51142 contains granules with a semi-amylopectin virtually identical to those reported by Nakamura *et al.* (2005) and Deschamps *et al.* (2008*a*). In their studies of nitrogen fixation in unicellular cyanobacteria, Schneegurt *et al.* (1994) noted that the carbodydrate granules were synthesized during the day and were being mobilized during the night. They also showed that nitrogen fixation occurred exclusively in darkness and was under circadian clock regulation. Nitrogenase, the enzyme of nitrogen fixation, is known to define an enzyme exquisitely sensitive to the presence of O_2 which inactivates it. Because cyanobacteria produce energy through oxygenic photosynthesis there is a conflict between energy production and its utilization for nitrogen fixation. Many cyanobacteria have resolved

this conflict through separating in space diazotrophy from oxygenic photosynthesis in distinct specialized cells within a multicellular filament. However, unicellular diazotrophic cyanobacteria of subgroup V are unable to do so and therefore have resorted to separate these processes in time through circadian clock regulation. Schneegurt *et al.* (1994) proposed that the energy stored in the carbohydrate granules is used both to supply the energy and reducing power required for nitrogenase and to lower the O_2 level further through respiration. Because diazotrophic unicellular cyanobacteria of subgroup V need to store significantly larger amounts of carbohydrates to feed cellular growth, division, and diazotrophy, Deschamps *et al.* (2008*a*) proposed that this yielded a selection pressure for the change of glycogen metabolism into the synthesis of semi-crystalline polymers. Indeed, this would enable the storage of larger amounts of osmotically inert carbon with a lower turnover during the light phase.

This could indeed explain the appearance of a starch-like structures in this particular taxonomic group which contains many important unicellular marine diazotrophic species. Interestingly Wing-Ming *et al.* (1994) also noted the presence of 'irregular polyglucan granules' in another subgroup V isolate, *Synechococcus* RF-1. However, they only noted the unusually large size of the granules without any detailed structural analysis.

Deschamps *et al.* (2008*a*) made a detailed structural characterization of the carbohydrate granules contained by a marine unicellular cyanobacterium Clg1 isolated by Falcon *et al.* (2004) related to both the genus *Cyanobacterium* and *Crocosphaera* both of subgroup V. Their attention was drawn by the presence of significantly larger granules than those present in *Cyanothece* sp. strain ATCC 51142. A very detailed characterization of the granules was made. Two polysaccharide fractions resembling amylopectin and amylose were purified with chain-length and mass distributions undistinguishable from the plant starch fractions. In addition, the granules displayed wide-angle powder X-ray diffraction patterns reminiscent of cereal starches (the so-called A-type diffraction pattern; Buléon *et al.*, 1998) demonstrating the presence of the same 3-D spatial organization of the amylopectin crystals. Moreover, small-angle X-ray scattering demonstrated the presence of the same 9 nm value that typifies the unit amylopectin cluster size (Deschamps *et al.*, 2008*a*). The carbohydrate granules of *Cyanothece* sp. strain ATCC 51142 also displayed an A-type diffraction pattern, further proving that this storage polysaccharide had properties much closer to amylopectin than glycogen (Deschamps *et al.*, 2008*a*). The presence of amylose in the Clg1 starch prompted Deschamps *et al.* (2008*a*) to look for the enzyme of amylose biosynthesis. GBSSI, an enzyme of cyanobacterial phylogeny never previously reported within cyanobacteria, was thus found bound to the starch granules and was demonstrated to synthesize amylose *in vitro* (Deschamps *et al.*, 2008*a*). Interestingly, GBSSI was more highly selective for ADP-glucose than the Archaeplastidal enzymes which, in most instances, proved to prefer either ADP-glucose or UDP-glucose but, nevertheless, were able to polymerize amylose from both. We believe this to reflect a distinct history of the Archaeplastidal enzymes. There is thus now enough evidence to support a cyanobacterial origin to starch. In addition, Deschamps *et al.* (2008*a*) clearly proposed that the plastid ancestor was indeed a cyanobacterial ancestor of subgroup V.

Reconstructing starch metabolism in the common ancestor of Archaeplastida

If one accepts a simple vertical inheritance model for the genes of starch metabolism, the monophyly of Archaeplastida allows for the reconstruction of a minimal gene set that must have been present in the ancestor of Archaeplastida to explain the present distribution of genes involved in storage polysaccharide metabolism in the three Archaeplastida lineages. This minimal gene set is displayed in Table 1. In reconstructing this set, Deschamps *et al.* (2008*a*) have minimized the number of genes originating from the green lineage to those that clearly displayed a common unique origin in phylogenetic trees, as we believe that most isoforms were generated by gene duplication when the Chloroplastidae and Rhodophyceae diverged. Table I also displays the phylogenetic (host, cyanobiont or unknown) origin of the pathway enzymes. We chose as a paradigm of the status of storage polysaccharide metabolism of the eukaryote host the enzyme network exemplified in *Entamoeba histolytica* (Loftus *et al.*, 2005). The relevance of this choice by the finding of a richer set and diversity of important enzymes in amoebas has previously been discussed. The starch metabolism network of *Crocosphaera watsonii* was chosen as a model subgroup V starch-accumulating diazotrophic cyanobacterium. Table 1 clearly shows that the starch metabolism network of Rhodophyceae and Chloroplastida define a very similar mosaic of enzymes of host and cyanobiont origin (Deschamps *et al.*, 2008*a*). In addition in phylogenetic trees, the common chloroplastidal and rhodophycean enzyme sequences display a common origin (Coppin *et al.*, 2005; Patron and Keeling, 2005; Deschamps *et al.*, 2008*a*). These observations are in complete agreement with Archaeplastida monophyly. The only difference consists of the presence of enzymes of ADP-glucose synthesis and utilization in Chloroplastida and the sole presence of the glycogen synthase from heterotrophic eukaryotes in Rhodophyceae (Table 1). However, the common ancestor must have contained all of these distinctive enzymes. Some enzymes, such as GWD, PWD, α-amylase, and pullulanase, have unknown phylogenetic origins.

Despite all efforts, it has not been possible to locate GWD or PWD-like sequences in lineages independent from the Archaeplastida. It is quite possible that this function evolved shortly after endosymbiosis in the host cytosol. GWD is responsible for amylopectin phosphorylation and thereby initiates starch degradation. There is good evidence for the presence of a glycogen phosphorylation pathway in heterotrophic eukaryotes and no evidence for such a pathway in cyanobacteria. The precise function of this pathway is unknown. Nevertheless, a dysfunctional phosphoglucan

phosphatase activity carried by a mutant laforin protein may help explain why highly phosphorylated anomalous glycogen (called Lafora bodies) accumulates during Lafora's disease in humans. This accumulation concerns many different tissues and organs, including the brain, thereby yielding a fatal progressive myoclonic epilepsy (Tagliabracci *et al.*, 2008). However, the nature of the enzyme responsible for glycogen phosphorylation in humans is unknown. Human laforin is known, on the other hand, to complement the defect in starch mobilization in *Arabidopsis* due to a mutation in the related sex4 protein (Gentry *et al.*, 2007; Kotting *et al.*, 2009). The sex4-laforin function is known to be required during starch mobilization to dephosphorylate the phosphoglucans generated by GWD (Kotting *et al.*, 2009). It is possible that GWD evolved because this unknown kinase may have been unsuitable to phosphorylate the more hydrophobic crystalline amylopectin clusters. The Lafora protein was then immediately recruited by GWD to establish a novel pathway of starch mobilization in the host cytosol. This invention can be seen as host mediated since it appeared in the cytosol and entirely relies on other components of host phylogeny such as laforin and β-amylase. Clearly, a better knowledge of the function of laforin in glycogen metabolism is required before one can suggest useful scenarios for the appearance of this very important and intriguing pathway.

Subcellular localization of storage polysaccharides in the common ancestor of Archaeplastida

The minimal ancestral enzyme set does not tell us where the enzymes were located shortly after endosymbiosis. Three mutually exclusive scenarios can be considered. In a first scenario, both the cyanobiont and the host cytosol synthesized storage polysaccharides, in a second scenario, only the cyanobiont synthesized such polysaccharides while, in a third scenario, only the cytosol contained this material. We argue that the third scenario defines the only plausible situation. There are several complex reasons for this that are outlined below.

First, Henrissat *et al.* (2002) after making a gene content survey of the genomes of several pathogenic bacteria noted that a strong correlation existed between glycogen metabolism loss and a highly dependent parasitic relationship with the infected host. Becoming an endosymbiont would, according to this view, automatically lead to the loss of storage polysaccharide metabolism by the symbiont. This seems largely confirmed by surveying most endosymbiont genomes (Gil *et al.*, 2004). There is only one known example in the literature which is independent of primary endosymbiosis of the plastid that is based on photosynthate export from a cyanobacterium engulfed by a protist. *Paulinella* carries two cyanobionts (called chromatophores) which are replicated with the host and cannot live as independent organisms (Bodyl *et al.*, 2007). The chromatophore genome has recently been sequenced (Nowack *et al.*, 2008). It apparently still contains over 1000 genes and displays a typical cyanobacterial cell morphology and organization. Interestingly, another heterotrophic species related to *Paulinella* that does not shelter chromatophores seems to prey on cyanobacteria that ressemble the latter. This observation pleads for a phagotrophic origin of these symbionts.

The *Paulinella* genome was proven to lack enzymes of glycogen metabolism while these are universally present in the genomes of free-living cyanobacteria (Nowack *et al.*, 2008). Although one might argue that these genes could have been, at least in theory, transferred to the host nucleus, the present gene content of the chromatophore does not support transfers as extensive as those observed during the evolution of plastids, although some of the experienced gene losses would have prevented independent life of the chromatophore.

A second argument comes from a close examination of the phylogenetic origin of the minimal ancestral enzyme set displayed in Table 1. This ancestral metabolism includes a nearly complete set of host glycogen metabolism (with the noticeable exception of indirect debranching enzymes which were substituted by direct DBEs). However, important enzymes of starch biosynthesis and degradation in cyanobacteria are missing. This involves three branching enzymes and two phosphorylases and possibly more if it is considered that isoamylases and soluble starch synthases may be of chlamydial rather than cyanobacterial origin (see below). If the gene losses occurred before routine targeting of cytosolic proteins evolved then the cyanobiont would not have been able to sustain storage polysaccharide metabolism. On the other hand, if the losses occurred after such a system became routine, one can argue that a duplicated branching enzyme gene of host origin may have had its product targeted to the cyanobiont and substituted for the three cyanobacterial enzymes which were subsequently lost. In addition, the starch phosphorylase would have required a similar replacement by an enzyme of host origin. In this case, it could be argued that these enzymes are certainly not functionally equivalent to the cyanobacterial phosphorylase. Indeed, the latter seems able to attack solid cyanobacterial starch directly and release glucose-1-P, a feat which cannot be achieved by the enzymes of eukaryotic origin (Dauvillée *et al.*, 2006).

A third argument comes from the unexplainable complexity of the Chloroplastida starch metabolism network (Table 1). Why many enzymes, but not all, have experienced in green algae one to two rounds of gene duplications followed by enzyme subfunctionalizations remains a complete mystery. Rhodophyceae which contain equally complex starch granules have not undergone such gene duplications and subfunctionalizations (Table 1). It must be stressed that, in the earliest diverging green algae, only the starch pathway seems to stand out by a high level of functional redundancy and subfunctionalizations. Other pathways have indeed not yet reached the level of complexity that can sometimes be found in terrestrial plants. There are only two alternatives to be considered at this point: either starch synthesis was exclusively cytosolic as we proposed or storage polysaccharides were synthesized both in the plastid and cytosol in contradiction to our hypothesis. In the first

case, the cyanobiont would, very early, have lost its ability to store starch, while, in the second, the latter would have been maintained.

If on the one hand, one assumes that the cyanobiont has lost the ability to store starch, then the plastidial localization of starch in green algae will have to be generated through a rewiring of the whole starch metabolism network from the cytosol to the plastid. In following sections, the means by which this could be achieved will be detailed. Suffice to say now that this process does explain the amount of gene duplications and enzyme subfunctionalizations that is seen in the Chloroplastida network.

If, on the other hand, storage polysaccharide metabolism had been maintained in the cyanobiont then the transfer to the nucleus of the genes required for their synthesis and degradation would have occurred one gene at a time. If the transferred gene by chance acquired a transit peptide then the corresponding gene on the cyanobacterial genome would have been lost. This process which is similar to that experienced by Calvin cycle genes or any other photosynthesis genes whose products has remained in the cyanobiont does not require any gene duplication and enzyme subfunctionalization and these are indeed not observed as extensively in such networks.

Because the selective complexity of the starch metabolism network (at variance with land plant pathways, only starch metabolism displays a high level of gene redundancy in green algae) can only be explained through the complex rewiring mechanism that selectively takes place if the cyanobiont had lost the ability to store starch, we believe this pleads for an ancient cytosolic pathway.

Less convincing arguments than the three outlined above equally support the cytosolic localization of the ancient network. Among these is the fact that two out the three Archaeplastidal lineages still synthesize, today, their storage polysaccharides exclusively in the cytosol and that the lineage thought more closely to resemble the ancestral Archaeplastida (the Glaucophyta) is among them.

Compartmentalization of the ancient pathway of starch metabolism in the common ancestor of Archaeplastida

If the storage polysaccharides were located in the cytosol, then most of the enzymes of starch synthesis and degradation must have been expressed in this compartment. This includes both the UDP-glucose and ADP-glucose requiring starch synthases and all other enzymes of bacterial or host phylogeny. How could the enzymes initially encoded by the cyanobiont genome be expressed in the cytosol at such an early stage of endosymbiosis? The process by which genes are transferred from the cyanobiont genome to the host nucleus is called endosymbiotic gene transfer (EGT) and is one of the major causes of lateral gene transfers in eukaryotic genomes. To the naive reader this would define an unusual and mysterious phenomenon that is expected to occur at exceptionally low frequency. In fact, EGTs are likely to occur at quite high frequencies. In yeast, marker genes that allow growth only when expressed in the cytosol were introduced into the mitochondrial genome. Mutations at the corresponding nuclear locus were introduced that would lead to the absence of growth on selective media. Restoration of growth due to the transfer of the mitochondrial copy in the yeast nuclear genome was observed at the frequency of 10^{-6}, a frequency comparable with that of spontaneous mutations in a given gene (Thorsness and Fox, 1990). Similar experimental results with similar frequencies were obtained with plastidial markers in tobacco (Stegemann *et al.*, 2003). Of course EGTs followed by expression of a protein in the cytosol will probably be observed at lower frequencies, since the organelle DNA inserted in the nucleus must by chance be located downstream of active promoter sequences. Nevertheless, such results leave little doubt that these events were indeed sufficiently frequent. Of course, to explain that the nuclear DNA is not filled with organelle sequences, one has to imagine that most of these events will be counter-selected and that losses of such nuclear sequences will be at least as frequent.

Nevertheless, EGTs followed by cytosolic expression of proteins from the cyanobiont will define the first kind of EGT recorded and can take place immediately at the time of endosymbiosis (or even before if the host preys through phagocytosis on the future endosymbionts). The more classical type of EGT that requires the expression of the protein product in the cyanobiont (such as enzymes of photosynthesis) will have to await the later development of a complex protein targeting machinery able routinely to readdress the corresponding proteins to the evolving organelle. Thus, if the cytosolic expression of the cyanobacterial genes of starch metabolism gives some advantage the corresponding EGT will be selected and maintained. The problem therefore is to understand what would have been the advantage for the host cytosol to harbour a dual substrate biochemical pathway based on both ADP-glucose and UDP-glucose. Indeed if one assumes that ADP-glucose pyrophosphorylase, the enzyme of ADP-glucose synthesis was also subjected to EGT and the enzyme transferred to the cytosol, then it becomes very hard to understand what benefit the cell would get from having cytosolic glucose-1-P funnelled into both UDP-glucose and ADP-glucose. However, we argue that ADP-glucose pyrophosphorylase is a highly unlikely target for cytosolic expression since its substrate and allosteric effectors couples the enzyme tightly to photosynthesis. Indeed the regulation of the enzyme through the 3-PGA/Pi ratio has been maintained throughout evolution (Ballicora *et al.*, 2003). If the coupling of the enzyme activity to the Calvin cycle and to ATP production is maintained, then it is more reasonable to consider that it remained expressed in the cyanobiont from the cyanobacterial genome and that it did not define a likely candidate for an early EGT leading to its expression in the cytosol. Therefore, for the enzymes of cyanobacterial origin to be immediately functional for starch synthesis in the cytosol, one has to assume that the ADP-glucose produced within the cyanobiont was exported from the cyanobiont to

the cytosol. For such a transport to be effective Deschamps *et al.* (2008*a*) proposed that a nucleotide sugar translocator (NST) of host origin able to exchange ADP-glucose for AMP or ADP was targeted to the inner membrane. The flux created by this proposal is displayed in Fig. 5. It is the only hypothesis that yields a selective advantage for the presence of a cytosolic dual substrate pathway for starch biosynthesis. Indeed, the flux depicted in Fig. 5 is no more nor less than the biochemical flux through which photosynthate was exported from the cyanobiont to its host. This would thus be the carbon flux on which the success of plastid endosymbiosis would rest.

The proposed flux of carbon through starch metabolism explains the establishment of plastid endosymbiosis

At the core of plastid endosymbiosis lies a biochemical link whereby the carbon produced through photosynthesis by the cyanobiont is exported to the eukaryotic cytosol to feed the host metabolism. This link was essential for the establishment of the plastid and introduction of oxygenic photosynthesis in eukaryotes. The problem with the establishment of such a link relies on the fact that both partners of endosymbiosis are both completely unrelated and independent implying that there is no cross-talk or possibility of regulation between the two unrelated biochemical networks. Yet the carbon must be exported only when the cyanobiont is able to supply photosynthate and used only when the host needs it. If carbon flows out the cyanobiont in the form of just any metabolite translocated at any time because of the targeting of some host transporter on the cyanobiont inner membrane then the net result for the homeostasis of the cyanobiont's metabolism is likely to be toxic and, at worst, lethal. A sudden burst of metabolites originating from the cyanobiont's metabolism into the host cytosol can, depending on the nature of the latter, also affect negatively the host physiology. Clearly some buffer is needed between the two unrelated

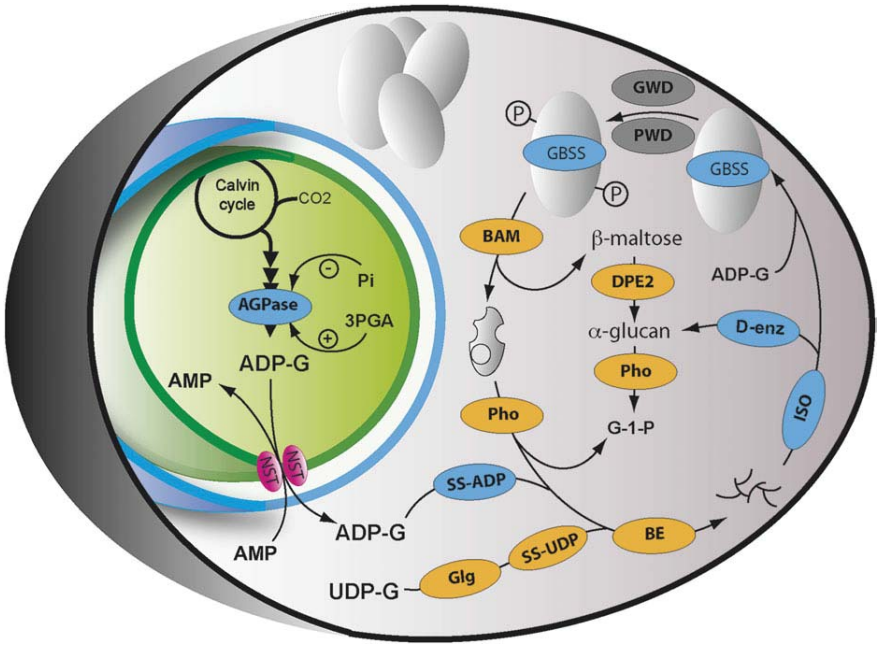


Fig. 5. Reconstruction of starch metabolism in the common ancestor of Archaeplastida. The cyanobiont is displayed in green with an emphasis on the inner (solid green line) selective membrane. The outer membrane is represented by a solid blue line and is thought to be much less selective and permeable to most metabolites including sugar-nucleotides. Enzymes are coloured with respect to their phylogenetic origin as bacterial (blue), host (orange), or unknown (grey) with the following symbols: AGPase, ADP-glucose pyrophosphorylase; SS-ADP, ADP-Glc requiring soluble starch synthase; SS-UDP, UDP-Glc requiring soluble starch synthase; Glg, glycogenin; Pho, phosphorylase; BE, branching enzyme; ISO, isoamylase; D-enz, disproportionating enzyme (α-1,4 glucanotransferase); DPE2, amylomaltase; BAM, β-amylase; GBSS, granule-bond starch synthase; GWD, glucan water dikinase; PWD, phosphoglucan water dikinase; NST, nucleotide sugar translocator. The carbon fixed by photosynthesis enters the Calvin cycle in the cyanobiont. The carbon committed to storage flows selectively through ADP-glucose pyrophosphorylase. The ADP-glucose accumulates in the cyanobiont to high levels because of the absence of storage polysaccharides within the latter. The ADP-Glc is exported in exchange for AMP and orthophosphate by an NST recruited from the host endomembrane system as described in Colleoni *et al.* (2010). Within the cytosol, the exported photosynthate is polymerized into starch thanks to the ADP-glucose-specific soluble starch synthase. The carbon is then mobilized through a completely host-driven polysaccharide degradation machinery following host needs. This machinery involves the GWD–PWD starch phosphorylation pathway followed by β-amylase digestion and maltose degradation. Degraded starch is symbolized by a smaller eroded and pierced granule.

biochemical networks because of their asynchronous supply and demand for carbon. It is very hard to imagine a flux better adapted to such constraints than that displayed in Fig. 5. ADP-glucose is one of the very few metabolites that can freely flow out of the cyanobiont without affecting it. Indeed, the carbon flowing through ADP-glucose pyrophosphorylase is that part of bacterial metabolism which is committed to storage, that is to escape cyanobacterial metabolism and become at least temporarily unavailable in solid starch. Translocating ADP-glucose out of the cyanobiont and having the carbon stored in the cytosol will make very little difference. In addition, as ADP-glucose enters the cytosol it has virtually no impact on host metabolism since this glycosyl-nucleotide is not produced nor recognized by eukaryotes. Because of this absence of recognition, ADP-glucose's only fate will be to be used through a bacterial enzyme that has been transferred by EGT to the host cytosol. The GT5 ADP-glucose requiring starch synthase will transfer glucose to the non-reducing end of the outer chains of storage polysaccharides and release ADP. ADP will re-enter the cyanobiont through the exchange reaction mediated through the ADP-glucose NST working on the cyanobiont's inner membrane. This will either be in the form of ADP or AMP since most known NSTs are documented to exchange the glycosylnucleotide with the corresponding nucleotide monophosphate. However, in the case of the exchange with AMP, the question of the missing orthophosphate's fate remains to be fully addressed. The cycle indeed needs to be balanced with respect to both purine nucleotides and orthophosphate to result in the net incorporation of one glucose residue to a cytosolic storage polysaccharide. This glucose comes from carbon fixation from the cyanobiont while the energy required to polymerize it into storage comes from cyanobacterial photophosphorylation responsible for producing ADP-glucose, the activated form of the carbohydrate.

Having the carbon stored in glycogen or starch will efficiently buffer demand and supply for carbon. Indeed the pathway of storage polysaccharide breakdown is entirely under the control of host enzymes and will therefore respond specifically to host demand. We believe that having both the first ADP-glucose requiring starch synthase working in the cytosol together with the ADP-glucose NST targeted to the inner cyanobiont membrane would have been sufficient successfully to establish the endosymbiotic link. This would have initially led to the synthesis of additional amounts of cytosolic glycogen. The switch to starch would have been brought about later thanks to EGTs of other cyanobacterial or chlamydial genes (see below) such as isoamylase responsible for the synthesis of crystalline amylopectin. This was followed by EGT of GBSSI a cyanobacterial gene responsible for amylose synthesis (an enzyme initially specific for ADP-glucose) within the solid starch granule (GBSSI is unable to use glycogen; Dauvillée *et al.*, 1999). These EGTs were facilitated by the fact that the endosymbiont was related to extant subgroup V cyanobacteria and thus contained the required genes. However, during this whole process, the cyanobiont itself remained unable to synthesize storage polysaccharides possibly because it had already lost critical genes such as the starch branching enzymes and starch phosphorylases.

The host, however, remained able to feed carbon into glycogen or starch from UDP-glucose. Indeed cytosolic glycogen was initially present in the host to enable it to siphon out of cellular metabolism the excess carbon produced through host metabolism for later use. This would have remained important since the host probably adopted first a mixotrophic cycle involving both phagocytosis and the use of photosynthate provided by its endosymbiont. Phagocytosis is expected to impact specifically and dramatically storage polysaccharide metabolism. Indeed large amounts of carbon need to be stored in a relatively short time in an otherwise oligotrophic environment. It is thought that the host has evolved complex regulatory mechanisms impacting the UDP-glucose specific glycogen synthase and triggering glycogen accumulation under such circumstances. The phagotroph model heterotrophic dinoflagellate *Crypthecodinium cohnii* was recently demonstrated, unlike other non-phagotrophic microorganisms such as yeasts or bacteria, to accumulate glycogen during the log phase as soon as the substrate becomes available and not at the transition between the log and stationary phases (Deschamps *et al.*, 2008*d*). This behaviour was interpreted as a response to nutrient bursts mimicking the consequences of prey ingestion. This useful physiological function and the concomitant regulation were maintained throughout and explains the presence of a dual substrate pathway.

The change from cytosolic glycogen to starch was brought about because semi-crystalline polysaccharides were better adapted to trap the vast amounts of glucose that became available through endosymbiosis. This also ensured that a permanent strong carbon sink existed in the cytosol even if vast amounts of glucose had already been stored. The switch to starch required the only innovation in the storage polysaccharide pathways that came about during plastid endosymbiosis: the evolution of GWD. This innovation would have been required if one considers that cyanobacterial starch phosphorylase genes which, unlike the corresponding eukaryotic enzymes, display the ability to digest solid starch, had already been lost. As mentioned earlier this innovation was built on a pre-existing eukaryotic pathway of glycogen phosphorylation whose precise function in glycogen metabolism requires further research (Tagliabracci *et al.*, 2008).

Discovering the missing link of eukaryotic photosynthesis

When the flux displayed in Fig. 5 was first proposed, there was no evidence for the existence of an ancient ADP-glucose translocator. Nevertheless, this ancient NST would have been central to the success of endosymbiosis as it established the first link between the endosymbiont and its host. Weber *et al.* (2006) published phylogenies of the major carbon transporters of the inner plastid membrane of

Rhodophyceae and Chloroplastida. These plastidial phosphate translocators (pPTs) exchange a number of distinct triose-phosphates, hexose-phosphates, xylulose-phosphate, and PEP for orthophosphate. All these diverse pPTs from both Rhodophyceae and Chloroplastida can be traced back to a unique ancestor that was proposed to define the translocator that was responsible for supplying the host cytosol with photosynthate from the cyanobiont during endosymbiosis (Weber *et al.*, 2006). The common ancestor of the pPTs defined a sister lineage to a group of transporters from the eukaryotic host endomembrane system. Weber *et al.* (2006) thus suggested that these host transporters were recruited and targeted to the cyanobiont inner membrane to establish the endosymbiotic link. However, the eukaryotic endomembrane transporters all defined nucleotide sugar translocators (NSTs) that exchanged glycosyl-nucleotides for the corresponding nucleotide monophosphate and not sugar-phosphate-orthophosphate translocators (Rollwitz *et al.*, 2006). In addition, the particular NST3 group concerned (family NST3 according to Martinez Duncker *et al.*, 2003) defines the only NST family reported to contain, among others, purine sugar nucleotide transporters such as GDP-mannose, GDP-arabinose or GDP-fucose. It thus occurred to us that if the carbon translocator responsible for establishing the endosymbiotic link was so close phylogenetically to purine sugar nucleotide transporters then it might very well have defined the ADP-glucose translocator required by the flux proposed in Fig. 5. NSTs, in general, display high selectivity with respect to the nature of the base and modest selectivity with respect to the sugar (Handford *et al.*, 2006). ADP-glucose is not reported in eukaryotes, but other purine-nucleotide sugars such as GDP-mannose, GDP-arabinose or GDP-fucose are common. If such a GDP-sugar translocator displayed a fortuitous ability to translocate ADP-glucose, it could have been recruited on the cyanobiont's inner membrane during endosymbiosis. This would imply that if ancient NST3-like translocators displayed such properties then maybe extant NST3 transporters should still display them. Colleoni *et al.* (2010) very recently investigated the kinetic properties of two GDP-mannose transporters from yeast and plants in yeast membrane-derived liposomes. They demonstrated that AMP acted at physiological concentrations as a very potent competitive inhibitor of both the yeast and the plant GDP-mannose translocator. In addition, they demonstrated that the plant enzyme was able to exchange ADP-glucose for AMP at rates similar to those of the GDP-mannose GMP exchange (Colleoni *et al.*, 2010). Nevertheless, the plant enzyme displayed a relatively high K_m for ADP-glucose (7 mM). Would such a low affinity for ADP-glucose have allowed the export of photosynthate from the cyanobiont? In the discussion of their findings Colleoni *et al.* (2010) argue that, on the periplasmic side of the cyanobiont's inner membrane, the high sensitivity of the translocators to AMP would have been sufficient to outcompete GMP and GDP-mannose. On the cyanobiont's side, the authors argue that ADP-glucose would have defined the only purine nucleotide sugar available. They also emphasize that the concentrations of the latter would have been likely to exceed 1 mM because of a block in ADP-glucose utilization. Indeed, having lost the ability to synthesize storage polysaccharides, the cyanobiont can be compared to a mutant blocked in the utilization of ADP-glucose. Therefore, the present biochemical properties of Golgi GDP-mannose translocators are sufficient to generate an efflux of ADP-glucose, provided, of course, the protein is correctly targeted to the symbiont's inner membrane (see below).

The results reported by Colleoni *et al.* (2010) do not imply that the ancestors of the corresponding GDP-mannose translocators defined the hypothetical missing link for the establishment of eukaryotic photosynthesis. In fact, probably any NST3 purine nucleotide sugar transporter would have been suited for this purpose provided it displayed a sufficiently high affinity for ADP-glucose. How then, can the disappearance of this translocator and its replacement by the pPTs be explained? Colleoni *et al.* (2010) proposed that, as the metabolic integration of the evolving plastid proceeded, an opportunity arose to duplicate the genes corresponding to existing plastid inner membrane transporters and to evolve more integrated routes of carbon exchange through the evolution of novel activities. The ADP-glucose translocator and its derived pPT family of transporters co-existed as long as starch was being polymerized in the cytosol from ADP-glucose. However, when the three Archaeplastida lineages diverged, this ability was lost by the Rhodophyceae and Glaucophyta when both lineages lost the ability to elongate starch from ADP-glucose. As to the Chloroplastida, the ADP-glucose transporter was lost at the final stage of starch metabolism rewiring to the chloroplast.

Early targeting of the ADP-glucose translocator to the cyanobiont inner membrane

One of the essential requirements to be met for the first carbon translocators was that they should be correctly targeted to the cyanobiont's inner membrane. Most researchers agree that the present sophisticated machinery of protein targeting to plastids took time to evolve. Yet a novel set of proteins probably needed to find its way to the endosymbiont's inner membrane to tap photosynthate into the periplasm. The latter would then leak to the cytosol, the outer membrane being much less selective.

Several options can be considered at this stage. However, one should also consider, in addition, that the the plant TPT (triose phosphate translocator) localizes to mitochondrial membranes when expressed in yeast without its transit peptide (Loddenkötter *et al.*, 1993). This could suggest that whatever mechanism or built in structural features the TPT uses to localize to mitochondrial membranes, these may also have been working for the localization of the TPT ancestors to the cyanobiont's inner membrane.

There is presently no consensus on the nature of early protein targeting to the cyanobiont or the cyanobiont inner membrane. Some authors support vesicular transport

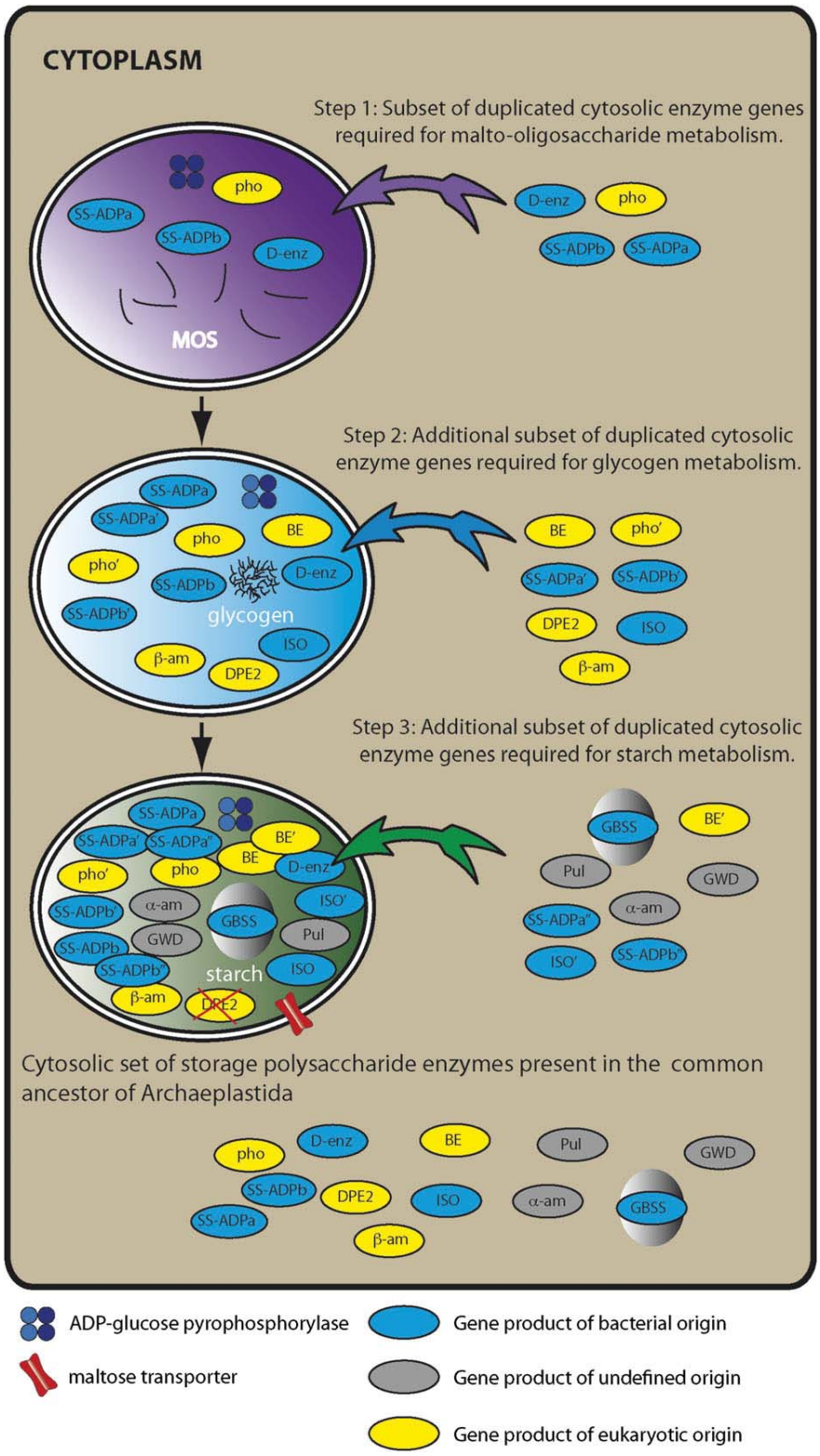


Fig. 6. Rewiring of starch metabolism in the emerging Chloroplastida. As detailed in Deschamps *et al.* (2008*c*), evolution of novel chlorophyll-rich LHCs in Chloroplastida possibly increased the photooxidative stresses experienced by the evolving chloroplasts. To obviate some of these stresses the presence of storage polysaccharides or oligosaccharides within the plastid became desirable. Because the rewiring of a 12 component network in a single step defines a very unlikely event we propose that this evolution involved two workable intermediate stages defined by the accumulation of a small pool of maltooligosaccharides and a larger pool of glycogen which co-evolved with the increasing complexity and requirements of the LHC antennae (symbolized by the purple phycobilisome

(Villarejo *et al.*, 2005) while others propose an early origin for a simplified version of the TIC-TOC import machinery (Bodyl *et al.*, 2009). Indeed, to target a protein to the cyanobiont inner membrane through a primitive TOC system would have only required an accidental duplication of a suitable NST gene that was accompanied by the substitution of the signal sequence by a pre-existing mitochondrial leader-like sequence. The primitive TOC system could have been defined by the sole cyanobacterial Omp85 protein whose orientation was reversed in the endosymbiont's outer membrane, as initially suggested by Bolter *et al.* (1998) and reviewed by Bodyl *et al.* (2009). This reversion would have been readily generated by the cytosolic expression of a copy of Omp85 that was transferred to the host nucleus (Bodyl *et al.*, 2009). Despite the fact that the recruited NST originally came from the host endomembrane system, we favour the primitive TOC hypothesis. This seems to be suggested by the fact that all extant pTPs are nowadays targeted by the TIC-TOC machinery. This primitive version of TOC would have been sufficient to direct a protein to the cyanobiont periplasm or the inner membrane. This suggests that both recruitment of the endomembrane NST and inversion of Omp85 were very early events, since both of these were simultaneously required to initiate endosymbiosis. It also implies that gene duplication of the recruited NST was a prerequisite for its use on the cyanobiont's inner membrane, since it had to be transported by an altogether different route than that taken by the original transporter targeted to the ER and Golgi.

Another possibility would be that the transporter was targeted to plastids through a selective modification of vesicular transport, a mechanism which is considered by some to have defined the first kind of transport machinery (Villarejo *et al.*, 2005). In this respect, the ER retention signal found in many pPT proteins (Martinez-Duncker *et al.*, 2003; Colleoni *et al.*, 2010) might have played a specific role to avoid the transporter being directed to the Golgi which defines the home of all purine nucleotide sugar transporters presently documented. However, in this case, the duplicated copies of the ancestral transporter would have switched later to transport through the TIC-TOC protein import machinery.

The rewiring of starch metabolism to chloroplasts involves extensive gene duplications and enzyme subfunctionalizations

The simple cytosolic pathway depicted in Fig. 5 generated the corresponding simple floridean starch synthesis and degradation pathways through the selective losses of ADP-glucose pyrophosphorylase, of the ADP-glucose translocator, and of the ADP-glucose specific starch synthases. Indeed the rhodophycean pathway is characterized by one isoform for each step of starch synthesis and degradation (Table 1). A simple pathway was also observed in secondary endosymbiosis lineages such as the apicomplexan parasites (Coppin *et al.*, 2005) and is suspected to be at work in the Glaucophyta (Plancke *et al.*, 2008). The Chloroplastida on the other hand experienced a rewiring of starch metabolism back to the compartment where it originated: the chloroplast (Deschamps *et al.*, 2008*b*). The means by which a gene will have its product rewired to plastids are quite clear: it must by chance acquire a transit peptide or another adequate targeting sequence and it will then be expressed in the organelle. However, an isolated enzyme from a biochemical pathway rewired by chance to a novel location has little chance to yield a benefit to the cell. Most likely its isolated presence will yield a penalty and will be counter-selected. The means by which an entire pathway becomes targeted to plastids are thus obscure.

This problem was first raised by Michels and Opperdoes (1991) when they were trying to understand how sections of the glycolytic pathway could have been rewired to the peroxisome, thereby generating the glycosome in trypanosomes. Nevertheless, biochemical pathways do get redirected to novel locations. As Martin (2010) recently pointed out, it is reasonable to assume that dual targeting and mistargeting of small amounts of enzyme activities could explain the means by which pathways can be reconstructed since an entirely isolated enzyme would thus not exist. It was argued that the redirected enzyme will find minute yet sufficient amounts of mistargeted partners from its original pathway to ensure maintenance of the redirected enzyme until the next wave of transit peptide acquisitions (Martin, 2010). We agree that very simple pathways consisting of

containing plastid, the blue-green intermediate with both LHCs and phycobilisomes and the green chloroplast). Each stage is thought to have lasted long enough for the rewired enzymes to have become optimized with respect to the type of products they were making. The movement from one stage to another (from oligosaccharides to glycogen or from glycogen to starch) can be initiated through a single (for instance BE to generate glycogen from the pre-existing MOS metabolism machinery) or at most two gene duplications (for instance starch synthase and phosphorylase to intiate MOS metabolism from ADP-glucose in the cyanobiont) with concomitant acquisition of transit peptides. The nature of the network desirable at each stage is deduced from that exemplified in bacteria for maltooligosaccharides and in amoebas for glycogen (although the elongation enzyme would be defined by the ADP-glucose specific starch sytnhases). The cytosolic network of the archaeplastidal ancestor is displayed at the bottom of the figure with enzymes of bacterial and host origin respectively in blue and yellow. This cytosolic network will be finally lost (with the exception of cytosolic phosporylase and DPE2) at the very last stage of the evolution process when the maltose transporter appeared (in red) and the plastidial DPE2 was lost. The rationale for enzyme subfunctionalization is detailed in the text.

Table 2. The identity and number of isoforms of the starch metabolism network rewired to plastids at each stage (MOS, glycogen, and plastidial starch) of the reconstruction are indicated in blue

To obtain the final theoretical numbers displayed in red, the number of plastidial enzymes (in blue) transferred at each stage must be summed together and with the number of enzymes of the ancestral cytosolic network (in orange) taking into account the selective losses that occurred mostly when cytosolic starch was finally lost. Gene losses are symbolized by the minus symbol followed by the number of genes lost that is coloured according to the cytosolic (orange) or plastidial (blue) localization of the product lost. Innovations are symbolized by the + symbol followed in bold by a number that is coloured according to its cytosolic (orange) or plastidial localization. The identity of the enzymes transferred at each stage corresponds to the catalytic activities that are known to be at work for MOS metabolism in bacteria, for glycogen metabolism in eukaryotes, and for starch metabolism. When several isoforms of the ancestral cytosolic network correspond to an activity required at a particular stage, it is assumed that the genes corresponding to all isoforms will be duplicated and their products targeted to the organelle. GWD, dpe2, and β-amylase are considered as enzymes of the eukaryote glycogen metabolism network either on the basis of their existence in extant non-opistokont eukaryotes or on the basis of the existence of a glycogen phosphorylation pathway in eukaryotes.

Activity	Ancient cytosolic Starch network	MOS network	Glycogen network	Plastidial starch network	Theoretical distribution	*O. taurii* network
ADP-glucose PPase	1	0	0	0	1	2
Starch synthase	2	2	2	2–2	6	5
GBSS I	1	0	0	1–1	1	1
Branching enzyme	1	0	1	1–1	2	2
Isoamylase	1	0	1	1–1	2	3
Phosphorylase	1	1	1	0	3	2
Glucanotransferase	1	1	0	0–1	1	1
Transglucosidase	1	0	1	–1	1	1
b-amylase	1	0	1	1–1	2	2
GWD	+1	0	1	1–1	2	2
MEX	0	0	0	+1	1	1

a few distinct steps where a defined balance of enzyme activities is, in addition, not critical may be rewired in this fashion. However, we disagree that such a process would apply to starch metabolism as a whole, although it might apply to sections of it. First of all, starch metabolism defines a non-linear biochemical pathway where, for instance, the product of elongation becomes a substrate for branching which regenerates a different type of substrate for the elongation enzyme in a more cyclic than a linear fashion (Ball and Morell, 2003). Polysaccharide debranching also produces substrates for further elongation and branching and debranching (Ball and Morell, 2003). In addition, the balance of enzyme activities is certainly critical and the presence of multienzyme complexes may be required to ensure that the right stochiometry and balance of enzyme activities is indeed achieved (Tetlow *et al.*, 2004*b*; Hennen-Bierwagen *et al.*, 2008).

Everything we know concerning starch metabolism tells us that the simultaneous acquisition of a transit peptide by all the required enzymes and the achievement of the proper activity balance in one step are absolutely impossible. Yet unlike most biochemical pathways, the same set or subset of enzymes can yield several very different end-products that are indeed found in different lineages. In fact, storage polysaccharides can be found either in the form of unbranched malto-oligosaccharides, in the form of glycogen or in the form of starch. Deschamps *et al.* (2008*c*) argued that this flexibility of the pathway enabled the evolving green algae to redirect storage polysaccharide metabolism in three steps of increasing biochemical complexity corresponding to the three different types of end-products. Interestingly these three forms of storage glucans correspond to three pools of carbohydrates of progressively increasing size. Linear MOS consist of α-1,4 glucan chains up to a degree of polymerization of 20. Chains of longer size will spontaneously align and crystallize making them inaccessible and potentially cytotoxic. Nevertheless, these carbohydrate stores will increase the amount of glucose readily accessible to cellular metabolism. Small MOS on the other hand cannot be accumulated to very high levels as they are quite active osmotically. Glycogen is comparatively much less active osmotically and readily hydrosoluble. Glycogen will thus define a stage where larger pools of carbohydrates can be stored in an osmotically tolerable form. Finally, starch, being semi-crystalline and solid, will define a stage where huge amounts of carbohydrates can be stored in an osmotically inert form; although mobilization of these stores will be more complex than those of glycogen.

Whatever reason(s) prompted the return of storage glucan(s) to plastids, this (these) reason(s) seem to have acted selectively in the emerging Chloroplastida lineage (Deschamps *et al.*, 2008*b*). Indeed, three events seemed to correlate with the emergence of the Chloroplastida: the evolution of novel chlorophyll *b*-containing light-harvesting complexes, the return of starch metabolism in plastids, and an increased complexity of starch metabolism. A bioinformatic analysis carried out in six green alga genomes demonstrated that, in the earliest diverging prasinophyte algae (the mamelliales), the full complexity of starch metabolism witnessed in higher plants is already present (Deschamps *et al.*, 2008*b*). Deschamps *et al.* (2008*c*) proposed that the evolution of novel LHCs in green algae

might have propelled the return of storage glucans to plastids. Indeed, higher plants defective for the import of ATP at night were demonstrated to experience photooxidative stresses that depended on the size of the plastidial starch pools (Reinhold *et al.*, 2007). This was due to the accumulation of protoporphyrin IX, an intermediate of chlorophyll biosynthesis, due to the high requirement of magnesium chelatase for ATP (Reinhold *et al.*, 2007). The flux to chlorophyll is suspected to have increased in Chloroplastida which have substituted the phycobilisomes by chlorophyll-containing light-harvesting complexes. Hence the need for a plastidial source of ATP at night would have increased (Deschamps *et al.*, 2008*c*).

Whether or not resistance to photooxidative stresses prompted the return of storage glucans to plastids, this return seems in any case to have required an increase in the pathway's complexity. Deschamps *et al.* (2008*b, c*) argue that the return to the chloroplast and starch metabolism complexity are mechanistically linked. The problems inherent to a rewiring of a whole suite of enzymes have been outlined above. However, such a return can be imagined if the whole process happened in sequential steps. These entail two fully functional intermediate stages involving first MOS production then glycogen synthesis (Deschamps *et al.*, 2008*c*). Until starch biosynthesis was finally achieved the plastidial carbohydrate pools remained significantly smaller than those present in the form of cytosolic starch. However, the appearance of starch in the plastid correlated with the loss of this polymer in the cytosol as two major pools of storage carbohydrates in distinct cellular compartments have never been reported. The key to understand the extent of subfunctionalizations witnessed in Chloroplastida is to imagine that the two intermediate stages, the accumulation of malto-oligosaccharides or glycogen lasted long enough to subject the rewired enzymes to mutations that would have optimized their activities with respect to the products they now synthesized. For instance, if a starch synthase gene from the cytosolic starch metabolism network was duplicated and acquired a transit peptide for its localization within plastids at the time of MOS synthesis, then this gene would then accumulate mutations optimizing its function for that purpose. This would entail, in this case, the ability to prime the reaction at increased rates and to elongate short oligosaccharides that remain hydrosoluble. When a selection pressure reappeared to increase the plastidial storage carbohydrate pools further by evolving to the synthesis of glycogen then it is easier to imagine that an enzyme from the cytosolic starch metabolism network was duplicated and its product redirected to plastids. It would probably have taken longer to accumulate mutations in the aforementioned MOS synthase in order to turn the enzyme into a polysaccharide synthase rather than just duplicate a cytosolic enzyme gene whose product was already able to cope with branched substrates. When the novel polysaccharide synthase landed within the plastids it will not have led to the loss of the MOS synthase. Indeed, the latter would have been more efficient in glucan priming and the synthesis of short oligosaccharides leading to subfunctionalization of the two enzymes into their respective specializations. Such a reasoning applies to all enzymes of starch metabolism and the reader is referred to a recent review for further details (Deschamps *et al.*, 2008*c*).

The general consequence of this kind of networking is that the amount of gene duplications and subfunctionalizations experienced will depend, on the one hand, on the initial complexity of the cytosolic starch metabolism pathway and, on the other hand, on the number of times a certain type of activity is required to move from one stage to the next. For instance, elongation by glucan synthases is required at all three stages while branching is only required for glycogen and starch synthesis. The expected numbers of enzymes corresponding to starch synthase or branching enzyme isoforms will be, respectively, of six and two, since the initial cytosolic network contained, respectively, two and one isoforms of each. Similarly, debranching will only be required for two stages while disproportionating enzyme or GBSSI will only be required, respectively, at the MOS or starch stage yielding two or one isoforms. As mentioned previously, the appearance of starch in plastids correlated with its loss in the cytosol. Most of the genes of the cytosolic pathway were lost except those that were recruited to establish the cytosolic pathway of maltose degradation (cytosolic phosphorylase and dpe2) (reviewed by Fettke *et al.*, 2009). The establishment of starch in the plastid therefore required the evolution of a novel transporter exporting maltose, the major product of starch catabolism to the cytosol. This mex (maltose export) protein is apparently the only innovation of the Chloroplastidae with respect to starch metabolism as it is not found elsewhere (Niittylä *et al.*, 2004).

Figure 6 summarizes the rewiring process while Table 2 displays the predictions of isoform numbers with this evolution scenario. One must not forget that in addition to the process of duplication and enzyme subfunctionalization pertaining to the rewiring mechanism that we propose, other reasons may have prompted a minor proportion of enzyme subfunctionalizations. For instance, isoamylase was evidently duplicated and subfunctionalized in the cytosolic starch metabolism network independently of the rewiring mechanism outlined above. Similarly ADP-glucose pyrophosphorylase was also duplicated and subfunctionalized into large and small subunits despite its maintenance in the cyanobiont. In these two cases, however, strong functional constraints would have favoured such processes. For instance, cytosolic subfunctionalized isoamylases were required as debranching enzymes played essential functions during both synthesis and degradation of starch polysaccharides. Occasional duplications and subfunctionalizations are indeed found in many pathways. It is argued, however, that starch metabolism in Chloroplastida differ from all other pathways by the sheer magnitude of this phenomenon. Indeed this pathway stands out as the only one displaying this type of complexity among prasinophytes. In addition, there is general good agreement between the theoretical isoform numbers deduced in Table 2 from the reconstruction mechanism that is proposed here and those

observed in the Chloroplastida starch metabolism network. Such an unexpected agreement supports our initial proposal that the ancient starch metabolism pathway was indeed cytosolic.

Amylose synthesis and the presence of the pyrenoidal starch sheath

Many green and red algae and some of the secondary endosymbiosis lines contain pyrenoids in the plastid stroma. Pyrenoids are insoluble Rubisco aggregates that are seemingly involved in the CO_2 concentration mechanism (CCM) operating in these cells (Süss *et al.*, 1995; for a review see Giordano *et al.*, 2005). In most green algae and in hornworts, in porphyridiales red algae and in cryptophytes, the pyrenoid is surrounded by a starch sheath (named pyrenoidal starch) in addition to dispersed granules found either within the plastid of the green algae (often named 'stromal starch'), in the periplastidial space of the cryptophytes or in the cytosol of the porphyridiales red algae (Izumo *et al.*, 2007).

In the case of red algae and cryptophytes, the starch remains separated from the pyrenoid by the two rhodoplast membranes or the two inner membranes from the cryptophyte secondary plastids, respectively. The shape of such granules is typically distorted and results from the physical interaction between the starch sheath and the underlying pyrenoid structure. The growth of the starch sheath was shown to be induced in *Chlamydomonas* when switching from high to low CO_2 while stromal starch synthesis is favoured in the presence of high CO_2 and (or) acetate (Villarejo *et al.*, 1996). It has been demonstrated through the use of starch defective mutants in *Chlamydomonas* and *Chlorella* that starch is not required for normal induction of the CCM (Plumed *et al.*, 1996; Villarejo *et al.*, 1996). Hence the starch sheath most probably comes as a consequence of localized CO_2 fixation by pyrenoidal Rubisco yielding concentration gradients of important metabolites such as 3-PGA and hexoses rather than being an integral part of the CCM required for pyrenoid confinement and function.

The presence of amylose in starch could be an important feature required for the synthesis of an abundant pyrenoidal starch sheath. Indeed all photosynthetic eukaryotes that synthesize a pyrenoidal starch sheath contain amylose. The porphyridiales which define the only red algae whose floridean starch is synthesized in close association with the pyrenoid also defines the only group that contains amylose within the Rhodophyceae. Delrue *et al.* (1992) noted that *Chlamydomonas* mutants defective for GBSSI and, consequently, amylose synthesis displayed an altered pyrenoidal starch sheath. Because GBSS is responsible for building long chains within both amylose and amylopectin it is possible that these chains allows the starch granules to be more 'plastic' and adopt those shapes that are compatible with a close physical association to the pyrenoid. SSIII might play a similar and partly redundant function in Chloroplastida (Maddelein *et al.*, 1994). Interestingly both SSIII and GBSSI are induced when switching to low CO_2 in *Chlamydomonas* and *Chlorella* (Miura *et al.*, 2004; Oyama *et al.*, 2006). Hence it would be tempting to suggest that amylose and thus GBSSI was maintained in starch-storing algae for the purpose of pyrenoidal starch synthesis.

It must be stressed, however, that neither subgroup V cyanobacteria nor glaucophytes nor terrestrial plants other than hornworts are reported to contain pyrenoids; yet these organisms have maintained amylose and GBSSI. Nevertheless the maintenance of some 'plasticity' in starch granule morphogenesis which is exemplified by the pyrenoidal starch sheath might define one of the major selection pressures for the maintenance of amylose in Archaeplastida. This could be important in various other instances where substrate concentration gradients could lead to localized amylose synthesis.

Chlamydial genes in the starch pathway: evidence for ancient lateral gene transfers

One of the most surprising recent findings of phylogenetic inference consist in the unexpected relationship existing between the genes of Archaeplastida and Chlamydiae (reviewed in Horn, 2008). Chlamydiae define obligate intracellular gram-negative bacterial pathogens initially documented in animals and more recently found as 'symbionts' of protists notably amoebas (Horn, 2008). Some of these protists are also thought to be able to act as reservoirs of pathogenic Chlamydiae. Chlamydiae enters cells at the 'elementary body' stage through phagocytosis and multiply only within host vacuoles in the form of reticulate bodies. Chlamydiae typically import their ATP from the host cytosol through an ATP/ADP translocase. An increasing diversity of *Chlamydia*-like microorganisms have recently been identified, often showing greater metabolic capabilities (and genomes) than the human pathogens but with similar obligate intracellular life cycles typified by the same basic energy parasitism (Horn, 2008).

Bioinformatic analysis have yielded 150 out of 2031 protein coding sequences which in the genome of *Pseudochlamydia amoebophila* display phylogenies grouping the latter with the Archaeplastida (Horn *et al.*, 2004; Horn, 2008). The more recent acquisition of the red alga *Cyanidioschizon merolae* and of diatom genomes sequence reveal between 39–55 genes in these genomes of probable Chlamydial ancestry (Huang and Gogarten, 2007; Becker *et al.*, 2008; Moustafa *et al.*, 2008). The data are suggestive of a very ancient origin of this particular relationship probably at the time of endosymbiosis.

However, it is very hard to decide if the Chlamydiae donated genes to the Archaeplastida or if they took up such genes during this ancient relationship. One obvious case where the Chlamydia donated a function which can be considered as important for the establishment of endosymbiosis has been documented when analysing the phylogeny of the NTT transporters (Linka *et al.*, 2003). These transporters have been shown to import ATP at night within

plastids (Reinhold *et al.*, 2007). Such a function derived from the ATP/ADP translocase at the heart of the parasitic way of life of the Chlamydiae (Trentmann *et al.*, 2007) would indeed have been of importance for a cyanobiont that, we argue, was deprived of storage polysaccharides. The starch pathway contains two candidates for horizontal transfer of Chlamydial genes: the isoamylases (Huang and Gogarten, 2007; Becker *et al.*, 2008; Moustafa *et al.*, 2008) and one particular soluble starch synthase (the ancestor of SSIII and SSIV) (Moustafa *et al.*, 2008). Isoamylase might very well define a case of LGT (lateral gene transfer) similar to the ATP/ADP translocase. The transfer of the isoamylase gene which switched glycogen to starch synthesis was of importance since it increased the pool of polysaccharides resulting from the export of photosynthate to the host cytosol. As to the soluble starch synthase, phylogenies support an LGT of an enzyme related to extant chloroplastidal SSIII-SSIV from Protochlamydia to plants (Moustafa *et al.*, 2008). However, in this case, only the ancestral Chlamydiae display convincing evidence for this transfer and the more classic animal pathogen-type of Chlamydiae proved to contain an unrelated enzyme.

One of the most parsimonious explanations for all these surprising observations would be that the protist lineage, which would enter into a symbiosis with a particular cyanobacterium, was routinely infected by an ancestor of extant *Chlamydia* (Huang and Gogarten, 2007). During integration of the cyanobiont the recurrent presence of chlamydial parasites expanded the repertoire of useful LGTs, thereby facilitating establishment of endosymbiosis. For instance, a chlamydial parasite donated its ATP/ADP translocase enabling the cyanobiont to withstand in darkness the negative consequences of the loss of its storage polysaccharides. In addition, *Chlamydia* parasites may have been the source of the Archaeplastida isoamylases turning the pool of cytosolic glycogen into starch, thereby further increasing photosynthate export.

Present-day plants and algae are not known to be subjected to infection by *Chlamydia*. Therefore, this particular influence of Chlamydiae-related parasites was terminated at a very early stage of Archaeplastida evolution, possibly before the divergence of Chloroplastida, Rhodophyceae, and Glaucophyta. According to Huang and Gogarten (2007), this happened when the Archaeplastida's ancestor switched from a mixotrophic way of life involving both phagocytosis and taping carbon and energy from its cyanobiont to an autotrophic way of life. This could have resulted very simply by the loss of phagocytosis. This, in turn, would have prevented penetration of the energy parasites which are known to use phagocytosis to enter their hosts. The building of rigid and complex cell walls would further strengthen this immunity to infection by *Chlamydia*.

The finding of isoamylase as a bona fide chlamydial gene has important consequences on our understanding of the evolution of starch metabolism in Archaeplastida. Indeed the finding of true starch and of GBSSI in unicellular diazotrophic cyanobacteria would tend to suggest that the cyanobiont might have donated the critical genes required for turning glycogen into starch. While this remains valid when one envisions the synthesis of amylose by GBSSI, the switch of hydrosoluble glycogen to starch in the archaeplastidal cytosol did not result from an LGT from a cyanobacterial debranching enzyme gene but rather by that of the corresponding chlamydial gene. This, in turn, could mean that semi-crystalline polymers may be synthesized by a different pathway in cyanobacteria. In any case this argues that semi-crystalline storage polysaccharides evolved independently in cyanobacteria and Archaeplastida. A critical evaluation of the phylogeny of debranching enzymes in apicomplexa parasites further suggests that the archaeplastidal isoamylase was not the source for the debranching enzyme gene in starch accumulating secondary endosymbiosis lineages (Coppin *et al.*, 2005; Deschamps *et al.*, 2008*c*). There again the switch from glycogen to starch may have occurred independently in the cytosol or periplast of these organisms.

General conclusion: starch metabolism evolution gives an unexpected window on those fundamental events that shaped the Archaeplastida

When we first pondered the evolution of the starch pathway in Archaeplastida we were faced with the problem of fitting a dual substrate storage polysaccharide metabolism relying on two types of glycosyl-nucleotides (ADP-glucose and UDP-glucose) in a single compartment defined by the common ancestor's cytosol. This could only be achieved and make any kind of physiological sense if ADP-glucose was exported from the cyanobiont. We thus stumbled on what actually could define the major metabolic link of plastid endosymbiosis as the only plausible solution to our compartmentalization problem. In that sense, examining the origin of starch metabolism in Archaeplastida gave us an unexpected window on the critical early steps that led to the establishment of the cyanobiont. The proposal that we make of an ancient cytosolic starch metabolism network also gives us the reasons underlying the otherwise unexplainable complexity of starch metabolism that characterize selectively the Chloroplastida compared with the Rhodophyceae and Glaucophyta. The history of the starch metabolism network in vascular plants is thus highly complex and fascinating. When going through this complex history it appears that synthesis of semi-crystalline starch-like polysaccharides evolved several times independently from glycogen metabolism: it appeared in unicellular diazotrophic cyanobacteria, it evolved in the cytosol of the Archaeplastida through the use of enzymes which are of chlamydial rather than cyanobacterial origin and it appeared again in the cytoplasm of the common ancestor of dinoflagellates and apicomplexa. Hence the fact that the cyanobiont ancestor was a starch accumulator may not have been a requirement for the evolution of semi-crystalline

polysaccharides per se in Archaeplastida. Nevertheless, there is one important component of starch which required the presence of such ancestors. Amylose is, in all cases, known to depend on the presence of GBSSI, an enzyme which, apart from Archaeplastida and a few secondary endosymbiosis derivatives, is only found to date in such unicellular diazotrophic starch-accumulating cyanobacteria. Because GBSSI has been proven to be significantly active only when associated with semi-crystalline amylopectin packaged into solid starch granules, the cyanobiont ancestors will have had to have been starch accumulators.

Another consequence from the complex history of starch metabolism is that the function of an enzyme seems to be just as much dependent on the historical record of the network to which this enzyme belongs as on its catalytic activity per se. A deep understanding of starch metabolism will require an understanding of the function of an enzyme through the entire evolutionary process: that is in cyanobacteria or chlamydia for those enzymes that originated from these organisms, in the cytosol of the heterotrophic eukaryotes for those of eukaryotic ancestry, after the merging of the pathways in the first archaeplastidal cell, and within each of the major archaeplastidal lines as well as their secondary endosymbiosis derivatives.

Acknowledgements

This work was performed in the Unité de Glycobiologie Structurale et Fonctionnelle funded by the Region Nord Pas de Calais, the European Union, the French Ministry of Education, the CNRS and ANR grant 'starchevol'. In memory of Jean Montreuil recently deceased who fathered the field of Glycobiology in France.

References

Adl SM, Simpson AG, Farmer MA, *et al*. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* **52,** 399–451.

Alonso-Casajús N, Dauvillée D, Viale AM, Muñoz FJ, Baroja-Fernández E, Morán-Zorzano MT, Eydallin G, Ball S, Pozueta-Romero J. 2006. The role of glycogen phosphorylase in glycogen breakdown and structure in *Escherichia coli*. *Journal of Bacteriology* **188,** 5266–5272.

Aury JM, Jaillon O, Duret L, *et al*. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444,** 171–178.

Baldauf SL. 2003. The deep roots of eukaryotes. *Science* **300,** 1701–1703.

Ball SG, Guan HP, James M, Myers A, Keeling P, Mouille G, Buléon A, Colonna P, Preiss J. 1996. From glycogen to amylopectin: a model explaining the biogenesis of the plant starch granule. *Cell* **86,** 349–352.

Ball S, van de Wal M, Visser R. 1998. Progress in understanding the biosynthesis of amylose. *Trends in Plant Science* **3,** 462–467.

Ball SG, Morell MK. 2003. Starch biosynthesis. *Annual Review of Plant Biology* **54,** 207–233.

Ballicora MA, Iglesias AA, Preiss J. 2003. ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogensynthesis. *Microbiology and Molecular Biology Reviews* **67,** 213–225.

Becker B, Hoef-Emden K, Melkonian M. 2008. Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evolutionary Biology* **8,** 203.

Blennow A, Nielsen TH, Baunsgaard L, Mikkelsen R, Engelsen SB. 2002. Starch phosphorylation: a new front line in starch research. *Trends in Plant Science* **7,** 445–450.

Bodyl A, Makiewicz P, Stiller JW. 2007. The intracellular cyanobacterial of *Paulinella chromatophora*: endosymbionts or organelles? *Trends in Microbiology* **15,** 295–296.

Bodyl A, Mackiewicz P, Stiller JW. 2009. Early steps in plastid evolution: current ideas and controversies. *Bioessays* **31,** 1219–1232.

Bolter B, Soll J, Schulz A, Hinnah S, Wagner R. 1998. Origin of a chloroplast protein importer. *Proceedings of the National Academy of Sciences, USA* **95,** 15831–15836.

Boos W, Shuman H. 1998. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiology and Molecular Biology Reviews* **62,** 204–229.

Buléon A, Colonna P, Planchot V, Ball S. 1998. Starch granules: structure and biosynthesis. *International Journal of Biological Macromolecules* **23,** 85–112.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active Enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* **37,** D233–D238.

Carlton JM, Hirt RP, Silva JC, *et al*. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis.*. *Science* **315,** 207–212.

Cavalier-Smith T. 2006. Cell evolution and Earth history: stasis and revolution. *Philosophical Transactions of the Royal Society B, Biological Sciences* **361,** 969–1006.

Colleoni C, Dauvillée D, Mouille G, Morell M, Samuel M, Slomiany MC, Liénard L, Wattebled F, D'Hulst C, Ball S. 1999. Biochemical characterization of the *Chlamydomonas reinhardtii* alpha-1,4 glucanotransferase supports a direct function in amylopectin biosynthesis. *Plant Physiology* **120,** 1005–1014.

Colleoni C, Linka M, Deschamps P, Handford MG, Dupree P, Weber APM, Ball SG. 2010. Identification of a missing link required for the endosymbiotic origin of eukaryotic photosynthesis. *Molecular Biology and Evolution* **27,** 2691–2701.

Comparat-Moss S, Denyer K. 2009. The evolution of the starch biosynthetic pathway in cereals and other grasses. *Journal of Experimental Botany* **60,** 2481–2492.

Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobillard MO, Buleon A, Ball S, Tomavo S. 2005. Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *Journal of Molecular Evolution* **60,** 257–267.

Dauvillée D, Colleoni C, Shaw E, *et al*. 1999. Novel starch-like polysaccharides are synthesized by a soluble form of granule-bound starch synthase in glycogen accumulating mutants of Chlamydomonas reinhardtii. *Plant Physiology* **119,** 321–330.

Dauvillée D, Kinderf IS, Li Z, Kosar-Hashemi B, Samuel MS, Rampling L, Ball S, Matthew MK. 2005. Role of the *E. coli glgX* gene in glycogen metabolism. *Journal of Bacteriology* **187,** 1465–1473.

Dauvillée D, Chochois V, Steup M, *et al*. 2006. Plastidial phosphorylase is required for normal starch granule biogenesis in the monocellular alga *Chlamydomonas reinhardtii*. *The Plant Journal* **48,** 274–285.

Dauvillée D, Deschamps P, Ral JP, Plancke C, Putaux JL, Devassine J, Durand-Terrasson A, Devin A, Ball SG. 2009. Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate *Crypthecodinium cohnii*. *Proceedings of the National Academy of Sciences, USA* **106,** 21126–21130.

De Koning AP, Keeling PJ. 2006. The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biology* **4,** 12.

Delrue B, Fontaine T, Routier F, Decq A, Wieruszeski JM, van den Koornhuyse N, Maddelein ML, Fournet B, Ball S. 1992. Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin. *Journal of Bacteriology* **174,** 3612–3620.

Derelle E, Ferraz C, Rombault S, *et al*. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences, USA* **103,** 11647–11652.

Deschamps P, Colleoni C, Nakamura Y, *et al*. 2008*a*. Metabolic symbiosis and the birth of the plant kingdom. *Molecular Biology and Evolution* **25,** 536–548.

Deschamps P, Guillebeault D, Devassine J, *et al*. 2008*d*. The heterotrophic dinoflagellate *Crypthecodinium cohnii* defines a model genetic system to investigate cytoplasmic starch synthesis. *Eukaryotic Cell* **7,** 247–257.

Deschamps P, Haferkamp I, Dauvillee D, *et al*. 2006. Nature of the periplastidial pathway of starch synthesis in the cryptophyte *Guillardia theta*. *Eukaryotic Cell* **5,** 954–963.

Deschamps P, Haferkamp I, d'Hulst C, Neuhaus E, Ball S. 2008*c*. The relocation of starch metabolism to chloroplasts: when, why and how. *Trends in Plant Science* **13,** 1802–1816.

Deschamps P, Moreau H, Worden AZ, Dauvillée D, Ball SG. 2008*b*. Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. *Genetics* **178,** 2373–2387.

Edner C, Li J, Albrecht T, *et al*. 2007. Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial beta-amylases. *Plant Physiology* **145,** 17–28.

Eichinger L, Pachebat JA, Glöckner G, *et al*. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435,** 43–57.

Eisen JA, Coyne RS, Wu M, *et al*. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biology* **4,** 1620–1642.

Falcon L, Lindwall I, Bauer K, Bergman B, Carpenter EJ. 2004. Ultrastructure of unicellular N-2 fixing cyanobacteria from the tropical North Atlantic and subtropical North Pacific Oceans. *Journal of Phycology* **40,** 1074–1078.

Fettke J, Hejazi M, Smirnova J, Höchel E, Stage M, Steup M. 2009. Eukaryotic starch degradation: integration of plastidial and cytosolic pathways. *Journal of Experimental Botany* **60,** 2907–2922.

François J, Parrou JL. 2001. Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* **25,** 125–145.

Gentry MS, Dowen RH, Worby CA, Mattoo S, Ecker JR, Dixon JE. 2007. The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. *The Journal of Cell Biology* **178,** 477–488.

Gil R, Latorre A, Moya A. 2004. Bacterial endosymbionts of insects: insights from comparative genomics. *Environmental Microbiology* **6,** 1109–1122.

Giordano M, Beardall J, Raven J. 2005. CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annual Reviews of Plant Biology* **56,** 99–131.

Hamana K, Aizaki T, Arai E, Saito A, Uchikata K, Ohnishi H. 2004. Distribution of norspermidine as a cellular polyamine within microgreen algae including the non-photosynthetic achlorophyllous *Polytoma, Polytomella, Prototheca,* and *Helicosporidium*. *Journal of General and Applied Microbiology* **50,** 289–295.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic 'supergroups'. *Proceedings of the National Academy of Sciences, USA* **106,** 3859–3864.

Handford M, Rodriguez-Furlán C, Orellana A. 2006. Nucleotide-sugar transporters: structure, function and roles *in vivo*. *Brazilian Journal of Medical and Biological Research* **39,** 1149–1158.

Hennen-Bierwagen TA, Liu F, Marsh RS, Kim S, Gan Q, Tetlow IJ, Emes MJ, James MG, Myers AM. 2008. Starch biosynthetic enzymes from developing *Zea mays* endosperm associate in multisubunit complexes. *Plant Physiology* **146,** 1892–1908.

Henrissat B, Deleury E, Coutinho PM. 2002. Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? *Trends in Genetics* **18,** 437–440.

Honda D, Yokota A, Sugiyama J. 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *Journal of Molecular Evolution* **48,** 723–739.

Horn M. 2008. Chlamydiae as symbionts in eukaryotes. *Annual Review of Microbiology* **62,** 113–131.

Horn M, Collingro A, Schmitz-Esser S, *et al*. 2004. Illuminating the evolutionary history of Chlamydiae. *Science* **304,** 728–730.

Huang J, Gogarten JP. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biology* **8,** R99.

Izumo A, Fujiwara S, Oyama Y, Satoh A, Fujita N, Nakamura Y, Tsuzuki M. 2007. Physicochemical properties of starch in *Chlorella*

change depending on the CO_2 concentration during growth: comparison of structure and properties of pyrenoid and stroma starch. *Plant Science* **172,** 1138–1147.

James MG, Robertson DS, Myers AM. 1995. Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. *The Plant Cell* **7,** 417–429.

Keeling PJ. 2009. Chomalveolates and the evolution of plastids by secondary endosymbiosis. *Journal of Eukaryotic Microbiology* **56,** 1–8.

Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biology* **11,** 209–221.

Kotting O, Santelia D, Edner C, *et al*. 2009. STARCH-EXCESS4 is a laforin-like phosphoglucan phosphatase required for starch degradation in *Arabidopsis thaliana*. The *Plant Cell* **21,** 334–346.

Krebs EG. 1983. Historical perspectives on protein phosphorylation and a classification system for protein kinases. *Philosophical Transactions of the Royal Society B: Biological Sciences* **302,** 3–11.

Lin T-P, Caspar T, Somerville C, Preiss J. 1988. Isolation and characterisation of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh. lacking ADP-glucose pyrophosphorylase activity. *Plant Physiology* **86,** 1131–1135.

Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, Urbany C, Seil I, Kusch J, Neuhaus HE. 2003. Phylogenetic relationship of nonmitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* **306,** 27–35.

Loddenkötter B, Kammerer B, Fischer K, Flügge UI. 1993. Expression of the functional mature chloroplast triose phosphate translocator in yeast internal membranes and purification of the histidine-tagged protein by a single metal-affinity chromatography step. *Proceedings of the National Academy of Sciences, USA* **90,** 2155–2159.

Loftus B, Anderson I, Davies R, *et al*. 2005. The genome of the protist parasite. *Entamoeba histolytica*. *Nature* **433,** 865–868.

Maddelein ML, Libessart N, Bellanger F, Delrue B, D'Hulst C, van den Koornhuyse N, Fontaine T, Wieruszeski JM, Decq A, Ball S. 1994. Toward an understanding of the biogenesis of the starch granule. Determination of granule-bound and soluble starch synthase functions in amylopectin synthesis. *Journal of Biological Chemistry* **269,** 25150–25157.

Martin W. 2010. Evolutinary origins of metabolic compartmentalization in eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365,** 847–855.

Martinez-Duncker I, Mollicone R, Codogno P, Oriol R. 2003. The nucleotide-sugar transporter family: a phylogenetic approach. *Biochimie* **85,** 245–260.

Matsuzaki M, Misumi O, Shin-i T, *et al*. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428,** 653–657.

McFadden GI, van Dooren GG. 2004. Evolution: red algal genome affirms a common origin of all plastids. *Current Biology* **14,** 514–516.

Miao X, Wu Q, Wu G, Zhao N. 2003. Changes in photosynthesis and pigmentation in an agp deletion mutant of the cyanobacterium *Synechocystis* sp. *Biotechnology Letters* **25,** 391–396.

Miao X, Wu Q, Wu G, Zhao N. 2006. Sucrose accumulation in salt-stressed cells of agp gene deletion-mutant in cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiology Letters* **218,** 71–77.

Michels PAM, Opperdoes FR. 1991. The evolutionary origin of glycosomes. *Parasitology Today* **7,** 105–109.

Miura K, Yamano T, Yoshioka S, *et al*. 2004. Expression profiling-based identification of CO_2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant Physiology* **135,** 1595–1607.

Morell MK, Myers AM. 2005. Towards the rational design of cereal starches. *Current Opinion in Plant Biology* **8,** 204–210.

Mouille G, Maddelein ML, Libessart N, Talaga P, Decq A, Delrue B, Ball S. 1996. Preamylopectin processing: a mandatory step for starch biosynthesis in plants. *The Plant Cell* **8,** 1353–1366.

Moustafa A, Reyes-Prieto A, Bhattacharya D. 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* **3,** e2205.

Myers AM, Morell MK, James MG, Ball SG. 2000. Recent progress toward understanding biosynthesis of the amylopectin crystal. *Plant Physiology* **122,** 989–998.

Nakamura Y, Kubo A, Shimamune T, Matsuda T, Harada K, Satoh H. 1997. Correlation between activities of starch debranching enzymes and α-polyglucan structure in endosperms of sugary-1 mutants of rice. *The Plant Journal* **12,** 143–153.

Nakamura Y, Takahashi J, Sakurai A, *et al*. 2005. Some cyanobacteria synthesize semi-amylopectin type α-polyglucans instead of glycogen. *Plant and Cell Physiology* **46,** 539–545.

Niittylä T, Messerli G, Trevisan M, Chen J, Smith AM, Zeeman SC. 2004. A previously unknown maltose transporter essential for starch degradation in leaves. *Science* **303,** 87–89.

Nowack ECM, Melkonian M, Glöckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Current Biology* **18,** 410–418.

Oyama Y, Izumo A, Fujiwara S, Shimonaga T, Nakamura Y, Tsuzuki M. 2006. Granule-bound starch synthase cDNA in *Chlorella kessleri* 11h: cloning and regulation of expression by CO_2 concentration. *Planta* **224,** 646–654.

Patron NJ, Keeling PK. 2005. Common evolutionary origin of starch biosynthetic enzymes in green and red algae. *Journal of Phycology* **41,** 1131–1141.

Plancke C, Colleoni C, Deschamps P, *et al*. 2008. The pathway of starch synthesis in the model glaucophyte *Cyanophora paradoxa*. *Eukaryotic Cell* **7,** 247–257.

Plumed MP, Villarejo A, de los Rios A, Garcia-Reina G, Ramazanov Z. 1996. The CO_2-concentrating mechanism in a starchless mutant of the green unicellular alga *Chlorella pyrenoidosa*. *Planta* **200,** 28–31.

Pombert J-F, Keeling PJ. 2010. The mitochondrial genome of the entomoparasitic green alga *Helicosporidium*. *PLoS ONE* **5,** e8954.

Preiss J. 1984. Bacterial glycogen synthesis and its regulation. *Annual Review of Microbiology* **38,** 419–458.

Ral JP, Derelle E, Ferraz C, *et al*. 2004. Starch division and partitioning a mechanism for granule propagation and maintenance in the picophytoplanktonic green alga *Ostreococcus tauri*. *Plant Physiology* **136,** 3333–3340.

Raven JA, Beardall J, Flynn KJ, Maberly SC. 2009. Phagotrophy in the origins of photosynthesis in eukaryotes and as a complementary mode of nutrition in phototrophs: relation to Darwin's insectivorous plants. *Journal of Experimental Botany* **60,** 3975–3987.

Reinhold T, Alawady A, Grimm B, *et al*. 2007. Limitation of nocturnal import of ATP into Arabidopsis chloroplasts leads to photooxidative damage. *The Plant Journal* **50,** 293–304.

Reyes-Sosa FM, Molina-Heredia FP, De la Rosa MA. 2010. A novel α-amylase from the cyanobacterium *Nostoc* sp. PCC7119. *Applied Microbiology and Biotechnology* **86,** 131–141.

Richards T, Cavalier-Smith T. 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436,** 1113–1118.

Ritte G, Lloyd JR, Eckermann N, Rottmann A, Kossmann J, Steup M. 2002. The starch-related R1 protein is an α-glucan, water dikinase. *Proceedings of the National Academy of Sciences, USA* **99,** 7166–7171.

Roach PJ. 2002. Glycogen and its metabolism. *Current Molecular Medicine* **2,** 101–120.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology* **15,** 1325–1330.

Rollwitz I, Santaella M, Hille D, Flügge U-I, Fischer K. 2006. Characterization of AtNST-KT1, a novel UDP-galactose transporter from *Arabidopsis thaliana*. *FEBS Letters* **580,** 4246–4251.

Shadan S. 2007. Evolutionary genetics: you are what you ate. *Nature* **449,** 155.

Scheidig A, Frèohlich A, Schulze S, Lloyd JR, Kossmann J. 2002. Downregulation of a chloroplast-targeted beta-amylase leads to a starch-excess phenotype in leaves. *The Plant Journal* **30,** 581–591.

Schneegurt MA, Sherman DM, Nayar S, Sherman LA. 1994. Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium *Cyanothece* sp. strain ATCC 51142. *Journal of Bacteriology* **176,** 1586–1597.

Schneegurt MA, Sherman DM, Sherman LA. 1997. Composition of carbohydrate granules of the cyanobacterium, *Cyanothece* sp. strain ATCC51142. *Archives of Microbiology* **167,** 89–98.

Shearer J, Graham TE. 2002. New perspectives on the storage and organization of muscle glycogen. *Journal of Applied Physiology* **27,** 179–203.

Shimonaga T, Fujiwara S, Kaneko M, Izumo A, Nihei S, Francisco BP, Satoh A, Fujita N, Nakamura Y, Tsuzuki M. 2007. Variation in storage alpha-polyglucans of red algae: amylose and semi-amylopectin types in porphyridium and glycogen type in cyanidium. *Marine Biotechnology* **9,** 192–202.

Shimonaga T, Konishi M, Oyama Y, *et al*. 2008. Variations in storage α-polyglucans of the Porphyridiales (Bangiophycidae, Rhodophyta). *Plant and Cell Physiology* **49,** 103–116.

Song J, Xu Q, Olsen R, Loomis WF, Shaulsky G, Kuspa A, Sucgang R. 2005. Comparing the *Dictyostelium* and *Entamoeba* genomes reveals an ancient split in the Conosa lineage. *PLoS Computational Biology* **1,** 579–584.

Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences, USA* **100,** 8828–8833.

Süss K-H, Prokhorenko I, Adler K. 1995. *In situ* association of Calvin cycle enzymes, ribulose-1,5-bisphosphate carboxylase/oxygenase activase, ferredoxin-$NADP^+$ reductase, and nitrite reductase with thylakoid and pyrenoid membranes of *Chlamydomonas reinhardtii* chloroplasts as revealed by immunoelectron microscopy. *Plant Physiology* **107,** 1387–1397.

Tagliabracci VS, Girard JM, Segvich D, Meyer C, Turnbull J, Zhao X, Minassian BA, Depaoli-Roach AA, Roach PJ. 2008. Abnormal metabolism of glycogen phosphate as a cause for lafora disease. *Journal of Biological Chemistry* **283,** 33816–33825.

Teste MA, Enjalbert B, Parrou JL, Francois JM. 2000. The *Saccharomyces cerevisiae YPR184w* gene encodes the glycogen debranching enzyme. *FEMS Microbiology Letters* **193,** 105–110.

Tetlow IJ, Morell MK, Emes MJ. 2004*a*. Recent developments in understanding the regulation of starch metabolism in higher plants. *Journal of Experimental Botany* **55,** 2131–2145.

Tetlow IJ, Wait R, Lu ZX, Akkasaeng R, Bowsher CG, Esposito S, Kosar-Hashemi B, Morell MK, Emes MJ. 2004*b*. Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein–protein interactions. *The Plant Cell* **16,** 694–708.

Thorsness PE, Fox TD. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* **346,** 376–379.

Tomlinson K, Denyer K. 2003. Starch synthesis in cereal grains. *Advances in Botanical Research* **40,** 1–61.

Trentmann O, Horn M, van Scheltinga ACT, Neuhaus HE, Haferkamp I. 2007. Enlightening energy parasitism by analysis of an ATP/ADP transporter from chlamydiae. *PLos Biology* **5,** e231.

Ugalde JE, Parodi AJ, Ugalde RA. 2003. *De novo* synthesis of bacterial glycogen: *Agrobacterium tumefaciens* glycogen synthase is involved in glucan initiation and elongation *Proceedings of the National Academy of Sciences*, *USA* **100,** 10659–10663.

Villarejo A, Buren S, Larsson S, *et al*. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nature Cell Biology* **7,** 1224–1231.

Villarejo A, Martínez F, Plumed MP, Ramazanov Z. 1996. The induction of the CO_2 concentrating mechanism in a starch-less mutant of *Chlamydomonas reinhardtii*. *Physiologia Plantarum* **98,** 798–802.

Viola R, Nyvall P, Pedersen M. 2001. The unique features of starch metabolism in red algae. *Proceedings of the Royal Society B: Biological Sciences* **268,** 1417–1422.

Wang Z, Wilson WA, Fujino MA, Roach PJ. 2001. Antagonistic controls of autophagy and glycogen accumulation by Snf1p, the yeast homolog of AMP-aativated protein kinase, and the cyclin-dependent kinase Pho85p. *Molecular and Cellular Biology* **21,** 5742–5752.

Wattebled F, Dong Y, Dumez S, *et al*. 2005. Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. *Plant Physiology* **138,** 184–195.

Weber APM, Linka M, Bhattacharya D. 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryotic Cell* **5,** 609–612.

Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J. 2010. Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiology Reviews* **34,** 952–985.

Wing-Ming C, Hsueh-Mei C, Hso-Freng Y, Jei-Fu S, Tan-Chi H. 1994. The aerobic nitrogen-fixing *Synechococcus* RF-1 containing uncommon polyglucan granules and multiple forms of α-amylase. *Current Microbiology* **29,** 201–205.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology and Evolution* **21,** 809–818.

Zabawinski C, Van Den Koornhuyse N, D'Hulst C, Schlichting R, Giersch C, Delrue B, Lacroix JM, Preiss J, Ball S. 2001. Starchless mutants of *Chlamydomonas reinhardtii* lack the small subunit of an heterotetrameric ADP-glucose pyrophosphorylase. *Journal of Bacteriology* **183,** 1069–1077.

Zeeman SC, Umemoto T, Lue W-L, Au-Yeung P, Martin C, Smith AM, Chen J. 1998. A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. *The Plant Cell* **10,** 1699–1712.

Zeeman SC, Smith SM, Smith AM. 2007. The diurnal metabolism of leaf starch. *Biochemical Journal* **401,** 13–28.

The Plant Cell, Vol. 25: 7–21, January 2013, www.plantcell.org

PERSPECTIVE

Metabolic Effectors Secreted by Bacterial Pathogens: Essential Facilitators of Plastid Endosymbiosis?[W][OA]

Steven G. Ball,[a,2] Agathe Subtil,[b,1] Debashish Bhattacharya,[c,1] Ahmed Moustafa,[d] Andreas P.M. Weber,[e] Lena Gehre,[b] Christophe Colleoni,[a] Maria-Cecilia Arias,[a] Ugo Cenci,[a] and David Dauvillée[a]

[a] Université des Sciences et Technologies de Lille, Unité de Glycobiologie Structurale et Fonctionnelle, Unité Mixte de Recherche 8576, Centre National de la Recherche Scientifique–Université des Sciences et Technologies de Lille, Cité Scientifique, 59655 Villeneuve d'Ascq cedex, France

[b] Institut Pasteur, Unité de Biologie des Interactions Cellulaires, Unité de Recherche Associée au Centre National de la Recherche Scientifique 2582, 75015 Paris, France

[c] Department of Ecology, Evolution, and Natural Resources and Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901

[d] Department of Biology and Biotechnology Graduate Program, American University in Cairo, New Cairo 11835, Egypt

[e] Institute for Plant Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, Germany

Under the endosymbiont hypothesis, over a billion years ago a heterotrophic eukaryote entered into a symbiotic relationship with a cyanobacterium (the cyanobiont). This partnership culminated in the plastid that has spread to forms as diverse as plants and diatoms. However, why primary plastid acquisition has not been repeated multiple times remains unclear. Here, we report a possible answer to this question by showing that primary plastid endosymbiosis was likely to have been primed by the secretion in the host cytosol of effector proteins from intracellular Chlamydiales pathogens. We provide evidence suggesting that the cyanobiont might have rescued its afflicted host by feeding photosynthetic carbon into a chlamydia-controlled assimilation pathway.

INTRODUCTION

Members of the photoautotrophic lineage, the Archaeplastida (also called Kingdom Plantae) consisting of the Chloroplastida (green algae and plants), Rhodophyceae (red algae), and Glaucophyta (glaucophytes) are the founding lineage of photosynthetic eukaryotes (Rodríguez-Ezpeleta et al., 2005, Reyes-Prieto et al., 2007; Chan et al., 2011). Despite the advantages offered by photosynthesis and the innumerable opportunities that presumably have existed for phagotrophic protists to recapitulate plastid acquisition, only one other case is known of primary plastid endosymbiosis, in the photosynthetic amoeba *Paulinella* (Nowack et al., 2008). To understand why primary endosymbiosis is exceedingly rare, clues have been sought in the nuclear genomes of Archaeplastida. The presence of a significant cyanobacterium-derived component in these genomes is known to have arisen from endosymbiotic gene transfer, which resulted in the movement of many cyanobiont genes to the host chromosomes (Martin et al., 2002; Reyes-Prieto et al., 2006; Moustafa and Bhattacharya, 2008). Recent analyses suggest that between 6 and 10% of nuclear genes in algae retain the signature of cyanobacterial origin with most encoding proteins that return to the plastid to express their function (Sato et al., 2005; Reyes-Prieto et al., 2006; Moustafa and Bhattacharya, 2008). The second major source of foreign genes in Archaeplastida is Chlamydia-like pathogens (Stephens et al., 1998; Huang and Gogarten, 2007; Moustafa et al., 2008; Becker et al., 2008; Fournier et al., 2009; reviewed in Horn, 2008). The existence of 30 to 50 cases of horizontal gene transfers (HGTs) from members of the order Chlamydiales suggests a significant role for these obligate intracellular bacteria in Archaeplastida evolution. This may have included metabolic integration of the newly acquired plastid (Huang and Gogarten, 2007; Becker et al., 2008; Moustafa et al., 2008; Fournier et al., 2009). This intriguing possibility is buttressed by the observation that a significant portion of the Chlamydiales-derived genes are shared by two or all three archaeplastidal lineages (i.e., green and red algae and Glaucophyta) and therefore must have been present in their common ancestor (Huang and Gogarten, 2007; Becker et al., 2008; Moustafa et al., 2008; Fournier et al., 2009; reviewed in Horn, 2008).

We show that the sole extant enzyme that could have been devoted to the assimilation of photosynthate in the host cytosol is an enzyme of chlamydial origin. Because such assimilation is necessary for the selection of plastid endosymbiosis, this enzyme had to be present at the onset of endosymbiosis in the host cytosol. This specific function defines a previously unsuspected

[1] These authors contributed equally to this work.
[2] Address correspondence to steven.ball@univ-lille1.fr.
[W] Online version contains Web-only data.
[OA] Open access articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.112.101329

PERSPECTIVE

class of effectors secreted by Chlamydiales bacteria. This finding suggests strongly that an intracellular bacterium ancestor to the extant order Chlamydiales, a cyanobacterium, and a eukaryotic host were indeed linked together in a stable tripartite symbiotic relationship that led to the establishment of the plant lineages.

RESULTS

Phylogenomic Approaches: Assessing the Chlamydialean Contribution to Archaeplastidial Genomes

We reanalyzed the contribution of Chlamydiales to the genomes of Archaeplastida using the additional data (e.g., >60k proteins from the red algae *Calliarthron tuberculosum* and *Porphyridium cruentum*; Chan et al., 2011) that have become available since previous analyses (Huang and Gogarten, 2007; Becker et al., 2008; Moustafa et al., 2008). The set of predicted proteins from all available Chlamydiales genomes (28,244 proteins) was used to query a local database that contained annotated protein sequences from RefSeq release 37 and genome data from other public databases (for details, see Methods). This phylogenomic analysis returned 23,470 RAxML maximum likelihood trees with clade support values for each phylogeny based on 100 bootstrap replicates. Sorting these trees at the bootstrap cutoff of 75% showed that the majority (67%) of all chlamydial proteins are of bacterial affiliation. However, 584 distinct chlamydial proteins formed a sister group to Archaeplastida (Figure 1A). The chlamydial proteins of eukaryote affiliation were grouped into clusters of protein families (conservatively, trees with greater than or equal to one protein shared among them were united into a cluster). This showed that the Archaeplastida were the major recipient of these lateral gene transfers (LTGs) with 48 protein families associated with this lineage (see Table 1 for a full list). A full display of the 48 phylogenetic trees can be found in Supplemental Data Set 1 online. Among these genes, we found 12 HGTs shared by two of the three archaeplastidial lineages and three of them shared by all three archaeplastidial clades (Figure 1B). Changing the bootstrap cutoff for Chlamydiales-Archaeplastida monophyly to 90% still yielded 30 protein families. At the 75% cutoff, within the Archaeplastida, the Chloroplastida (green algae and land plants) contain most of chlamydial genes followed by Rhodophyta (red algae). These numbers reflect the current availability of genome data (see Supplemental Table 1 online) with animals (976,563 proteins in the database) and Chloroplastida (506,307 proteins) being data rich and other groups, such as haptophytes (96,133 proteins), cryptophytes (38,939), and glaucophytes (57,737 proteins), currently being relatively data poor. The data-rich fungi (451,434 proteins) contain only 14 chlamydial genes and the Amoebozoa (131,139 proteins), which are frequently infected by chlamydial species, contain only 12 genes (Horn, 2008). Hence, the Archaeplastida are significantly enriched in chlamydial LGTs, with several dozen distinct genes from Chlamydiales present in this lineage.

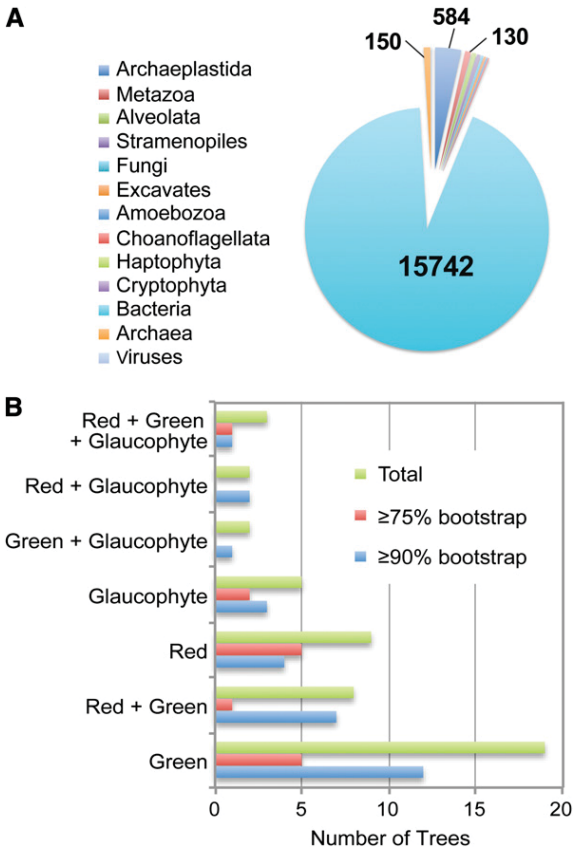


Figure 1. Chlamydial Genes in Archaeplastida.

(A) Distribution of BLASTx hits to proteins encoded on chlamydial genomes. The vast majority of proteins are, as expected, of bacterial affiliation, whereas the second largest class of hits is to Archaeplastida.
(B) The sister group relationships of Chlamydiales-derived proteins in Archaeplastida at bootstrap cutoff ≥90%, ≥75%, and when these numbers are added, including the four genes shared exclusively by Chlamydiales and Archaeplastida (see Table 1)

The Search for Parasite Candidate Genes That Could Have Participated in Plastid Endosymbiosis

The stable maintenance of energy parasites during the metabolic integration of plastids suggests that a chlamydia-like organism that infected the ancestor of the Archaeplastida adopted an essential function in establishment of the cyanobiont. This essential function likely existed at the time when the cyanobiont entered the host through phagocytosis and established the initial metabolic connection with the eukaryote. Recent phylogenetic and biochemical evidence suggest that a likely candidate was export of the bacterial-specific metabolite ADP-Glc from the cyanobiont to the host cytosol (Deschamps et al., 2008a) where the latter could be stored by conversion to glycogen. The immediate flow into the osmotically inert cytosolic carbon stores

PERSPECTIVE

Table 1. Minimal List of Candidate HGTs between Chlamydiales and Archaeplastida

Eggnog	Description	Pfam	RefSeq	Bootstrap	Topology
COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	PF01171	gi16752030	90	G
COG0039	Malate/lactate dehydrogenases	PF00056	gi15605100	90	G
COG0042	tRNA-dihydrouridine synthase	PF01207	gi46400854	90	G
COG0144	tRNA and rRNA cytosine-C5-methylases	PF01189	gi297620390	90	R+Gl
COG0162	Tyrosyl-tRNA synthetase	PF00579	gi15834952	90	G+R
COG0217	Uncharacterized conserved protein	PF01709	gi46400603	90	G+R
COG0275	Predicted *S*-adenosylmethionine–dependent methyltransferase involved in cell envelope biogenesis	PF01795	gi46399586	90	G
COG0304	3-Oxoacyl-(acyl-carrier-protein) synthase	PF00109	gi298537950	90	G+R
COG0306	Phosphate/sulfate permeases	PF01384	gi15605425	75	G
COG0324	tRNA δ(2)-isopentenylpyrophosphate transferase	PF01715	gi46400518	90	G
COG0448*	GlgC; ADP-Glc pyrophosphorylase[a]		gi297375973	NA	G
COG0517	FOG: CBS domain	PF01380	gi46401057	90	G
COG0527	Aspartokinases	PF00696	gi297376397	90	Gl
COG0545	FKBP-type peptidyl-prolyl *cis-trans* isomerase1	PF00254	gi281497802	75	R
COG0547	Anthranilate phosphoribosyltransferase	PF00591	gi123763216	90	G+R
COG0564	Pseudouridylate synthases, 23S RNA-specific	PF00849	gi281498107	75	R
COG0566	rRNA methylases	PF00588	gi46399416	90	G
COG0574	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	PF02896	gi281499037	90	Gl
COG0588	Phosphoglycerate mutase 1	PF00300	gi15834720	90	G
COG0590	Cytosine/adenosine deaminases	PF00383	gi281498821	75	G+R
COG0720	6-Pyruvoyl-tetrahydropterin synthase	PF01242	gi297376398	90	R+Gl
COG0812	UDP-*N*-acetylmuramate dehydrogenase	PF02873	gi15834838	90	G
COG0821	Enzyme involved in the deoxyxylulose pathway of isoprenoid biosynthesis	PF04551	gi46400015	75	G
COG1054	Predicted sulfur transferase	PF00581	gi46399653	90	G+Gl
COG1092	Predicted SAM-dependent methyltransferases	PF10672	gi281500146	75	Gl
COG1164	Oligoendopeptidase F	PF01432	gi46400453	90	R
COG1165	2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase	PF02776	gi336482699	75	R
COG1169	Isochorismate synthase	PF00425	gi336482698	90	R
COG1181	D-Ala-D-Ala ligase and related ATP-grasp enzymes	PF01820	gi46399673	90	G
COG1187	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases	PF00849	gi270284888	90	G
COG1212	CMP-2-keto-3-deoxyoctulosonic acid synthetase	PF02348	gi281500494	75	G
COG1218	3-Phosphoadenosine 5-phosphosulfate (PAPS) 3-phosphatase	PF00459	gi334695582	75	G
COG1477	Membrane-associated lipoprotein involved in thiamine biosynthesis	PF02424	gi333410213	90	R
COG1496	Uncharacterized conserved protein	PF02578	gi46401143	75	Gl
COG1523	GlgX; type II secretory pathway, pullulanase PulA and related glycosidases	PF00128	gi46400381	90	G+R+Gl
COG1576	Uncharacterized conserved protein	PF02590	gi46400983	75	R
COG1611	Predicted Rossmann fold nucleotide binding protein	PF03641	gi46399856	90	Gl
COG1878	Predicted metal-dependent hydrolase	PF04199	gi336483621	75	G
COG1947	4-Diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase	PF00288	gi46400864	90	G+R+Gl
COG2217	Cation transport ATPase	PF00122	gi15834725	90	G+R
COG2265	SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase	PF05958	gi46401273	90	G+R
COG3914	Predicted *O*-linked *N*-acetylglucosamine transferase, SPINDLY family	PF00534	gi336480908	75	R

(Continued)

PERSPECTIVE

Table 1. (continued).

Eggnog	Description	Pfam	RefSeq	Bootstrap	Topology
euNOG05012	2-C-Methyl-D-erythritol 4-phosphate cytidylyltransferase	PF01128	gi46445961	75	G+R+Gl
KOG4197	FOG: PPR repeat	PF01535	gi336482601	NA	G
NOG04817	Putative uncharacterized protein		gi281499473	90	G
NOG04985	Protein involved in glycerolipid metabolism	PF12695	gi29834553	90	R
NOG05008	Glycerol-3 protein	PF01553	gi46400593	NA	G+R+Gl
NOG87709	Glycosyltransferase family A (GT-A)	PF01704	gi46399599	NA	G+Gl

The Archaeplastida are named as follows: G, Chloroplastida; R, Rhodophyta; and Gl, Glaucophyta. These candidate HGTs were identified using an automated phylogenomic procedure that took as the selection criterion monophyly of Chlamydiales and Archaeplastida to the exclusion of any other clade (except algae that contain secondary plastids, such as chromalveolates and euglenids) with bootstrap values ≥90 or ≥75%. Trees that only contained Chlamydiales and Archaeplastida are marked with "NA." This list does not contain examples of chlamydial gene transfers that have a complex gene history, such as GlgA (this work) and the ATP translocator that has been previously described. The trees listed here are available in the same order in Supplemental Data Set 1 online. It should be noted that the putative gene functions result from an automated annotation procedure and should be interpreted with caution.

[a]The automated annotation GlgC-ADP-Glc pyrophosphorylase is justified by the presence of a GlgC-like domain within an otherwise unknown protein. This protein found in prasinophytes, stramenopiles, and chlamydiales does not define a classical and functional ADP-Glc pyrophosphorylase.

for delayed use by host metabolism was required because there was no possibility to adjust carbon supply and demand between the two partners (reviewed in Ball et al., 2011). This function must have required a nucleotide sugar translocator that was recruited to the cyanobacterial inner membrane from the host endomembrane system (Deschamps et al., 2008a; Colleoni et al., 2010). However, this scenario would also require a glucan synthase in the host cytosol able to use ADP-Glc to generate storage polysaccharide (Deschamps et al., 2008a; reviewed in Ball et al., 2011). The presence of this enzyme, deduced from pathway reconstruction, is surprising because eukaryotes lack the ability to synthesize or use the bacterium-specific ADP-Glc metabolite.

Extant green plants and algae synthesize starch from a minimum of five enzymes named GBSS, SSI, SSII, SSIII, and SSIV. Figure 2 shows that these enzymes fall into two major monophyletic groups consisting on the one hand of GBSSI, SSI, and SSII on their own (referred to as the GBSS-SSI-SSII clade) and on the other hand of SSIII, SSIV, and a number of related bacterial sequences from proteobacteria cyanobacteria and chlamydiales (the whole group being referred to as the SSIII-IV clade). The source of the GBSSI SSI SSII clade can be reasonably assumed to be the ancestor of the form of GBSSI found in extant cyanobacteria (Deschamps et al., 2008a; Ball et al., 2011). Indeed, if the entire clade is rooted with such a source in mind, then duplication of the chloroplastidial GBSSI followed by mutations that yielded soluble enzymes explain the selective appearance of SSI-SSII in the green algae. This must have coincided with the complex rewiring of starch metabolism from the cytosol to the chloroplast when the chloroplastidial lineage diverged (Deschamps et al., 2008b; reviewed in Ball et al., 2011). GBSSI has been recently discovered in a group of clade B unicellular diazotrophic cyanobacteria (Deschamps et al., 2008a) whose ancestors, according to Gupta (2009), could qualify as the plastid source. The biochemical properties of this diazotrophic cyanobacterium have been studied in detail (Deschamps et al., 2008a). The cyanobacterial enzyme has similar functions and biochemical properties to the plant enzyme, with one exception. This is its high selectivity for ADP-Glc; the cyanobacterial enzyme has no residual activity with UDP-Glc. This is in contrast with all archaeplastidial enzymes that use both purine and pyrimidine nucleotide sugars, unlike the soluble starch synthases of plants, red algae, and glaucophytes (Leloir et al., 1961; Deschamps et al., 2006; Shimonaga et al., 2007; Plancke et al., 2008). This observation is good agreement with both a proposed cyanobacterial origin of the *GBSSI* gene and an ancestral dual substrate pathway in the cytosol of the common ancestor. GBSSI is well known to display very low (nonphysiological) activity as a soluble enzyme (Dauvillée et al., 1999; Edwards et al., 1999). Since there are no reported examples of eukaryotes not derived from Archaeplastida that accumulate starch, we assume that at the very onset of plastid endosymbiosis the cytosolic pools consisted of glycogen that would not have supported GBSSI activity, thereby disqualifying it from early recruitment to polysaccharide biosynthesis. Because *SSI-SSII* originated from *GBSSI*, duplicated in the green lineage, and because analogous proteins have not been found elsewhere, despite searching hundreds of fully sequenced bacterial genomes, we also conclude that genes encoding SSI-SSII were not available at the onset of endosymbiosis.

This leaves the ancestor of the SSIII-SSIV clade as the only extant enzyme whose ancestor could have been recruited at the onset of plastid endosymbiosis. This possibility is further strengthened by the finding of a homologous sequence in the genome of the glaucophyte *Cyanophora paradoxa* where it appears to be involved in starch synthesis using ADP-Glc in the cytosol (Price et al., 2012).

Previously, SSIII-IV was reported as an enzyme of chlamydial origin from examination of the sequence of the amoebal parasite *Protochlamydia amoebophila* (Moustafa et al., 2008). However, inclusion of the more diverged enzymes from animal parasites and the new sequence information from additional genomes

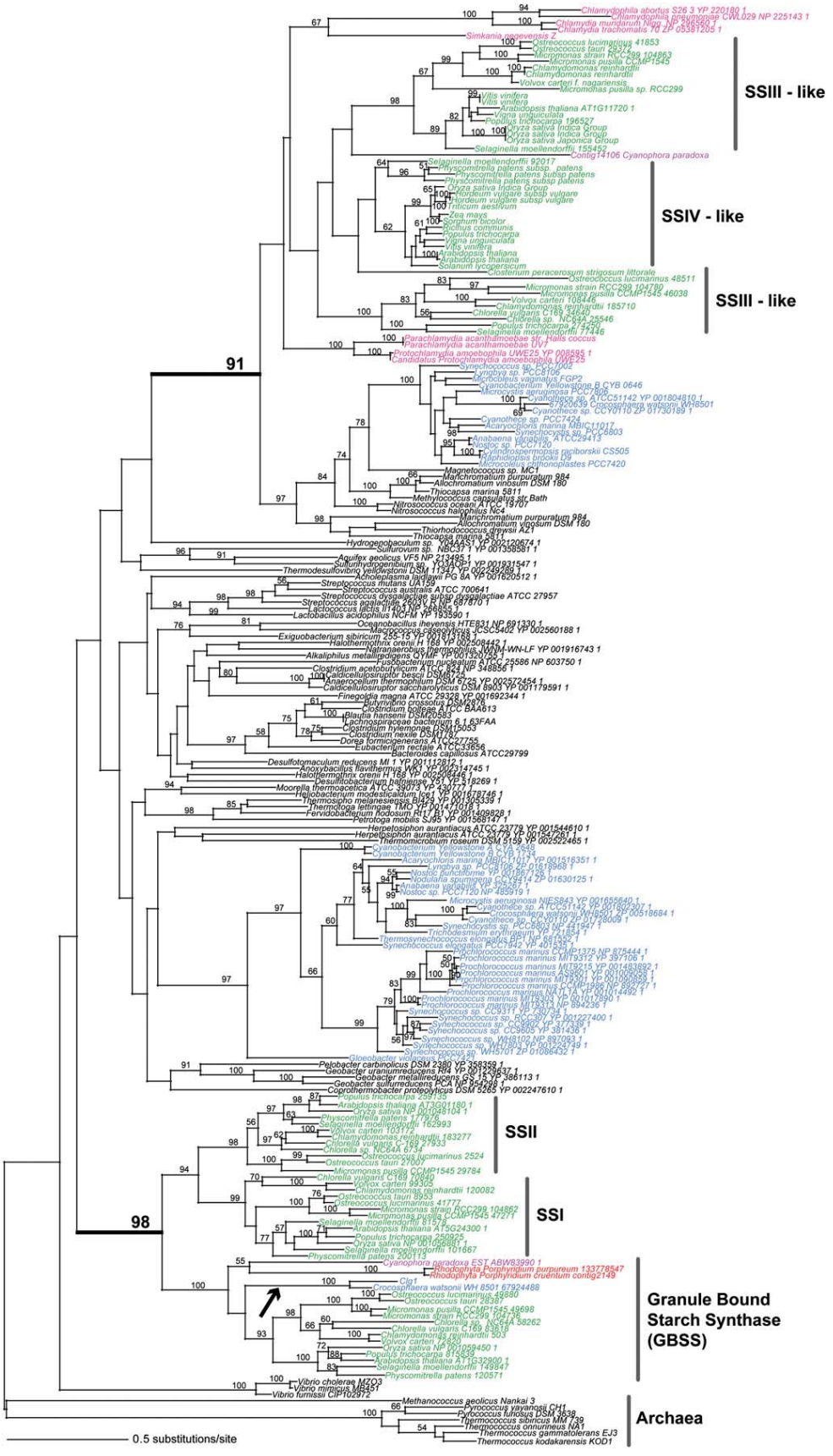


Figure 2. Phylogenetic Analysis of ADP-Glc–Specific Starch Synthases of Bacteria and Archaeplastida.

PERSPECTIVE

suggests a more complex SSIII-SSIV phylogeny (Figure 2; see Supplemental Figure 1, Supplemental Table 2, and Supplemental Methods 1 online for the complex phylogeny of the glucan synthase). Our results suggest strongly that the source of the entire SSIII-IV enzyme clade (inclusive of a small subset of proteins from proteobacteria and cyanobacteria) was an enzyme form that evolved in chlamydial ancestors. The phylogeny is complicated further by the presence of several distinct LGT events. Indeed, an additional LGT from Chlamydiales to proteobacteria is evidenced and well supported by the phylogenetic evidence. This sequence was subsequently passed on from proteobacteria to a subset of cyanobacteria. In all these bacteria, SSIII-IV is one of several glycogen synthase types present, while Chlamydiales define the only group of organisms that consistently have the SSIII-IV enzyme as their sole glucan synthase. The phylogeny is in agreement with a chlamydial source and its transfer from the parasites to a subset of proteobacteria and from there to a subset of cyanobacteria. In addition, the Archaeplastida have also received a copy of a gene encoding the chlamydial enzyme. However, erosion of phylogenetic signal and low bootstrap support does not allow one to distinguish between direct transfer of SSIII-IV from Chlamydiales from indirect transfer via a proteobacterial intermediate. Nevertheless, cyanobacteria can be safely ruled out as potential donors. This complexity suggests that the chlamydial enzyme probably harbors an unusual biochemical property responsible for its wide dissemination and success in the bacterial world.

Bioinformatic Effector Predictions

Why would a chlamydial glucan-synthase be present in the host cytosol at the onset of plastid endosymbiosis? A clue is provided by the fact that all chlamydia-like pathogens are known to secrete a large number of protein effectors into their host cytosol. Hence, if a particular effector played a key role in establishing symbiotic carbon flux, it would have engaged the parasite, the cyanobiont, and the host in a tripartite symbiosis, each genome being required for the coding of enzymes establishing the common essential carbon flux. Virulence effectors are secreted by Chlamydiales through the type III secretion system that has been present for >1 billion years in chlamydia-like parasites that infect protists (Horn et al., 2004; reviewed in Horn, 2008). This syringe-like macromolecular machine, known as the type three secretion system (TTS), spans the two bacterial membranes and the inclusion membrane of the host that surrounds the parasites. Hence, if enzymes of chlamydial glycogen metabolism were being secreted into the host cytosol, they could have participated in the key biochemical events that established plastid endosymbiosis.

It is noteworthy that the majority of Chlamydiales contain a full suite of enzymes for glycogen metabolism (Iliffe-Lee and McClarty, 2000), even though very few appear to produce glycogen autonomously or to accumulate the polymer internally. This is in contrast with the majority of intracellular pathogens that have lost the enzymes of glycogen metabolism (Henrissat et al., 2002). The relevant enzymes include ADP-Glc pyrophosphorylase (AGPase; GlgC), glycogen synthase (GlgA), glycogen branching enzyme (GlgB), one or more glycogen phosphorylases (GlgP and MalP), a GlgX-type of direct debranching enzyme related to plant isoamylases, and an amylomaltase required to metabolize the oligosaccharides released through GlgX (MalQ). Because the genes encoding these enzymes of glycogen metabolism have been retained in the highly streamlined and reduced genomes of the Chlamydiales (1 Mb for the animal parasites; Horn, 2008), they likely serve an important function in these organisms. One possibility is that some enzymes in this pathway are cytosolic effectors that manipulate host glycogen metabolism.

Based on the proposed relationship between cytosolic glycogen metabolism and establishment of the endosymbiotic link, we used four TTS prediction algorithms (Arnold et al., 2009; Löwer and Schneider, 2009; Samudrala et al., 2009; Wang et al., 2010) to study all the enzymes of glycogen metabolism identified in the genome sequences of environmental Chlamydiales.

Figure 2. (continued).

Unrooted GlgA (glycogen starch synthase) tree using the best-fit protein model (LG + G) and 1000 RAxML bootstrap replicates (alignments are given in Supplemental Data Sets 2 to 4 online). To identify the source of the ADPGlc-using glucan synthase involved in endosymbiosis, the phylogeny of archaeplastidal starch synthases compared with bacterial glycogen synthases was determined using maximum likelihood analysis with 1000 replica bootstrap analysis. Extant archaeplastidal candidates whose ancestors may have triggered the establishment of the symbiotic link are those enzymes that can be linked phylogenetically to ADP-Glc–requiring enzymes conserved in green algae and land plants (in green), which group into three classes known as granule-bound starch synthase (GBSS in the bottom part of figure), the soluble starch synthase (SS) SSI/SSII class (in the bottom part of the figure), and the SSIII/SSIV class (in the top and middle parts of the figure; reviewed in Ball and Morell (2003). The high bootstrap nodes unifying the GBSS-SSI-SSII and the SSIII-IV clades are highlighted in bold. Rhodophyceae starch synthase progenitors are not present in this tree because those organisms have lost the ability to polymerize starch from ADP-Glc and have retained only the very different host-derived soluble glycogen (starch) synthase that uses UDP-Glc (reviewed in Ball et al., 2011). However, GBSSI was transmitted to red algae (in red) and Glaucophyta (in dark blue) where it polymerizes amylose from UDP-Glc in the cytosol. These proteins still display the ability to use ADP-Glc. The source of the GBSSI, SSI, and SSII clade is proposed to be cyanobacterial and is displayed by an arrow (for reasons sustaining this source, see text and Deschamps et al., 2008a; Ball et al., 2011). As to the SSIII/SSIV clade, SSIII/SSIV was previously reported to result from an HGT from environmental chlamydia-like amoeba parasites *P. amoebophila* and *P. acanthamoeba* (Moustafa et al., 2008). However, the phylogeny displayed here suggests the presence of several LGT events not previously evidenced. Because of the importance of the issue, we devoted the Supplemental Methods 1 online to the detailed analysis of these results.

PERSPECTIVE

We focused on the amoebal parasites *P. amoebophila* and *Parachlamydia acanthamoeba* because these taxa are the most frequently recovered sisters to archaeplastidial enzymes of chlamydial origin (Moustafa et al., 2008). The results (see Supplemental Table 3 online), summarized in Table 2, show that the isoamylase encoded by *GlgX* was found by all prediction programs to be a putative effector, AGPase, the enzyme that synthesizes ADP-Glc (encoded by *GlgC*) was scored as a potential effector in either *P. amoebophila* or *P. acanthamoeba*, respectively, by the PETS and effective T3 programs. Glycogen phosphorylase in *P. acanthamoeba* was identified as a candidate effector by the effective T3 program, whereas MalQ in *P. amoebophila* was identified as a potential effector by both the SIEVE and PETS programs.

In Vivo Validation of Bioinformatic Predictions

To test these bioinformatic predictions, we relied on a previously established in vivo assay in *Shigella flexneri* (Subtil et al., 2001). This assay relies on the conservation of the structural determinants of the TTS within bacteria. *S. flexneri* is used to establish the presence of an N-terminal sequence that allows secretion through the TTS of a fused reporter gene, the calmodulin-dependent adenylate cyclase reporter from *Bordetella pertussis*. This system has been validated in screens for TTS in *Chlamydia pneumoniae* with a rate of false positives below 5% (Subtil et al., 2005). To test whether proteins of glycogen metabolism of *P. acanthamoeba* and *P. amoebophila* contained a functional TTS signal, we made fusions between the first 30 codons from each gene and the reporter gene and tested secretion of the reporter by *S. flexneri* (see Supplemental Table 4 online). The results shown in Supplemental Table 4 and Figure 3 not only confirm that GlgX (debranching enzyme) and GlgC/GlgP (ADP-Glc pyrophosphorylase and glycogen phosphorylase) are recognized as TTS substrates, but they also show that GlgA (glycogen synthase) carries a secretion signal recognized by *S. flexneri*. As a negative control for the secretion assay, we deleted the first 10 amino acids of GlgA and showed that the corresponding fusion protein was no longer secreted through the TTS (Figure 3).

Consequences of the Effector Nature of the Enzymes of Glycogen Metabolism of Chlamydiales on the Priming of Plastid Endosymbiosis

The bioinformatic predictions and experimental evidence implied that combinations of parasite enzymes could work in combination to hijack host metabolism. Chlamydia-like pathogens potentially use ADP-Glc pyrophosphorylase, glycogen phosphorylase, and glycogen synthase to increase flux to glycogen in the cytosol of their hosts and recover maltotetraose through the action of the GlgX-type of glycogen debranching enzyme. ADP-Glc is an energy-rich, bacterial-specific metabolite, unrecognized by eukaryotes, whose costly synthesis would be induced in the cytosol by these putative chlamydial effectors (Figure 4), likely thereby weakening host defense against the pathogen. Maltotetraose is normally not produced by eukaryotic cytosolic glycogen metabolism, and no MOS malto-oligosaccharides degrading enzyme other than the eukaryotic dpe2 type of amylomaltase is likely to be present in this compartment (although glucosidases have been

Table 2. Bioinformatic Predictions and in Vivo Validation of Chlamydial Putative Effectors

Protein	BPBAac	SIEVE	PETS	Effective T3	Secretion Assay in *Shigella*
P. amoebophila GlgA	−	−	−	−	Not tested (highly conserved)
P. acanthamoeba GlgA	−	−	−	−	+
P. amoebophila GlgB	−	−	−	+	−
P. acanthamoeba GlgB	−	−	+	-	−
P. amoebophila GlgC	N/A	N/A	N/A	N/A	N/A
P. acanthamoeba GlgC	−	−	+	+	+
P. amoebophila GlgP	−	−	−	−	Not tested
P. acanthamoeba GlgP-1	−	−	−	+	Chimera not expressed
P. acanthamoeba GlgP-2	−	−	−	−	+
P. amoebophila GlgX	+	+	+	+	+
P. acanthamoeba GlgX	N/A	N/A	N/A	N/A	N/A
P. amoebophila MalQ	−	+	+	−	+
P. acanthamoeba MalQ	−	−	−	−	−

The different putative effector prediction programs were used respectively with a cutoff of 0.4 (PETS), a SIEVE percentage score above 1, and an effector T3% score above 90% (Arnold et al., 2009; Löwer and Schneider, 2009; Samudrala et al., 2009; Wang et al., 2010). N/A refers to the presence of too many uncertainties on the translation start site to allow prediction or in vivo testing. We were unable to test GlgX from *P. acanthamoeba* and GlgC from *P. amoebophila* because of uncertainties in the reported N-terminal sequence deduced from the genome sequence. Because the N-terminal sequence of GlgA from *P. acanthamoeba* and *Protochlamydia* was highly conserved, we chose to test only the *Parachlamydia* sequence. GlgA, glycogen synthase; GlgB, glycogen branching enzyme; GlgC, ADP-Glc pyrophosphorylase; GlgX, debranching enzyme (isoamylase-like); MalQ, α-1,4 glucanotransferase (DPE2 maltase type).

PERSPECTIVE

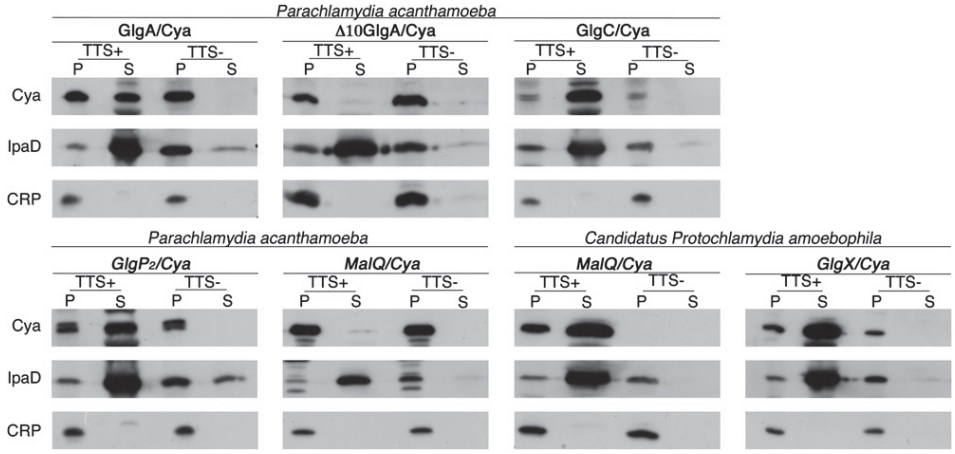


Figure 3. Chlamydial Enzymes Encode Functional TTS Signals.

The N-terminal 30 codons (see Supplemental Table 4 online) of the indicated genes were cloned upstream of *cya* and expressed in the *ipaB* and *mxiD S. flexneri* strains. The *ipaB* strain constitutively secretes TTS substrates (labeled TTS+), whereas the *mxiD* strain is defective for TTS (labeled TTS−) (Subtil et al., 2001). Exponential cultures expressing the fusion protein were fractionated into supernatants (S) and pellets (P). Samples were resolved by SDS-PAGE gels, transferred to a polyvinylidene difluoride (PVDF) membrane, and probed with anti-Cya to detect the chimera. Probing the membrane with anti-IpaD showed that TTS in the *ipaB* background was not impaired by transformation of the various constructs. Antibodies against the cAMP receptor protein (CRP) were used to control for bacterial lysis during fractionation. A GlgA construct in which the first 10 amino acids were deleted (Δ10GlgA/Cya) was included as a negative control. When expressed in the *ipaB* strain, the Δ10GlgA/Cya chimera was not secreted in the supernatant, in contrast with the GlgA/Cya chimera. In the *mxiB* (TTS defective) background, none of the chimeric proteins were recovered in the culture supernatant, demonstrating that their secretion in the *ipaB* background occurred by a TTS mechanism.

described in other eukaryotic compartments). Hence, the bacterial effectors would induce the synthesis of glycogen through ADP-Glc at the beginning of the infection cycle when the ratio of cytosolic ATP to Pi is high. At later stages when this ratio decreases, the presence of the unregulated glycogen phosphorylase of the chlamydial pathogen will bypass the highly regulated host enzyme and trigger conversion of the cytosolic glycogen pools to Glc-1-P and a carbon substrate (maltotetraose) that can be metabolized only by the pathogen. It is relatively straightforward to envision how introduction of a cyanobacterium into the host via phagocytosis would rapidly rescue a cell infected by such a pathogen and effectively convert the latter to a symbiont (Figure 5). The cyanobacterium would generate large amounts of ADP-Glc as a product of photosynthetic metabolism. This high-energy compound could flood the host metabolic system with substrate to fuel flux through glycogen and into the parasite, without draining host energy resources. Only one critical event would be required: recruitment of a host-derived nucleotide sugar translocator to the inner envelope of the cyanobacterium. Extant eukaryote transporters are able to facilitate movement of ADP-Glc across membranes even though they do not use that particular nucleotide sugar (Colleoni et al., 2010). The endosymbiotic link would thus be established within a tripartite system without the need for a novel protein or a HGT event followed by foreign gene expression in the host. Under this "ménage à trois" scenario, just two chlamydial functions are required to establish symbiosis: the glucan synthase that feeds the exported carbon into the glycogen pool (GlgA) and the GlgX debranching enzyme that generates the maltooligosaccharides required to feed the chlamydial symbionts, thereby maintaining its genes required for symbiosis. Hence, all other chlamydial glycogen metabolism effectors (GlgC, GlgP, and MalQ) may have been counter-selected after establishment of the tripartite symbiosis as they were no longer necessary following the switch from pathogenesis to symbiosis. Indeed, when the system had switched to symbiosis, it was no longer necessary for the chlamydia to tap ATP and Glc-1-P from host metabolism through the use of its GlgC (AGPase) effector, and there would likely be selection against unregulated breakdown of cytoplasmic glycogen pools by the chlamydial phosphorylase (GlgP). In agreement with this observation, the phylogenetic origin of other components of storage polysaccharide breakdown in extant Archaeplastida is of host (Eukaryotic) derivation (Deschamps et al., 2008a).

Isoamylase (GlgX) Defines a Chlamydial Effector Common to all Three Archaeplastidial Lineages

The isoamylase gene of Archaeplastida may have been derived from the chlamydial parasite's GlgX protein (Brinkman et al., 2002; Huang and Gogarten, 2007; Becker et al., 2008; Moustafa et al., 2008). The Chlamydial protein has been reported to display structural properties analogous to the *Escherichia coli* GlgX

PERSPECTIVE

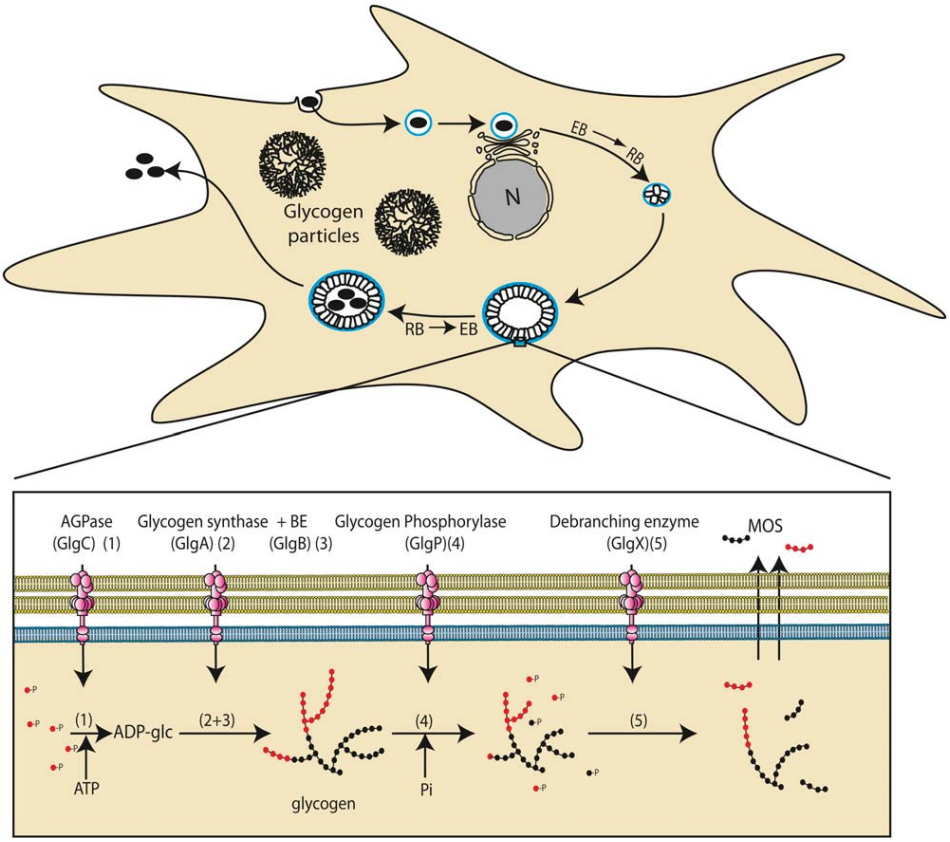


Figure 4. Glycogen Metabolism Putative Effectors in Extant and Ancient Chlamydiales.

The infection cycle of a typical eukaryote by a chlamydia-like organism is shown. Chlamydiales are internalized through phagocytosis. The pathogen modifies the surrounding vesicle into an inclusion membrane (shown in blue). These bacteria undergo two successive developmental forms: the infectious elementary body (EB) and the replicative reticulate body (RB). At the end of the cycle, reticulate bodies differentiate back to elementary bodies that are eventually released from the cell. Both elementary bodies and reticulate bodies possess a type III secretion apparatus that spans the two bacterial membranes and the inclusion membrane, allowing for the secretion of bacterial proteins directly into the host cytosol. In the enlargement, the inclusion membrane is depicted in blue, type III secretion apparatuses in pink, the eukaryote cytosol in yellow, and the inside of the chlamydial cell in white. The Glc residues of the host glycogen used as primers for synthesis are symbolized by black dots. Red dots represent accessible Glc, whereas the Glc-1-P generated by glycogen phosphorylase is represented by red dots with an attached "P." The sequential nature of the biochemical reactions is depicted as numbers 1 to 5. The enlargement displays the particular type of putative effector combination present in extant *P. acanthamoeba* and possibly in the ancient parasite that would have predisposed a host cell to endosymbiosis. Briefly, carbon is diverted to the glycogen pools through the chlamydial ADP-Glc pyrophosphorylase putative effector (GlgC) making use of the high host cytosolic ATP and Glc-1-P pools at the beginning of the infection cycle. Polymerization into glycogen from ADP-Glc occurs through the chlamydial glycogen synthase (GlgA). The host glycogen synthase only uses UDP-Glc and is not represented. Branching could occur through the action of either the host or parasite branching enzyme, although we did not find clear evidence for a type III secretion signal in the two parasite branching enzymes tested (GlgB), indicating that branching might solely involve host enzymes. When the orthophosphate concentration rises and the cytosolic ATP and Glc-1-P decreases as a consequence of the infection, parasite phosphorylase (GlgP) will recess the outer chains of glycogen, terminating four residues away from each branch. This will allow the action of GlgX, whose possible substrates are restricted to such chains. The maltotetraose generated by GlgX is normally not metabolized by eukaryotes in the cytosol and could therefore be a substrate for import and catabolism within the parasites. The transporter for the import of MOS in the parasite is presently unknown. We omitted MalQ from the drawing since we do not know if the chlamydial version of this enzyme would behave more like a maltase or disproportionate type of α-1,4 glucanotransferase.

PERSPECTIVE

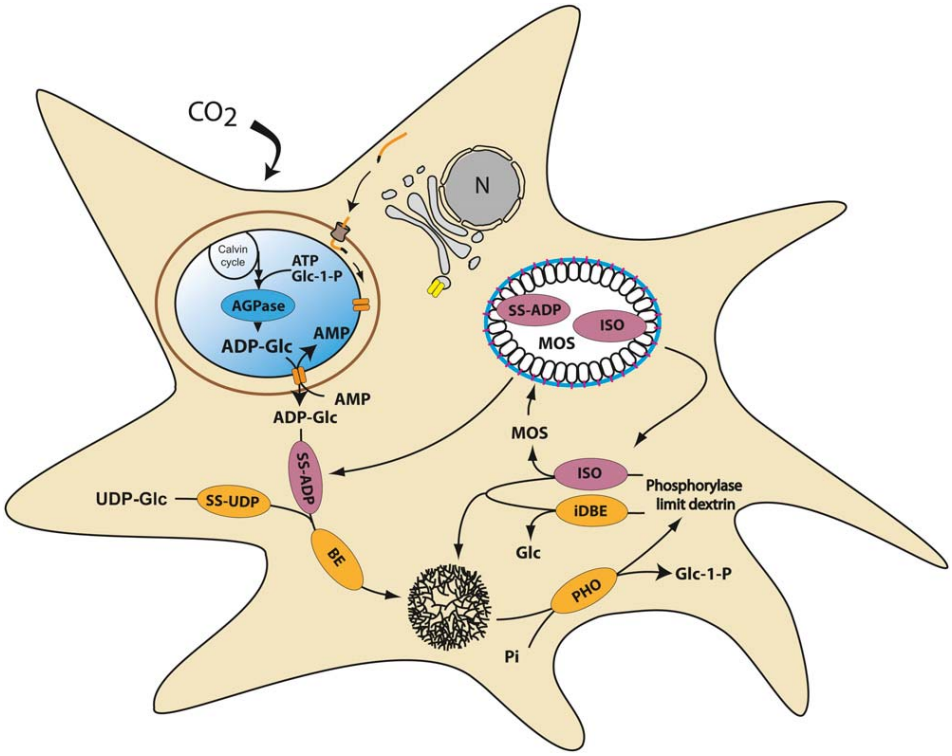


Figure 5. Ménage à Trois.

The interdependent symbiosis between the cyanobiont, its host, and a chlamydial parasite is shown, displaying an effector combination akin to that hypothesized in Figure 4. Enzymes are colored with respect to their phylogenetic origin: in orange for host enzymes of eukaryotic glycogen metabolism, in pink for chlamydial effectors, and in blue for cyanobacterial enzymes. Only those enzymes of chlamydial origin that are required to establish the symbiotic flux and feed carbon into the parasite or the host are displayed. The other enzymes depicted in Figure 4 do not interfere with this biochemical flux. The GlgX chlamydial debranching enzyme is represented here by the name "iso" to denote its present function in Archaeplastida (isoamylase). The cyanobiont in blue exports the bacterial specific metabolite ADP-Glc by recruiting a family III nucleotide sugar translocator (colored in orange) from the host endomembrane system (colored in yellow) as recently proposed (Weber et al., 2006; Deschamps et al., 2008a; Colleoni et al., 2010). A possible primitive TOC (translocon of the outer chloroplast membrane) targeting system is drawn in gray. The ADP-Glc would be funneled to glycogen by the chlamydial effector glycogen synthase, whereas the host would still be able to polymerize Glc into glycogen from UDP-Glc. The dramatic increase of the cytosolic glycogen pools would have benefitted the host through the increased production of Glc and Glc-1-P that can be further metabolized. However, it would have equally benefitted the parasite by increasing the supply of MOS that may only be metabolized by the latter. BE, branching enzyme; iDBE, indirect debranching enzyme; ISO, glgX type of direct debranching enzyme; PHO, phosphorylase; SS-ADP, ADP-specific glycogen synthase; SS-UDP, UDP-specific glycogen synthase.

protein, which is a maltotetraose-generating enzyme in glycogen catabolism (Jeanningros et al., 1976; Dauvillée et al., 2005). Among the plant isoamylase genes, all three isoamylases display a common origin (Figure 6; Abe et al., 1999; Hussain et al., 2003). Therefore, the anabolic function of isa1 (and isa2) in amylopectin synthesis likely evolved at a later stage, postendosymbiosis. As a result, we propose that glycogen accumulated in the ancestral symbiosis (Figure 5). In the phylogeny of GlgX, glycogen debranching enzymes (Figure 6; see Supplemental Figure 2 online), the enzymes from Archaeplastida and Chlamydiales, form a well-supported (maximum likelihood bootstrap support of 83%) monophyletic group to the exclusion of all other bacteria. This suggests a HGT event between these two lineages. The direction of gene transfer can be inferred from existing knowledge of storage polysaccharide metabolism. One of the major differences between bacterial and eukaryotic glycogen metabolism lies in the mode of hydrolysis of the α-1,6 linkages (Ball et al., 2011). Debranching enzymes that attack the α-1,6 branch directly and release debranched glucan chains are found only in bacteria and Archaea. Eukaryotes use the indirect debranching enzymes that produce Glc. No eukaryotes, with the notable exception of Archaeplastida, contain a direct type of debranching enzyme, such as GlgX or isoamylase. In addition, a candidate sequence encoding an indirect debranching enzyme was recently found in

PERSPECTIVE

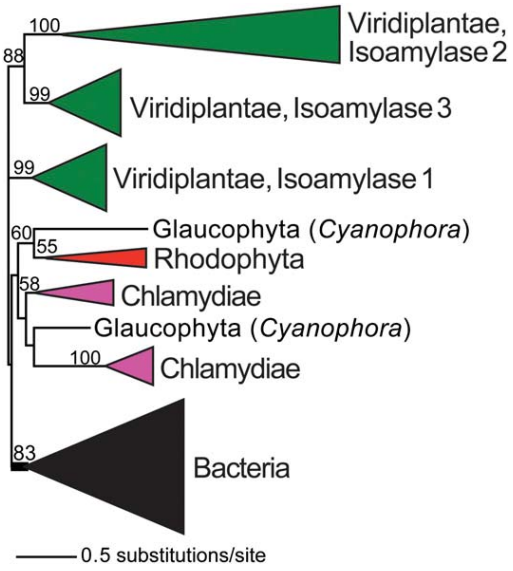


Figure 6. Phylogenetic Analysis of Bacterial and Archaeplastidal Glycogen Debranching Enzymes.

Starch debranching enzymes of Chloroplastida are known to play an important role both in polysaccharide synthesis and starch degradation. The maximum likelihood tree of these enzymes shows the three types of Chloroplastida (land plants and green algae) subunits (filled green triangles). Isoamylase1 (Isa1) is responsible for trimming misplaced preamylopectin branches that if unprocessed prevent polysaccharide crystallization, leading to glycogen formation, rather than starch. Isa2 is a noncatalytic subunit assembled in heteromultimers together with Isa1. Isa3 is involved in the breakdown of branches during starch degradation. The homologous enzymes of red algae (red filled triangle) and Glaucophyta (blue branches) are also shown in this tree. Analogous specialized roles in starch synthesis and degradation remain to be determined in these organisms. The novel glaucophyte sequences were isolated using RT-PCR. Alignments are shown in Supplemental Data Sets 3 to 5 online, and the full tree is displayed in Supplemental Figure 2 online.

the *C. paradoxa* (Glaucophyta) genome, strongly suggesting that the ancestral host of the primary endosymbiosis contained such an enzyme, enabling eukaryotic glycogen catabolism in the cytosol (Price et al., 2012). From these observations, we conclude that the direction of gene transfer was from Chlamydiales to the common ancestor of the Archaeplastida. However, the relative positions of the different Archaeplastida are unresolved in the tree. In particular, the node grouping the Chloroplastida has low bootstrap support, as is the case for other basal nodes in the Archaeplastida.A possible explanation for this lack of resolution is the complex evolutionary history of the Chloroplastida with respect to debranching enzymes. We propose that the chlamydial gene was initially used in the cytosol of the Archaeplastida common ancestor but later retargeted to the plastid in Chloroplastida (Deschamps et al., 2008b). Therefore, acceleration of sequence evolution likely happened twice in Chloroplastida and only once in the Rhodophyta and Glaucophyta. This could have led to a stronger erosion of the phylogenetic signal in the genes of Chloroplastida. Similar problems exist in the more complex GlgA phylogeny and in the GBSSI-SSI-SSII clade (see Supplemental Figure 1 online).

The ménage à trois tripartite symbiosis hypothesis predicts that only two of the glycogen metabolism effector enzymes from the parasites were required to prime endosymbiosis. These are the two chlamydial HGTs that are observed in the storage polysaccharide metabolism network of Archaeplastida. These define one of the four (five if the nucleotide transporter (NTT) is taken into account) established cases of chlamydial HGTs in all three archaeplastidal lineages and one of the 10 to 11 cases of HGTs common to two out of the three archaeplastidal lineages. The possibility that it is by coincidence that these two cases would simultaneously define the two critical components of symbiotic carbon flux and the two cases of documented common HGTs is remote, and we believe this to be as close as possible to experimental demonstration of the tripartite symbiosis hypothesis.

DISCUSSION

The Chlamydial Imprint on the Genomes of Archaeplastida

Forty-eight genes have been found in this study that are likely candidates for HGT from Chlamydiales to the common ancestor of the Archaeplastida. However, the nature of the filters used in our phylogenomic analysis have systematically excluded genes that do not satisfy the criterion of monophyly of Chlamydiales and Archaeplastida (and their secondary endosymbiosis derivatives) proteins to the exclusion of all other clades (at bootstrap 75 or higher). Genes displaying more complex phylogenies, such as the glucan synthase gene GlgA discussed here, have been excluded. Interestingly, the paradigm of a parasitic function transferred to Archaeplastida (the ATP import protein NTT of that gave the name “energy parasites” to the Chlamydiales) would also have been excluded because of additional complexities in the phylogeny of these transporters and restricted gene distribution. Hence, the impact of Chlamydiales on the plant genomes may still be underappreciated. This impact of on plant genomes seems modest (50 genes) and appears restricted when compared with that of Cyanobacteria (~1000 genes). Yet, this figure is of the same order of magnitude as the number of genes from Cyanobacteria that encode proteins not targeted to plastids and that are neither involved in photosynthesis nor in organelle maintenance and genome expression (the so-called nonplastid functions; Reyes-Prieto et al., 2006; Moustafa and Bhattacharya, 2008). In our view, the chlamydial symbiont compartment was likely not the site of an essential biochemical pathway and was maintained only because it encoded useful genes for the ménage à trois. Once all possible chlamydial genes had been transferred to the host, the compartment disappeared together with the chlamydial genome. It is well documented that

PERSPECTIVE

eukaryotic lineages that have lost a plastid, have, as a consequence, lost the majority of their genes of cyanobacterial origin. Indeed, the nuclear genome of *Cryptosporidium parvum*, an apicomplexa parasite that has lost the apicoplast, has only five genes of possible cyanobacterial (two genes) or plant ancestry (three genes) (Huang et al., 2004). In this context, the maintenance of a 50 gene signatures in the Archaeplastida genomes underlines the biochemical importance of chlamydial gene products in the maintenance of the symbiosis.

The Ménage à Trois Allowed Unprecedented Levels of Metabolic Integration

After metabolic stabilization of the tripartite system, a great deal of subsequent evolution, including extensive HGT, was necessary before reaching the true photoautotrophic eukaryote that was the direct ancestor of extant archaeplastidial lineages. During that time, enzyme functions derived from the host, the cyanobiont, and the chlamydial symbiont were likely integrated and optimized through a combination of HGTs and novel effector evolution in the cytosol or elsewhere, including the membranes and stroma of the cyanobiont. The continued presence of a functional, temperate chlamydial symbiont would have been required for so long as the chlamydiae encoded useful effector genes. However, HGTs to the host nucleus of these effector genes would have rendered the chlamydial partner dispensable, leading ultimately to loss of the symbiont.

We can therefore expect that, as the ancestral archaeplastidial lineages diversified, some lineages lost their chlamydial symbiont early on because the symbiotic effector genes had all been transferred to the host nucleus. However, the switch from pathogenesis to symbiosis would have prompted the positive selection of mutations of the N termini of chlamydial proteins turning them into new effector proteins. This would provide an opposing force to parasite loss as a result of HGT of effector genes to the nucleus. In some lineages, the generation of novel effectors would have kept ahead of the HGT process and ensured that the chlamydial symbiont was maintained by natural selection. It is plausible that such lineages would have been at a selective advantage because they displayed higher levels of metabolic integration. This metabolic integration may in part explain how the cyanobiont reached the status of a true cellular organelle. The race between evolution of new effectors and HGT to the nucleus also explains why different archeaplastidial lineages may not necessarily display the same number and type of chlamydial HGTs. Another process that may have altered the pathogen's signature in extant genomes consists of selective gene losses experienced by the three major archaeplastidial lineages. Of the three lineages, Rhodophyceae may be the clade that has experienced the greatest gene losses as it diverged from the other two archaeplastidial groups, while Glaucophyta may be the lineage that has maintained the highest number of genes from the ancestral protist. Starch metabolism offers an interesting example of this. For instance, Rhodophyceae have lost the ability to synthesize starch from ADP-Glc and as a consequence have lost the ancestral effector gene that established the tripartite symbiosis, while this gene was retained both in Chloroplastida and Glaucophyta. The combination of gene losses upon separation of the archaeplastidial lineages and of different timing of loss of the chlamydial symbiont can easily account for the pattern of chlamydial HGTs observed in Archaeplastida.

Our finding of at least 50 genes from Chlamydia (including the 48 shown in Table 1 and the GlgX debranching enzyme, the GlgA SSIII-SSIV, the plastid ATP translocator [NTT], and UhpC; Huang and Gogarten, 2007; Moustafa et al., 2008; Price et al., 2012) in algae and plants is consistent with long-term residency of the chlamydial symbiont in the archaeplastidial ancestor. This would provide continued selection for cyanobiont photosynthetic function to ameliorate deleterious effects of the energy parasite on host cell metabolism. Conversely, the photosynthetic function from the cyanobiont to provide abundant high-energy compounds to the tripartite system and its gradual modification to generate starch rather than glycogen would have significantly improved this partnership.

A surprising implication of our hypothesis is that the primary plastid endosymbiosis was not restricted to the host and cyanobacterial partners, nor based solely on the acquisition by the host of photosynthetic function (Huang and Gogarten, 2007; Becker et al., 2008; Moustafa et al., 2008). Rather, photoautotrophy in eukaryotes could have resulted from an increased tolerance to a particular chlamydial infection, thereby transforming an interaction from pathogenic to symbiotic (Dale et al., 2002; Wernegreen, 2004). A second aspect of our hypothesis is that the advantage conferred by endosymbiosis arose immediately after the capture of the cyanobacterium, rather than after extensive endosymbiotic gene transfer and gene activation had occurred in the host nucleus (Bodyl et al., 2007). It is noteworthy that the presence of an additional partner in endosymbiosis with the ability to secrete useful proteins in the cytosol not only expanded the repertoire of genes available for metabolic integration and production of a new organelle, but it greatly facilitated the process by making key elements available in the cytosol of the host. These factors might explain the singularity of the ancient plastid endosymbiosis (excluding *Paulinella*) (Nowack et al., 2008) that fundamentally changed the evolutionary trajectory of life on our planet.

METHODS

Phylogenomic Methods and TTS Prediction Analysis

The complete set of Chlamydiales proteins (28,244 proteins as of January 2011) available from RefSeq (Pruitt et al., 2007) was phylogenomically processed using iTree (Moustafa et al., 2010) against a comprehensive database, which was assembled from the complete RefSeq proteins and additional algal and microbial genomes from the Joint Genome Initiative. We also supplemented the database with recently published data from key lineages, such as red algae, dinoflagellates, and the glaucophyte *Cyanophora paradoxa* (Price et al., 2012). The generated phylogenetic trees were searched for trees that indicated sister relationship between

PERSPECTIVE

Chlamydiales and Archaeplastida using PhyloSort (Moustafa and Bhattacharya, 2008). Only trees with bootstrap values ≥75% were considered in subsequent analyses. The four methods of bioinformatic prediction of TTS signals are detailed elsewhere (Arnold et al., 2009; Löwer and Schneider, 2009; Samudrala et al., 2009; Wang et al., 2010).

Type III Secretion Assays

The first 30 codons of the genes under study were synthesized (Life Technologies) for cloning into puc19cya as described (Subtil et al., 2001). In three cases, the codon that initiates the translation could not be unambiguously identified, and more than 30 codons were included (*Parachlamydia acanthamoeba* GlgX, *Protochlamydia amoebophila* GlgP1, and α-1,4 glucanotransferase; see Supplemental Table 3 online). Constructs were made in *Escherichia coli* strain TG1, verified by sequencing, and transformed in *Shigella flexneri* strains SF401 and SF620 in which the *mxiD* and *ipaB* genes, respectively, have been inactivated (Subtil et al., 2001). Analysis of secretion in these strains was performed as described previously (Subtil et al., 2001). Antibodies against CRP, a cytosolic marker, were used to estimate the contamination of supernatant fractions with bacterial proteins as a result of bacterial lysis. Antibodies against IpaD, a type III secreted protein of *Shigella*, were used to verify that type III secretion occurred normally in the transformed *ipaB* strains.

Sequence Mining and Phylogenetic Tree Reconstruction

Amino acid sequences from *P. amoebophila* were first used as queries to look for similar protein sequences using BLASTP (Altschul et al., 1997) against a local genome database containing ~300 bacterial and ~60 eukaryotic complete annotated genomes. All homologous sequences (with an e-value inferior to 1.0e-05) were collected from each BLAST hit list and aligned together using the MAFFT software (Katoh et al., 2002). After automatic gaps removal in the resulting multiple alignments, maximum likelihood tree reconstructions were done using the software FastTree (Price et al., 2009). Trees were manually inspected to remove possible sequences duplicates, to refine sequence sampling, and to add additional sequences of interest that were not part the original set used for BLASTP. Trees were then realigned and recomputed with the TREEFINDER software (Jobb et al., 2004) using the LG + Gamma model. Protein sequences where collected from GenBank. Experimental trees (see Supplemental Figure 1 online) and candidate HGT trees (see Supplemental Data Set 1 online) were run with 100 bootstrap replicates. For the GlgA and GlgX trees (Figure 1; see Supplemental Figure 2 online), we used the ProtTest 2.4 to identify the best-fit model of protein evolution for our data (the LG + I + G + F model) and applied 1000 bootstrap replicates (http://www.ncbi.nlm.nih.gov/protein), the Joint Genome Institute (http://www.jgi.doe.gov/) and on genome home project websites for *Galdieria sulphuraria* (http://genomics.msu.edu/galdieria/), *Cyanidioschyzon merolae* (http://merolae.biol.s.u-tokyo.ac.jp/), and *Arabidopsis thaliana* (http://www.Arabidopsis.org/). Isoamylase sequences from *C. paradoxa* and GBSS sequences from unidentified Cyanobacterium Clg1 (GenBank taxonomy ID 197335) were produced in this study. The analysis of isoamylases from *P. acanthamoeba* showed that the presence of two open reading frames in the genome was due to an annotation error. We therefore decided to fuse these two open reading frames together.

Cloning of Isoamylase Sequences from Glaucophyta

mRNA were purified from 300 mL of exponential growth culture of *C. paradoxa* using the FastTrack 2.0 kit (Invitrogen). Reverse transcription reactions (200 ng of mRNA) were performed with SuperScript III and in the presence of GeneRacer oligo(dT) following the recommendations of the GeneRacer kit (Invitrogen). 3′ Rapid amplification of cDNA ends reactions were then performed using Platinum Taq DNA polymerase High Fidelity (Invitrogen) in the presence of 2% DMSO. Forward primers were designed according to available nucleotide sequences of contig 40664 (5′-CATCACCACACGAACTACACGGGGTGCGGGAACACGGTCAA-3′), contig 12528 (5′-GTCCGCGAGTTCAAGGAGAT-3′), and contig 38691 (5′-GTGCACGAGTTCAAGACGATGGTGCGGGAGCTGCACA-3′). PCR products were cloned in TOPO cloning vector (Invitrogen) and sent for sequencing.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Phylogenetic Analysis of the SSIII-IV ADP-Glc–Specific Glucan Synthase.

Supplemental Figure 2. Phylogenetic Analysis of the GlgX-Isoamylase Glycogen (Starch) Debranching Enzymes.

Supplemental Table 1. The Number of Protein Sequences in the Database That Was Used for the Phylogenomic Analysis to Uncover the Origins (Based on Phyla) of Chlamydiales Proteins.

Supplemental Table 2. Distribution of GT5-Glucan Synthases among SSIII-IV Containing Bacterial Species.

Supplemental Table 3. Bioinformatic Analysis of Candidate Chlamydial Effectors.

Supplemental Table 4. Sequence of the Inserts Cloned in the puc19cya Vector.

Supplemental Methods 1. The Complex Phylogeny of the SSIII-SSIV Clades.

Supplemental Data Set 1. Candidate HGT Trees.

Supplemental Data Set 2. GlgA Alignments.

Supplemental Data Set 3. GlgX Alignments.

Supplemental Data Set 4. GlgA Alignments.

Supplemental Data Set 5. GlgX Alignments.

ACKNOWLEDGMENTS

S.G.B. was supported by the Centre National de la Recherche Scientifique, the Université des Sciences et Technologies de Lille, the Région Nord Pas de Calais, and Agence Nationale de la Recherche grants ("starchevol" and "ménage à trois"). D.B. was partially supported by grants from the National Science Foundation (MGSP 0625440 and MCB 0946528). A.P.M.W. appreciates support by the German Research Foundation (CRC-TR1 and WE 2231/6-1). L.G. received financial support from Kurzfristige Auslandsstipendien (KWA) University of Vienna.

AUTHOR CONTRIBUTIONS

S.G.B. designed research and wrote the article. A.S. and L.G. performed the in vivo effector tests. D.B. and A.M. performed phylogenetic and phylogenomic analyses. A.P.M.W. analyzed the Chlamydial transporters. C.C., U.C., and M.-C.A. cloned the glaucophyte *GlgX* genes. D.D. characterized the green algal SSIII enzyme. A.S., D.B., A.P.M.W., and A.M. edited the article. A.S. and D.B. contributed equally to the work.

PERSPECTIVE

Received June 11, 2012; revised October 22, 2012; accepted January 13, 2013; published January 31, 2013.

REFERENCES

Abe, J.-I., Ushijima, C., and Hizukuri, S. (1999). Expression of the isoamylase gene of *Flavobacterium odoratum* KU in *Escherichia coli* and identification of essential residues of the enzyme by site-directed mutagenesis. Appl. Environ. Microbiol. **65:** 4163–4170.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Arnold, R., Brandmaier, S., Kleine, F., Tischler, P., Heinz, E., Behrens, S., Niinikoski, A., Mewes, H.W., Horn, M., and Rattei, T. (2009). Sequence-based prediction of type III secreted proteins. PLoS Pathog. **5:** e1000376.

Ball, S.G., Colleoni, C., Cenci, U., Raj, J.N., and Tirtiaux, C. (2011). The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J. Exp. Bot. **62:** 1775–1801.

Ball, S.G., and Morell, M.K. (2003). From bacterial to starch: Understanding the biogenesis of the plant starch granule. Annu. Rev. Plant Biol. **54:** 207–233.

Becker, B., Hoef-Emden, K., and Melkonian, M. (2008). Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. BMC Evol. Biol. **8:** 203.

Bodył, A., Mackiewicz, P., and Stiller, J.W. (2007). The intracellular cyanobacteria of *Paulinella chromatophora*: Endosymbionts or organelles? Trends Microbiol. **15:** 295–296.

Brinkman, F.S., et al. (2002). Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria and the chloroplast. Genome Res. **12:** 1159–1167.

Chan, C.X., Yang, E.C., Banerjee, T., Yoon, H.S., Martone, P.T., Estevez, J.M., and Bhattacharya, D. (2011). Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. Curr. Biol. **21:** 328–333.

Colleoni, C., Linka, M., Deschamps, P., Handford, M.G., Dupree, P., Weber, A.P.M., and Ball, S.G. (2010). Phylogenetic and biochemical evidence supports the recruitment of an ADP-glucose translocator for the export of photosynthate during plastid endosymbiosis. Mol. Biol. Evol. **27:** 2691–2701.

Dale, C., Plague, G.R., Wang, B., Ochman, H., and Moran, N.A. (2002). Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc. Natl. Acad. Sci. USA **99:** 12397–12402.

Dauvillée, D., Colleoni, C., Shaw, E., Mouille, G., D'Hulst, C., Morell, M., Samuel, M.S., Bouchet, B., Gallant, D.J., Sinskey, A., and Ball, S. (1999). Novel, starch-like polysaccharides are synthesized by an unbound form of granule-bound starch synthase in glycogen-accumulating mutants of *Chlamydomonas reinhardtii*. Plant Physiol. **119:** 321–330.

Dauvillée, D., Kinderf, I.S., Li, Z., Kosar-Hashemi, B., Samuel, M.S., Rampling, L., Ball, S., and Morell, M.K. (2005). Role of the *Escherichia coli* glgX gene in glycogen metabolism. J. Bacteriol. **187:** 1465–1473.

Deschamps, P., et al. (2008a). Metabolic symbiosis and the birth of the plant kingdom. Mol. Biol. Evol. **25:** 536–548.

Deschamps, P., et al. (2006). Nature of the periplastidial pathway of starch synthesis in the cryptophyte *Guillardia theta*. Eukaryot. Cell **5:** 954–963.

Deschamps, P., Haferkamp, I., d'Hulst, C., Neuhaus, H.E., and Ball, S.G. (2008b). The relocation of starch metabolism to chloroplasts: When, why and how. Trends Plant Sci. **13:** 574–582.

Edwards, A., Borthakur, A., Bornemann, S., Venail, J., Denyer, K., Waite, D., Fulton, D., Smith, A., and Martin, C. (1999). Specificity of starch synthase isoforms from potato. Eur. J. Biochem. **266:** 724–736.

Fournier, G.P., Huang, J., and Gogarten, J.P. (2009). Horizontal gene transfer from extinct and extant lineages: Biological innovation and the coral of life. Philos. Trans. R. Soc. Lond. B Biol. Sci. **364:** 2229–2239.

Gupta, R.S. (2009). Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. Int. J. Syst. Evol. Microbiol. **59:** 2510–2526.

Henrissat, B., Deleury, E., and Coutinho, P.M. (2002). Glycogen metabolism loss: A common marker of parasitic behaviour in bacteria? Trends Genet. **18:** 437–440.

Horn, M. (2008). *Chlamydiae* as symbionts in eukaryotes. Annu. Rev. Microbiol. **62:** 113–131.

Horn, M., et al. (2004). Illuminating the evolutionary history of chlamydiae. Science **304:** 728–730.

Huang, J., and Gogarten, J.P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol. **8:** R99.

Huang, J., Mullapudi, N., Lancto, C.A., Scott, M., Abrahamsen, M.S., and Kissinger, J.C. (2004). Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. Genome Biol. **5:** R88.

Hussain, H., Mant, A., Seale, R., Zeeman, S., Hinchliffe, E., Edwards, A., Hylton, C., Bornemann, S., Smith, A.M., Martin, C., and Bustos, R. (2003). Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. Plant Cell **15:** 133–149.

Iliffe-Lee, E.R., and McClarty, G. (2000). Regulation of carbon metabolism in *Chlamydia trachomatis*. Mol. Microbiol. **38:** 20–30.

Jeanningros, R., Creuzet-Sigal, N., Frixon, C., and Cattaneo, J. (1976). Purification and properties of a debranching enzyme from *Escherichia coli*. Biochim. Biophys. Acta **438:** 186–199.

Jobb, G., von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. **4:** 18.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30:** 3059–3066.

Leloir, L.F., De Fekete, M.A., and Cardini, C.E. (1961). Starch and oligosaccharide synthesis from uridine diphosphate glucose. J. Biol. Chem. **236:** 636–641.

Löwer, M., and Schneider, G. (2009). Prediction of type III secretion signals in genomes of gram-negative bacteria. PLoS ONE **4:** e5917.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA **99:** 12246–12251.

Moustafa, A., and Bhattacharya, D. (2008). PhyloSort: A user-friendly phylogenetic sorting tool and its application to estimating

the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. BMC Evol. Biol. **8:** 6.

Moustafa, A., Bhattacharya, D., and Allen, A.E. (2010). iTree: A high-throughput phylogenomic pipeline. Biomedical Engineering Conference (CIBEC), 2010 5th Cairo International, pp. 103–107.

Moustafa, A., Reyes-Prieto, A., and Bhattacharya, D. (2008). Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS ONE **3:** e2205.

Nowack, E.C., Melkonian, M., and Glöckner, G. (2008). Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. Curr. Biol. **18:** 410–418.

Plancke, C., et al. (2008). Pathway of cytosolic starch synthesis in the model glaucophyte *Cyanophora paradoxa.* Eukaryot. Cell **7:** 247–257.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. **26:** 1641–1650.

Price, D.C., et al. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science **335:** 843–847.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. **35** (Database issue)**:** D61–D65.

Reyes-Prieto, A., Hackett, J.D., Soares, M.B., Bonaldo, M.F., and Bhattacharya, D. (2006). Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr. Biol. **16:** 2320–2325.

Reyes-Prieto, A., Weber, A.P.M., and Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. **41:** 147–168.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15:** 1325–1330.

Samudrala, R., Heffron, F., and McDermott, J.E. (2009). Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. PLoS Pathog. **5:** e1000375.

Sato, N., Ishikawa, M., Fujiwara, M., and Sonoike, K. (2005). Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. Genome Inform. **16:** 56–68.

Shimonaga, T., Fujiwara, S., Kaneko, M., Izumo, A., Nihei, S., Francisco, P.B., Jr., Satoh, A., Fujita, N., Nakamura, Y., and Tsuzuki, M. (2007). Variation in storage alpha-polyglucans of red algae: Amylose and semi-amylopectin types in Porphyridium and glycogen type in Cyanidium. Mar. Biotechnol. (NY) **9:** 192–202.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., and Davis, R.W. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:** 754–759.

Subtil, A., Delevoye, C., Balañá, M.E., Tastevin, L., Perrinet, S., and Dautry-Varsat, A. (2005). A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates. Mol. Microbiol. **56:** 1636–1647.

Subtil, A., Parsot, C., and Dautry-Varsat, A. (2001). Secretion of predicted Inc proteins of *Chlamydia* pneumoniae by a heterologous type III machinery. Mol. Microbiol. **39:** 792–800.

Wang, Y., Zhang, Q., Sun, M.A., and Guo, D. (2011). High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. Bioinformatics **27:** 777–784.

Weber, A.P.M., Linka, M., and Bhattacharya, D. (2006). Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot. Cell **5:** 609–612.

Wernegreen, J.J. (2004). Endosymbiosis: Lessons in conflict resolution. PLoS Biol. **2:** E68.