Université des Sciences et Technologies de Lille

Ecole Doctorale Biologie-Santé

THESE DE DOCTORAT

Aspects cellulaires et moléculaires de la biologie

Génétique Microbienne

Présentée par

Derifa KADOUCHE

The comparative biochemistry of storage polysaccharide metabolism in Chlamydiales and Cyanobacteria: insights into the evolution of glycogen and starch metabolism in Eukaryotes

Directeur de Thèse : Pr Steven BALL

Equiipe de Génétique Microbienne UGSF UMR 8576

soutenue le 23 Novembre 2016

Les membres du jury:

Rapporteurs:
- Andreas WEBER, Professeur de l'université de Dusseldorf
- Wolfgang LÖEFFELHARDT, Professeur de l'Université de Vienne

Examinateurs:
- Agathe SUBTIL, Directrice de Recherche Institut Pasteur Paris
- Gilbert GREUB, Professeur au CHUV Lausanne
- Christophe COLLEONI, Professeur de l'Université Lille1
- Steven BALL, Professeur de l'Université Lille1

Président:
- Christophe D'HULST, Professeur de l'Université Lille1

Acknowledgments

First of all I would like to thank all members of my thesis comittee:

Professor Andreas WEBER, Professor Wolfgang LÖEFFELHARDT and Professor Gilbert GREUB for accepting to be members of my Ph. D jury and to examine my work.

Doctor Agathe SUBTIL for giving me both advice and coaching during my Ph. D and during my stay a few days at the Institute Pasteur Paris.

Professor Christophe D'HULST the comittee's chair for all his help during my Ph.D.

Je tiens particulièrement à remercier mes chefs le Professeur Steven Ball et le Professeur Christophe Colleoni, d'abord pour m'avoir acceptée dans l'équipe depuis mon master puis de m'avoir gardée pendant ma thèse et même après ma thèse pour un poste d'ATER.Tous les mots du monde ne suffisent pas pour exprimer ma reconnaissance et ma gratitude. Je vous remercie pour votre patience, votre bienveillance. Vous m'avez beaucoup appris et je me suis toujours senti soutenue et plus particulièrement dans les moments difficiles au cours de cette thèse. Je tiens à vous adresser mes sincères remerciements et ma gratitude.

Je tiens à remercier tous les membres de l'UGSF avec qui j'ai pu échanger scientifiquement ou personnellement, mais plus particulièrement l'équipe de génétique microbienne :

Ugo Cenci j'espère que t'es fier de ta petite stagiaire qui est arrivée en stage de Master 1 perdu et stressée, tu m'as beaucoup appris au cours de mon parcours, UN ENORME MERCI à toi.

Maria-Cecilia Arias je te remercie pour ton soutien et tes discussions entre deux manips ainsi que tes encouragements.

Cathy je te remercie pour ton soutien, tes discussions intéressantes et biensurs pour tes compagnes de mutagénèse qui nous ont permis à tous de faire notre travail ☺

Binquan Huang thanks for your help during the year that you spent with us in Lille.

Mathieu Ducatez merci pour ton aide durant mes années de thèse.

Sylvain Laurent oui oui, tu fais toujours partie de l'équipe, tu n'assumes toujours pas le fait d'être parti, je tiens à remercier l'ami avant le collègue, merci pour ton soutien, merci de m'avoir supporté mes plaintes, surtout en fin de thèse.

Justin Findinier je te remercie pour ton soutien, ta bonne humeur et merci de m'avoir supporté durant mes deux premières années de thèse au labo et même après être parti du labo.

Merci à tous mes amis de l'UGSF et autre : Steffi, Eudoxie, Maxime, Maud, Adeline, Hande, Thomas, Malika, Sara, Olivia, Ines, Imene, Soraya, Kenza ..… et la liste est longue, plus spécialement Cindy ta présence était d'une grande aide, nos pauses café nous permettaient de nous détendre un peu et avoir un peu plus d'énergie pour finir les journées de manips, alors merci.

Radia quoi dire à part you're the best BFF et ça sera forever binôme (je t'aime ;)).

Ma sœurette Ghenima je ne sais pas quoi te dire il n y'a pas assez de pages pour exprimer l'amour que j'ai pour toi et la chance que j'ai de t'avoir rencontré et de t'avoir dans ma vie. Si je suis arrivée jusque là, c'est en grande partie grâce à toi. D'une façon ou d'une autre tu as toujours été là quand j'avais besoin de toi et même quand je n'avais pas besoin de toi (hamlayghk).

Je remercie toute ma famille : frères et sœurs et même nièces et neveux pour leur soutien je n'y serai pas arrivé sans vous, je suis reconnaissante de vous avoir dans ma vie.

Le meilleur pour la fin ; mes parents, grâce à vous et vos encouragements, j'ai réussi à aller jusqu'au bout de mes rêves. Je ne serais pas où je suis si vous n'aviez pas été là, vous étiez loin de mes yeux mais jamais loin de mon cœur, malgré la distance vous m'avez toujours soutenu. Je suis reconnaissante de vous avoir comme parents, j'espère que vous êtes fier de la petite dernière autant que je suis fière de vous avoir comme parents.

ABBREVIATIONS

Gly: Glycogen

AP: Amylopectin

BLD: Beta Limit Dextrin

GlgA: Starch/glycogen synthase

GlgB: Branching Enzyme

GlgP: Starch/glycogen Phosphorylase

GlgX: Debranching Enzyme

iDBE: indirect Debranching Enzyme

MalQ: Amylomaltase

ADP: Adenosine Diphosphate

ADP-Glucose: Adenosine diphosphoglucose

UDP-glucose: Uridine diphosphoglucose

AMP: Adenosine monophosphate

ATP: Adenosyl tri phosphate

G1P: Glucose-1-Phosphate

DP: Degree of polymerization

DTT: 1,4-dithiothréitol

DNA: Deoxyribonucleic Acid

OD: Optic Density

IPTG: isopropyl β-D-1-thiogalactopyranoside

Amp: Ampicillin

Kan: Kanamycin

LB: Luria-Bertani

PCR: Polymerase Chain Reaction

WSP: Water soluble polysaccharide

Résumé

Le glycogène et l'amidon sont les formes de polysaccharides de réserve les plus répandues. Ils sont tous deux constitués de résidus de glucose liés en α-1,4et branchés en α-1,6. Le glycogène est présent sous forme de particules hydrosoluble dans le cytosol des cellules animales, fongiques et bactériennes. L'amidon, quant à lui, est retrouvé chez un petit groupe de cyanobactéries marines unicellulaires fixatrices d'azote appartenant à l'ordre de Chroococcales, et chez les eucaryotes photosynthétiques (plantes, algues rouges et Glaucophytes), et leurs dérivés après endosymbiose secondaire (alvéolés, cryptophytes). Bien qu'ils se distinguent fortement par leurs propriétés physico-chimiques, l'amidon a évolué à partir d'un métabolisme du glycogène préexistant. Il est apparu après l'endosymbiose primaire du plaste qui a eu lieu il y a plus d'un milliard d'années entre une cellule eucaryote et une cyanobactérie. Il a été proposé que l'endosymbiose primaire du plaste ait impliqué la manipulation du métabolisme des polysaccharides de réserves par une bactérie intracellulaire obligatoire pathogène qui appartient à l'ordre des Chlamydiales.

Afin de comprendre l'histoire évolutive des gènes du métabolisme du glycogène, et enquêter sur l'implication d'une bactérie de l'ordre des Chlamydiales dans l'endosymbiose plastidiale, nous nous sommes intéressés à l'évolution des polysaccharides chez les cyanobactéries et les Chlamydiales. Nous avons donc disséqué le fonctionnement de ce métabolisme chez *Cyanobacterium* sp CLg1, une cyanobactérie de l'ordre des Chroococcales qui accumule simultanément de l'amidon et du glycogène. Un mutant a été sélectionné après mutagenèse UV qui ne contient pas d'amidon, mais maintient la production normale de glycogène. Ce mutant est défectueux pour une des deux activités glycogène/amidon synthase: GlgA2. Nous avons démontré que la synthèse de l'amidon dépend d'une fonction fournie par GlgA2 qui ne peut pas être assurée par GlgA1. Notre travail suggère que *Cyanobacterium* sp. CLg1 affiche deux voies séparées pour la synthèse des polysaccharides, offrant la possibilité de réguler les deux voies au moins en partie indépendamment. De plus, nous montrons qu'outre l'action documentée des enzymes débranchantes dans la cristallisation de l'amidon, une propriété spécifique des enzymes d'élongation est aussi requise pour permettre la synthèse d'amidon plutôt que celle du glycogène.

D'autres part, nous avons étudié l'évolution du métabolisme des polysaccharides de réserve chez les Chlamydiales. La plupart des enzymes du métabolisme du glycogène des Chlamydiales ont été caractérisées (GlgC, GlgA, GlgP et GlgX). Nous avons également vérifié la sécrétion de toutes les enzymes de métabolisme du glycogène (GlgC, GlgA, GlgB, GlgP, GlgX et MalQ) par le Système de Sécrétion de Type Trois T3SS dans un système hétérologue de *Shigella flexenerii*. Des analyses phylogénétiques ont été effectués pour toutes les enzymes de métabolisme du glycogène des Chlamydiales. A l'exception de GlgC chez *Chlamydia trachomatis*, les enzymes de métabolisme du glycogène des Chlamydiales se sont avérée sécrétées par le T3SS. Les analyses phylogénétiques des différentes enzymes ont montré une origine monophylétique pour cinq d'entre elles (GlgC, GlgA, GlgB, GlgX et MalQ), tandis que GlgP s'est avéré être polyphylétique. Les résultats de la caractérisation de GlgC, GlgA, GlgX et GlgP renforcent notre théorie de ménage à trois et l'implication des Chlamydiales dans l'établissement de l'endosymbiose primaire du plaste. L'impact du métabolisme du glycogène des Chlamydiales sur l'évolution de celui des plantes et des animaux est discuté.

Mots clés : Glycogène, amidon, Chlamydiales, Cyanobactérie, Endosymbiose, évolution, phylogénie.

Abstract

Glycogen and starch are the most commonly found forms of storage polysaccharides. They both consist solely of glucose residues linked and branched respectively through α-1,4 and α-1,6 glycosidic bonds. Glycogen is present in the form of water soluble particles in the cytosol of animal fungal and bacterial cells, The distribution of starch, meanwhile, is restricted to a small group of Nitrogen fixing single-cell marine cyanobacteria belonging to the order of Chroococcales, and to photosynthetic eukaryotes such as the Archaeplastida (plants, red algae and Glaucophytes) and some of their secondary endosymbiosis derivatives (Alveolates, Cryptophyta). Although they considerably differ in their physicochemical properties, starch evolved in only a few steps in eukaryotes from the pre-existing eukaryotic glycogen metabolism. It appeared, after primary plastidial endosymbiosis which took place over one billion years ago between an ancestral cyanobacterium and a heterotrophic eukaryotic host. This endosymbiosis has been recently proposed in the "Ménage à Trois Hypothesis" (MATH) to have been triggered and facilitated through manipulation of glycogen metabolism by obligate intracellular bacteria pathogens, belonging to the order Chlamydiales.

In order to understand the evolutionary history of glycogen metabolism genes, and to investigate the possible nature of the Chlamydiales involvement in primary plastid endosymbiosis, we probed the evolution of storage polysaccharide metabolism in both extant Chroococcales and Chlamydiales. In the recently axenized, *Cyanobacterium* sp CLg1 that accumulates both glycogen and starch, a starchless mutant was selected after UV mutagenesis. This mutant was selectively impaired in starch synthesis, although it maintained the normal production of glycogen. The mutant was found to be defective for one the two glycogen/starch synthases: GlgA2. We thus showed that starch synthesis depends on a function provided by GlgA2 that cannot be supplied by GlgA1. Our work further suggests that *Cyanobacterium* sp. CLg1 displays two separate pathways for polysaccharide synthesis, affording the possibility to regulate both of these pathways at least partly independently. In addition to the known implication of debranching enzymes in starch granule aggregation, we now report that specific intrinsic properties of the elongation enzyme are also required to synthesize starch rather than glycogen.

On the other hand, we investigated the evolution of storage polysaccharide metabolism in Chlamydiales. Several recombinant enzymes of glycogen metabolism were thus characterized (GlgC, GlgA, GlgP and GlgX). In addition, we investigated the secretion of all the glycogen metabolism enzymes (GlgC, GlgA, GlgB, GlgP, GlgX and MalQ) by the Type Three Secretion System T3SS in a heterologous system of *Shigella flexenerii*. Phylogenetic analysis was carried out for all the glycogen metabolism enzymes of Chlamydiales. Most of the glycogen metabolism enzymes except for glgC in *Chlamydia trachomatis* were found to be secreted by the T3SS. The phylogenetic analysis of the different enzymes showed a common monophyletic origin for five of them (GlgC, GlgA, GlgB, GlgX and MalQ), while GlgP was shown to be polyphyletic. Our characterization of GlgC, GlgA, GlgX and GlgP strengthens the MATH and the implication of the Chlamydiales in the establishement of primary plastidial endosymbiosis. The importance of our findings with respect to the evolution of glycogen metabolism in animals and plants is discussed.

Key Words:

Glycogen, starch, Chlamydiales, Cyanobacteria, Endosymbiosis, evolution, phylogeny.

SOMMAIRE – TABLE OF CONTENTS

Acknowledgments
Abbreviations
Résumé
Abstract
Résumé Français
1. Introduction
1.1 Introduction générale 10
1.2. Résumé bibliographique : la biochimie comparée du métabolisme des polysaccharides de réserves chez les eucaryotes et les bactéries 12
2. Résultats
2.1. L'étude du métabolisme du glycogène chez les Chlamydiales met à jour un impact prédominant de ces pathogènes intracellulaires dans le métabolisme des polysaccharides de réserves 16
2.2. Evolution convergente du métabolisme de l'amidon chez les cyanobactéries et les Archaeplastida : l'amidon synthase GlgA2 définit de nouvelles conditions pour l'agrégation sélective de l'amidon relativement au glycogène 21
3. Conclusion générale 24

Introduction
The comparative biochemistry of glycogen metabolism in eukaryotes and bacteria 26

PART I: Investigating storage polysaccharide metabolism evolution in Chlamydiales

Foreword 31

I. Introduction 31

II. Materials and Methods 34

Strains and culture conditions 34
Gene cloning and sequencing 35
Protein expressions in *Escherichia coli* 36
Zymogram analysis 36
Purification of the recombinant proteins 37
Biochemical characterization of the glycogen debranching activities 37
Heterologous secretion assay in *Shigella flexneri* 37
Western blot analysis of the secreted proteins 38
Complementation of the Arabidopsis thaliana isoamylase mutants 38

III. Results:

Storage polysaccharide metabolism gene content 39
Induction of glycogen synthesis and sites of storage polysaccharide accumulation in infected *Acanthamoeba castellanii* 41
Phylogenetic analysis of Chlamydiales glycogen metabolism genes 42
Investigating the effector nature of glycogen metabolism enzymes in a semi *in vitro* system from *Shigella flexenerii* 49
Characterization of the GlgC and GlgA recombinant proteins from Chlamydiales 51
Characterization of the GlgP and GlgX recombinant proteins of Chlamydiales 56
The Protochlamydia amoebophila isoamylase restores leaf starch synthesis in glycogen accumulating mutants of Arabidopsis thaliana 63

IV. Discussion

The prokaryotic ancestry of the animal glycogen phosphorylases may explain its sensitivity to AMP allosteric activation 65
Lateral gene transfer of GH77 α-1,4 glucanotransferase from Spirochetes may explain the evolution of cytosolic α-amylase in the eukaryotic domain 66
Lateral gene transfer of chlamydial GH13 isoamylase may explain the evolution of starch in the cytosol of the last common ancestor of Archaeplastida 68
Biochemical properties of recombinant GlgA of Parachlamydiaceae and Simkaniaceae supports the Ménage à Trois hypothesis 70
A novel pathway of glycogen synthesis operates in Waddliaceae and Criblamydiaceae 71

References 72
Supplementary figures 75

PART II: Investigating storage polysaccharide metabolism evolution in Cyanobacteria

Foreword 79

Results: Publications.
Characterization of function of the GlgA2 glycogen/starch synthase in *Cyanobacterium* sp. Clg1 highlights convergent evolution of glycogen metabolism into starch granule aggregation 80

General conclusions 120

References 122

RESUME FRANÇAIS

1. Introduction

1.1. Introduction générale

L'acquisition du trafic membranaire et de l'endocytose par les ancêtres archéens de la lignée eucaryote [1] a rendu possible l'internalisation accidentelle dans celles-ci de bactéries entières. Ces ancêtres du domaine eucaryote [2] ont donc très tôt été confrontés à la colonisation pathologique de leurs cytosols par des bactéries libres. Les lignées eucaryotes en émergence ont, de fait, rétorqué en évoluant des mécanismes de défense antibacteriens, tels que la synthèse des ROS et de peptides antimicrobiens, ou encore l'évolution de l'autophagie ou de l'acidification des vacuoles [3]. Ces défenses antibactériennes ont permis la sélection du mode de vie phagotrophe chez les eucaryotes en émergence. Nous pensons que c'est ce mode de prédation original, régulant la taille des populations procaryotes, qui a précipité la diversification du domaine eucaryote. Dans cette optique, la phagocytose serait responsable du succès écologique initial de ce domaine. A leur tour, les bactéries ont évolué des mécanismes leur permettant de résister à ces défenses, par l'entremise de la sécrétion d'effecteurs protéiques dans le cytosol de l'hôte infecté. Certaines de ces lignées bactériennes se sont ensuite spécialisées, jusqu'à devenir des pathogènes intracellulaires obligatoires, incapables de se répliquer à l'extérieur de leur hôte. Les plus célèbres d'entre eux aujourd'hui définissent les rickettsiales et les chlamydiales dont les ancêtres remontraient à l'émergence même des eucaryotes. Ces dernières ont évolué des effecteurs protéiques, leur permettant non plus seulement de résister aux mécanismes eucaryotes de défense antibactériens, mais aussi de se multiplier de manière tempérée ou virulente dans leur hôte, en fonction de l'état physiologique de ces derniers. La phagotrophie requiert toutefois la présence d'une taille et d'une mobilité accrues de la cellule, ce qui, pour des raisons énergétiques, a nécessité l'acquisition préalable de la phosphorylation oxydative par des ancêtres archéens notoirement anaérobies. La complexité et la nécessaire compartimentation de ce métabolisme ont interdit l'acquisition des gènes nécessaires par simples transferts latéraux multiples de sources bactériennes diverses. La transition de la pathogénécité à la symbiose, d'un ancêtre tempéré des Rickettsiales, contenant les complexes protéiques et les structures membranaires requises, explique particulièrement bien l'endosymbiose de la future mitochondrie de même que le signal phylogénomique émanant des Rickettsiales dans le génome mitochondrial et nucléaire eucaryote [4]. En effet, ce mécanisme autorise le transfert en bloc de la machinerie et de l'architecture nécessaire pour la phosphorylation oxydative. Mon laboratoire d'accueil a récemment proposé que les endosymbioses impliquant un partenaire bactérien internalisé et

un hôte eucaryote ne sont possibles que si la bactérie impliquée est préadaptée à la vie intracellulaire ; en d'autres termes si la bactérie dérive d'un clade pathogène intracellulaire obligatoire ou facultatif. Les endosymbioses sont très communes dans la nature et appartiennent pour la plupart à des clades reconnus comme dérivant de bactéries intracellulaires facultatives ou obligatoires. Toutefois celles qui aboutissent à un nouvel organite cellulaire totalement intégré sont excessivement rares. En effet seuls la mitochondrie et le plaste (chloroplaste) ont atteint ce niveau d'intégration métabolique. L'endosymbiose primaire du plaste définit l'évènement d'acquisition de la photosynthèse oxygénique par les eucaryotes. Elle a impliqué un ancêtre eucaryote phagotrophe et une cyanobactérie pratiquant la photosynthèse. Les cyanobactéries ne sont nullement reconnues comme des pathogènes intracellulaires obligatoires ou facultatifs mais plutôt comme des producteurs primaires, archétypes des bactéries libres, incapables de se répliquer, ou même de survivre dans le cytosol des eucaryotes après phagocytose. Comment donc expliquer que de tels organismes aient pu être stabilisés et évoluer de manière analogue à la mitochondrie en un véritable organite cellulaire. Une partie de la réponse a été apportée par la découverte par 5 équipes différentes d'un signal phylogenomique chlamydien dans les génomes des eucaryotes descendant de l'endosymbiose primaire du plaste [5, pour revue voir 6]. Or, ni les plantes terrestres ni les algues rouges et vertes ne sont reconnues comme sujettes à infection par ces pathogènes intracellulaires obligatoires. Malgré cela, une cinquantaine de gènes d'origine chlamydienne, dont la moitié est partagée par au moins deux des trois lignées issues de l'endosymbiose primaire du plaste, sont sélectivement retrouvés chez les eucaryotes photosynthétiques [7]. Les chercheurs ont donc logiquement proposé que des pathogènes intracellulaires obligatoires ancêtres des chlamydiales infectaient de manière routinière l'hôte eucaryote phagotrophe qui a internalisé la cyanobactérie, et auraient assisté activement l'établissement de la cyanobactérie [7,8]. Le laboratoire de génétique microbienne de l'UGSF dans lequel j'ai effectué ma thèse, a pu préciser la nature moléculaire des effecteurs cytosoliques et des transporteurs chlamydiens supposés impliqués, proposant et expliquant le rôle des chlamydiales dans une symbiose tripartite appelée « ménage à trois », qui a permis l'assimilation du carbone d'origine photosynthétique dans le cytosol eucaryote [7, 10]. Les effecteurs chlamydiens cytosoliques soupçonnés impliqués sont deux enzymes du métabolisme du glycogène : la glycogène synthase (GlgA) et l'enzyme débranchante (GlgX). Les gènes codant ces protéines sont également retrouvés dans les cyanobactéries accumulatrices d'amidon ou de glycogène. Les voies métaboliques proposées sont suffisamment détaillées pour en éprouver la réalité par voie expérimentale. Ce laboratoire a

comme but, à moyen terme, de réaliser une endosymbiose expérimentale contrôlée, impliquant d'une part, une amibe coloniale modèle phagotrophe (*Dictyostelium discoideum*) et une cyanobactérie modèle (*Synechocystis* sp. PCC 6803). Dans ce contexte très général, le but que j'ai poursuivi dans ma thèse est double : 1°) J'ai voulu préciser les mécanismes moléculaires régissant la biosynthèse de l'amidon chez la cyanobactérie modèle *Cyanobacterium* sp CLg1 2°) J'ai voulu tenter de comprendre comment les Chlamydiales en général manipulent le métabolisme carboné de leur hôte en étudiant les propriétés biochimiques des enzymes recombinantes correspondant aux effecteurs du métabolisme du glycogène.

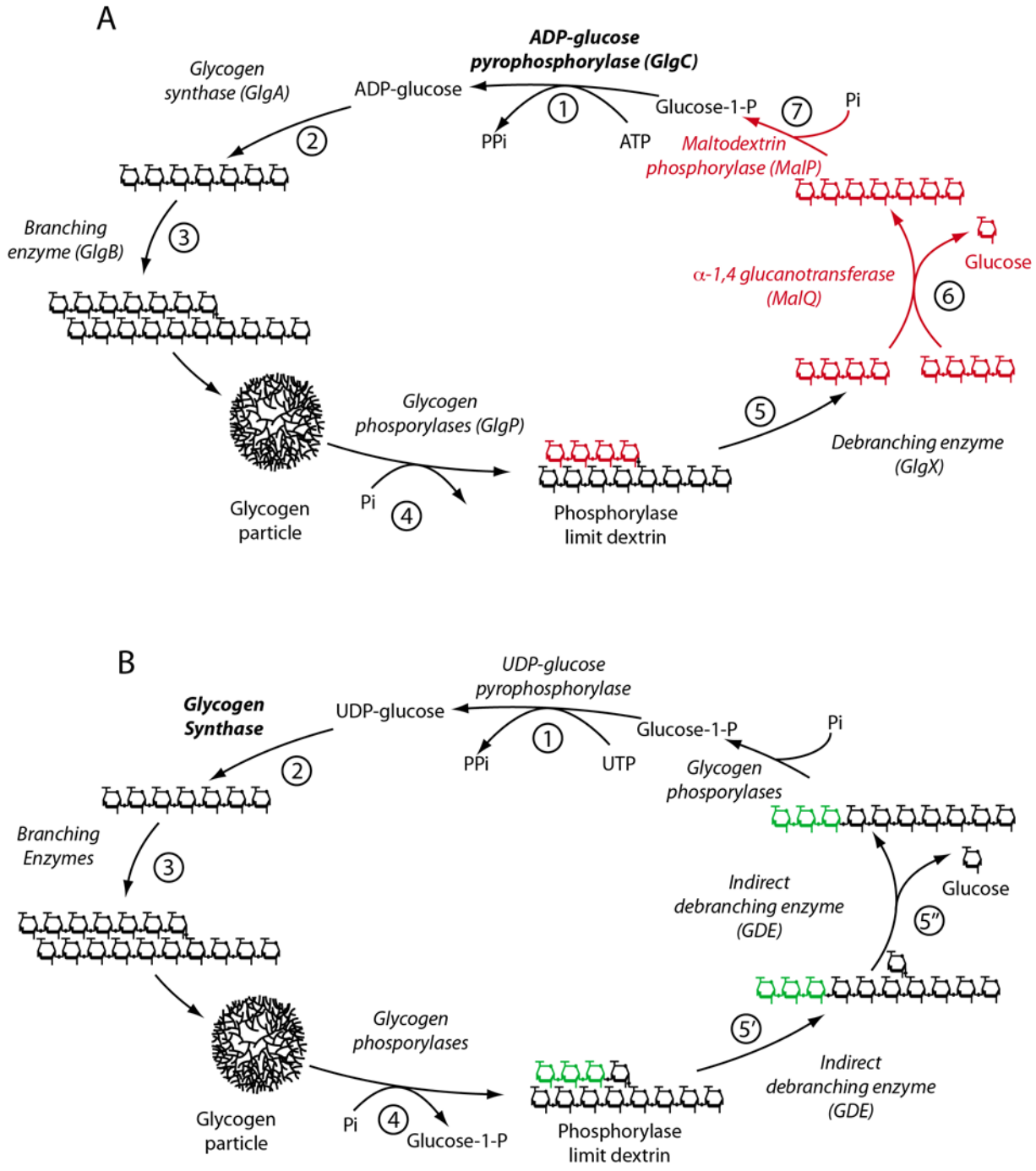
Avant de résumer mes résultats expérimentaux sur ces deux aspects je passerai en revue de manière très succincte la biochimie comparée de la structure et de la synthèse des polymères de réserves chez les bactéries et les eucaryotes et je résumerai ensuite mes résultats obtenus sur les cyanobactéries d'une part et les chlamydiales d'autre part.

1.2. Résumé bibliographique : la biochimie comparée du métabolisme des polysaccharides de réserves chez les eucaryotes et les bactéries

Figure R1 Comparaison des métabolismes du glycogène chez les bactéries et les eucaryotes. (voir le texte ci-dessous pour des explications plus détaillées)
Les chaines du glycogène sont représentées en noir alors que les chaines libérées lors du catabolisme sont représentées en rouge les enzymes considérées comme faisant partie du métabolisme du glycogène, au sens strict, sont schématisées en noir alors que celles qui font également et surtout partie du métabolisme des malto-oligosaccharides sont illustrées en rouge. Les segments de chaines transférées par l'enzyme de débranchement indirect sont représentés en vert. Les étapes « limitantes » qui contrôlent les flux dans ces voies sont illustrées en gras.

(A) Métabolisme du glycogène chez les bactéries

(B) Métabolisme du glycogène chez les eucaryotes.



Les eucaryotes et les eubactéries accumulent le plus fréquemment des α-1,4 glucanes ramifiés en α-1,6. La disposition symétrique de la ramification α-1,6, à raison de deux chaines en moyenne par branche chez le polysaccharide accumulé par les eubactéries et les eucaryotes non photosynthétiques conduit à la synthèse de particules de glycogène de dimension limitée, soluble dans l'eau et rapidement mobilisable par les enzymes de dégradation [12].. Cette structure nécessite un nombre limité de fonctions pour sa biosynthèse. En effet l'édification de l'architecture du grain de glycogène requiert seulement une étape d'élongation suivie par

une réaction la ramification. L'élongation consiste à un transfert de glucose à partir d'un glucosyl-nucléotide sur l'extrémité non-réductrice d'un α-1,4 glucane en croissance (Figure R1 panneaux A et B réaction 2). La ramification s'opère par un clivage hydrolytique et un transfert en α-1,6 du segment clivé (Figure R1 panneaux A et B réaction 3). Des différences importantes subsistent entre eubactéries et eucaryotes non photosynthétiques sur le plan notamment de l'identité du glucosylnucléotide utilisé (UDP-glucose chez les eucaryotes non photosynthétiques (Figure R1 panneau B) et ADP-glucose chez les eubactéries (panneau A)) ainsi que sur la nature de l'étape limitante sur laquelle s'exerce la régulation métabolique [13]. Dans le cas des eubactéries la synthèse d'ADP-glucose par l'ADP-glucose pyrophosphorylase (Figure R1 panneau A réaction 1) se réalise par l'entremise de régulations allostériques sur l'enzyme. Chez les mammifères ou la levure, la phosphorylation des glycogènes synthétases et phosphorylases induisent la dégradation ou la synthèse du polysaccharide [12,13]. Le catabolisme du glycogène s'effectue dans tous les cas par la glycogène phosphorylase. Cette enzyme utilise l'orthophosphate et le glucose des chaines externes de la particule de glycogène pour libérer du glucose-1-P [12] (Figure R1 panneaux A et B respectivement réactions 7 et 4) Cette manière de faire permet de récupérer 50 % des ATPs dépensés pour l'activation du glucose lors de la synthèse. La phosphorylase s'arrête systématiquement à 4 résidus de glucose de distance d'une ramification qui ne peut être attaquée par cette enzyme. Ce produit ultime de dégradation appelé phosphorylase limite dextrine (Figure R1 panneaux A et B) doit ensuite être déramifié pour poursuivre la dégradation de la particule. Les bactéries et les eucaryotes se distinguent par le mécanisme de déramification. Les bactéries attaquent directement la liaison α-1,6- par hydrolyse, à l'aide d'une enzyme appartenant à la famille CAZy GH13 [13]. L'enzyme de débranchement directe bactérienne est souvent appelée GlgX du nom du gène codant la protéine chez *E. coli*. Ce mode de débranchement libère un malto-oligosaccharide de 4 résidus glucose de long : le maltotetraose. Le maltotetraose n'est pas inerte sur le plan osmotique et doit être métabolisé par des enzymes du métabolisme des malto-oligosaccharides. Ces derniers sont le plus souvent constitués du couple malto-dextrine phosphorylase (MalP) (GT35) et α-1,4 glucanotransferase (MalQ, ou encore amylomaltase) (GH77). Les GH77 disproportionnent les oligosaccharides en transférant des segments de longueurs variables d'un oligosaccharide donneur à un oligosaccharide receveur conduisant à un allongement net du receveur au dépens du donneur. L'oligosaccharide receveur allongé sera le substrat d'une malto-oligosaccharide phosphorylase (maltodextrin phsophorylase) qui dégradera ce dernier jusqu'à 4 résidus de

distance de son extrémité réductrice. Le relargage du maltotetraose requiert donc un couplage du métabolisme du glycogène bactérien et des maltooligosaccharides.

Chez les eucaryotes tout se passe sur la particule de glycogène et aucun malto-oligosaccharides n'est largué dans le cytosol [12, 13]. Une enzyme bi-fonctionnelle appelée enzyme de débranchement indirecte constituée de deux domaines aux propriétés catalytiques différentes va agir pour produire d'une part du glucose, et d'autre part des chaines externes rallongées, mais toujours liées à la particule de glycogène. Ces dernières redeviennent donc un substrat pour la glycogène phosphorylase et il n'y a pas de malto-dextrine phosphorylase. Deux réactions se produisent consécutivement. Un domaine GH13 (et non GH77) de type α-1,4 glucanotransférase va d'abord couper la liaison α-1,4 précédant la branche qui soutient la chaine de 4 résidus de glucose laissée par la glycogène phosphorylase. La transferase va ensuite transférer le maltotriose (en vert sur Figure R1 panneau B réaction 5') sur une chaine de DP4 voisine dans la particule. La chaine externe de DP7 devient à nouveau un substrat pour la phosphorylase. Le domaine α-glucosidase et plus particulièrement α-1,6- glucosidase (GH133) se met ensuite en fonction sur le glucose unique démasqué à l'ancienne branche, le libérant ainsi dans le cytosol.

Même si les produits de la réaction (glucose-1-P majoritairement et glucose) sont les mêmes pour les bactéries et les eucaryotes, les mécanismes et la nature des enzymes impliquées sont différents. Tout se passe comme si l'inertie osmotique des polysaccharides de réserve était beaucoup plus importante à préserver dans le cytosol eucaryote par comparaison à celui des bactéries.

En résumé nous observerons que les métabolismes du glycogène se ressemblent dans les domaines eucaryotes et bactériens. Ils différent cependant sur 3 plans : 1°) sur celui de la nature du substrat activé (ADP-Glc ou UDP-Glc 2°) celui de la nature de l'étape limitant le flux carboné (synthèse du substrat activé (bactéries) ou élongation (eucaryotes) et les modalités de régulation qui en découlent 3°) sur la présence (bactéries) ou l'absence (eucaryotes) de couplage au métabolisme des malto-oligosaccharides

2. Résultats

2.1. L'étude du métabolisme du glycogène chez les Chlamydiales met à jour un impact prédominant de ces pathogènes intracellulaires dans l'évolution du métabolisme des polysaccharides de réserves.

Voici près de 5 ans mon laboratoire d'accueil s'est intéressé à l'impact des chlamydiales dans l'évolution du métabolisme des polysaccharides de réserves chez les Archaeplastida. Celui-ci avait montré que des gènes codant deux enzymes, GlgA et GlgX des chlamydiales avaient été transférés à l'ancêtre commun des Archaeplastida [7]. Rappelons que GlgX est l'ancêtre de la proteine Isa1, qui permet la cristallisation de l'amidon chez les microalgues et plantes vertes, et probablement aussi chez les deux autres lignées archaeplastides : les algues rouges et glaucophytes. Cette découverte s'inscrit dans un contexte plus large, où un signal phylogénétique chlamydien a été mis en évidence par un nombre toujours croissant d'études (pour revue voir [6]). L'importance du signal oscille, selon les auteurs et l'époque (les études anciennes bien sûr mettant à jour comparativement moins de transferts), entre 30 à 70 gènes qui signent un échange entre Chlamydiales et Archaeplastida [5, 7, 8, 11, 20]. Peter Gogarten, le premier, a proposé que ce signal résulte d'un rôle actif de ces pathogènes dans le processus d'endosymbiose plastidiale [8]. Mon directeur de thèse a proposé que les trois organismes étaient en fait unis par une symbiose tripartite où l'organisme hôte, le pathogène et le cyanobionte (la cyanobactérie endosymbiotique) partageaient le codage d'enzymes clés qui étaient directement responsables de l'assimilation du carbone photosynthétique [7]. GlgA jouerait dans cette perspective un rôle essentiel puisque cette enzyme, utilisant chez les bactéries l'ADP-glucose, serait capable d'utiliser un métabolite spécifiquement bactérien en provenance du plaste naissant et de le polymériser dans le pool de glycogène cytosolique de la cellule hôte, effectuant par cette entremise, la première connexion métabolique à la base du lien symbiotique [7]. Ce mécanisme proposé, et supporté par toutes les données acquises, depuis 2005, relatives à l'évolution de la synthèse de l'amidon, requiert que l'enzyme chlamydienne soit présente dans le cytosol de l'hôte, et non pas uniquement dans celui de la bactérie résidant dans l'inclusion. Dans ce schéma, d'une part GlgA, et aussi GlgX, se devaient d'être des effecteurs secrétés par le pathogène dans le cytosol et d'autre part un transporteur de nucleotide-sucres devait transporter l'ADP-glucose produit dans le stroma du cyanobionte dans le cytosol de l'hôte eucaryote. Ces prédictions ont été vérifiées au cours de trois études expérimentales différentes [7, 1, 21]. Par ailleurs chez *C. trachomatis*, la présence d'un transporteur de nucleotide sucre

sur la membrane d'inclusion s'est avérée requise pour la polymérisation du glycogène au sein même de l'inclusion [21]. L'hypothèse du ménage à trois proposée par mon équipe d'accueil est très détaillée au niveau moléculaire. C'est la confirmation expérimentale de ses prédictions très précises qui en font la force. Au cours de la collaboration, mise en place entre l'équipe de génétique microbienne de l'UGSF et le laboratoire de Debashish Bhattacharya de l'université Rutgers, Ahmed Moustafa a utilisé son pipe-line de phylogénomique pour compter les nombres de transferts de gènes chlamydiens relativement aux Archaeplastida d'une part et relativement aux champignons et animaux d'autre part [7]. Ces comparaisons sont pertinentes dans la mesure où les Archaeplastida, les champignons et les animaux ont été suffisamment explorés par le séquençage de génomes entiers. Les filtres utilisés ont restitué une soixantaine de transferts latéraux (LGT) unissant les Chlamydiales et les Archaeplastida tandis qu'une trentaine et douzaine de LGTs étaient repérés respectivement entre Chlamydiales, animaux et champignons. La douzaine de LGTs repérés au sein du règne fongique constitue le bruit de fonds dans ce type d'études, dans la mesure où comme les archaeplastides, les champignons ne sont pas sujet à des infections aujourd'hui par les chlamydiales, du fait de la présence d'une paroi infranchissable par ces pathogènes. La douzaine de LGTs repérés définissent des transferts très probablement antérieurs à l'évolution de la paroi fongique, c'est-à-dire antérieurs à la diversification des champignons et partagés avec les amibes et autres eucaryotes. Par contre, la trentaine de transferts visualisés dans la lignée animale constitue un signal significatif qui atteste de l'interaction constante des Chlamydiales avec cette lignée. Il n'en reste pas moins que les Archaeplastida ont accumulé plus de deux fois ce nombre sur un temps beaucoup plus court, dans la mesure où très rapidement la synthèse de parois végétatives a empêché l'interaction des Archaeplastida avec les Chlamydiales qui n'infectent plus ces lignées aujourd'hui. Dans cette deuxième partie de ma thèse, j'ai participé à un projet de grande envergure visant à préciser l'impact des Chlamydiales sur l'évolution des métabolismes des polysaccharides de réserves des eucaryotes en général et des plantes et animaux en particulier. Le métabolisme du glycogène implique 6 gènes chez les Chlamydiales et travailler sur cet échantillon restreint nous paraissait d'autant plus justifié que nous maitrisons parfaitement le fonctionnement de ces enzymes, ce qui nous permet d'avoir un regard fonctionnel sur les LGTs mis en évidence. L'équipe de génétique microbienne a donc choisi d'exprimer chez *E. coli* les 6 enzymes (GlgA, GlgX, MalQ, GlgP, GlgC, GlgB correspondant respectivement à la glycogène synthase, l'enzyme de débranchement, l'amylomaltase, la glycogène phosphorylase, l'ADP-glucose pyrophosphorylase, et l'enzyme de branchement). Ces enzymes ont été exprimées à partir de 6 espèces modèles de bactéries

illustrant la biodiversité de l'ordre des Chlamydiales [19] à savoir *Simkania negevensis*, (Simkaniaceae), *Chlamydia trachomatis*, (Chlamydiaceae), *Protochlamydia amoebophila*, (Parachlamydiaceae), *Parachlamydia acanthamoeba*, (Parachlamydiaceae), *Waddlia chondrophila*, (Waddliaceae), *Estrella lausanensis*, (Criblamydiaceae). Les clonages préliminaires et les infections préalables aux observations de microscopie électronique, ont été réalisés par Mathieu Ducatez et Sébastien Aebi, au sein du laboratoire de Gilbert Greub au CHU de Lausanne, les analyses bioinformatiques et phylogéniques ont été menées par Maria-Cecilia Arias,Trestan Pillonel (CHU Lausanne) et Ugo Cenci, la caractérisation préliminaire de GlgA a été menée par Mathieu Ducatez et Christophe Colleoni. J'ai pour ma part mené les clonages d'expression et la caractérisation complète et définitive de GlgX de même que le test de sécrétion des protéines recombinantes dans le système semi *in vitro* de *Shigella* (en collaboration et sous la direction d'Agathe Subtil). L'expression de la protéine GlgX de *Protochlamydia amoebophila* a été menée par Binquan Huang qui a également effectué des observations préliminaires sur la protéine GlgC. Les analyses bioinformatiques ont mis en évidence la présence d'un réseau complet de gènes du métabolisme du glycogène chez l'ensemble des Chlamydiales à l'exception des Waddliaceae et Criblamydiaceae. Chez ces deux familles, l'ADP-glucose pyrophosphorylase a été perdue tandis que la glycogène synthase était fusionnée à la protéine GlgB d'origine. Cette fusion glycogène synthase-enzyme de branchement est unique chez les bactéries. L'apparition de la fusion corrèle avec la dispartition de GlgC et l'apparition d'un nouveau gène codant une nouvelle enzyme de branchement. Chez *Estrella lausanensis* la fusion ne semble pas active du fait d'un codon stop dont la présence a été confirmée par Mathieu Ducatez. Ces résultats couplés aux observations de microscopie électronique attestant la présence abondante de glycogène au sein des corps élémentaires de *Waddlia chondrophila* et d'*Estrella lausanensis* suggère fortement que MalQ et GlgB suffisent à assurer la synthèse du glycogène qui utiliserait dans ces conditions soit des malto-oligosaccharides cytosoliques importés soit encore les amorces fournies par GlgP qui peut, dans certains environnements, fonctionner dans le sens de la synthèse. Les analyses phylogéniques enracinent GlgC des Chlamydiales auprès de divers représentants des PVC (Planctomycetes Verrucomicrobia Chlamydiales) suggérant une origine ancienne de ce métabolisme avant même que les Chlamydiales adoptent un mode de vie parasitaire. Il en va de même de certaines des multiples séquences de GlgP. Parmi toutes les séquences disponibles pour les 6 gènes seul le gène GlgP semble polyphylétique relativement au groupe des Chlamydiales. Ce fait en soi est remarquable et montre une bonne conservation chez l'ensemble des Chlamydiales donc une importance fonctionnelle de ces protéines. Pour ce qui

est de GlgP, deux groupes de séquences pourraient d'ailleurs être toutes les deux ancestrales. Le fait le plus remarquable dans les analyses phylogéniques constitue, sans aucun, doute la parenté inattendue entre les phosphorylases de la famille des Chlamydiaceae et les animaux et choanoflagellés. L'ensemble des résultats acquis suggère un remplacement du gène eucaryote ancestral par celui des pathogènes. Ce remplacement a eu lieu avant la diversification des animaux, puisque les lignées unicellulaires à la base du règne animal sont aussi concernées. La direction du transfert a été vérifiée en construisant des arbres dépourvus des séquences procaryotes. Si le transfert avait été en direction des pathogènes, les séquences auraient dû se placer à côté de la lignée soeur des animaux c'est à dire des champignons. Cela n'a pas été le cas, confirmant que le gène eucaryote ancestral avait été remplacé par celui des bactéries intracellulaires. La glycogène phosphorylase animale et en particulier humaine définit l'enzyme la plus importante dans l'histoire des sciences biologiques. Son étude a donné 6 prix Nobels obtenus pour la première synthèse *in vitro* d'une macromolécule biologique, la découverte de la première enzyme allostérique, la découverte de la phosphorylation des protéines, des protéines kinases, des protéines phosphatases, de l'AMPc, des cascades de phosphorylation et du mécanisme d'action des hormones. C'est bien cette enzyme que nous rapportons aujourd'hui comme d'origine chlamydienne. L'étude comparative de l'enzyme animale et fongique avait laissé perplexe les chercheurs qui s'y sont intéressés. En effet l'enzyme fongique pas plus que les autres enzymes eucaryotes rapportées ne font preuve de l'interaction allostérique avec l'AMP. Le mécanisme de phosphorylation et son site semblent différents. La phosphorylase kinase n'existe pas chez les champignons. L'ensemble de ces résultats s'expliquent particulièrement bien par le transfert latéral chlamydien. En effet les enzymes procaryotes, contrairement aux eucaryotes, contiennent très souvent des sites allostériques d'interaction pour l'AMP. De même, le remplacement du gène eucaryote ancestral aurait requis l'évolution de nouveaux mécanismes de phosphorylation impliquant de nouvelles kinases.

Les études préliminaires menées sur l'ADP-glucose pyrophosphorylase chlamydienne suggère la présence d'une régulation similaire à celle observée chez les bactéries utilisant simultanément la glycolyse et la voie de Entner-Doudoroff. En effet, l'enzyme des Chlamydiales semble activée par le fructose 1,6 diphosphate, le fructose-6-P et le pyruvate et inhibée par l'AMP. Rappelons que l'enzyme d'*E. coli* n'est activée que par le fructose 1,6 diphosphate [13]. La glycogène synthase de *C. trachomatis* avait été au préalable montrée comme une enzyme bi-fonctionnelle, utilisant avec efficacité l'UDP-glucose dans le cytosol et l'inclusion et l'ADP-glucose dans la bactérie [21]. Les Simkaniaceae et *Protochlamydia*

amoebophila ont par contre une spécificité étroite pour l'ADP-glucose alors que les enzymes de *Parachlamydia acanthamoeba* et *Waddlia chondrophila* utilisent les deux substrats quoiqu'avec une préférence beaucoup plus marquée pour l'ADP-glucose. J'ai de mon côté vérifié que GlgA était bien toujours un effecteur chez l'ensemble des Chlamydiales alors que GlgC l'enzyme de synthèse de l'ADP-glucose ne s'est pas révélé effecteur pour *C. trachomatis* comme l'avait montré initialement Gehre et al., 2016 [21]. Ces résultats suggèrent que la nature effectrice de GlgC est rendue nécessaire par la spécificité de substrat de GlgA. Ce résultat est important, dans la mesure où il conforte l'idée de l'existence d'un métabolisme de l'ADP-Glc dans le cytosol eucaryote qui serait induit par la présence des pathogènes. Or cette propriété est requise dans l'hypothèse MATH (Ménage à Trois Hypothesis) [2, 9].

Je me suis ensuite concentrée sur l'étude détaillée des enzymes débranchantes des chlamydiales. J'ai purifié ces enzymes et testé leurs préférences de substrat vis-à-vis du glycogène de l'amylopectine du pullulan et des dextrines limites produites par la phosphorylase ou la β-amylase. J'ai ensuite engagé des études cinétiques visant à préciser quelles chaines étaient préférées par les différentes enzymes. De toutes ces études, nous pouvons conclure que la protéine GlgX d'*E. coli* produit très rapidement du maltotetraose à partir de chaines externes de 4 résidus de glucose de long et parait peu efficace pour une production fortement ralentie de chaines plus longues. Cette enzyme préfère le glycogène à l'amylopectine. La protéine Isa1 du maïs débranche avec une efficacité comparable les chaines externes courtes et longues et préfère l'amylopectine au glycogène en accord avec sa fonction dans l'agrégation de l'amylopectine. Les protéines GlgX de *Waddlia chondrophila* et *Estrella lausanensis* font preuve de cinétique de débranchement et de préférences analogue à l'enzyme d'*E. coli*. Ce mode de fonctionnement corrèle avec l'abondance de glycogène visualisé à l'intérieur de ces bactéries et présage d'une fonction adaptée à la synthèse de glycogène dans le compartiment bactérien. Par contre, la cinétique de débranchement et de préférences des enzymes de *Chlamydia trachomatis* et de *Protochlamydia amoebophila* sont en tous points identiques à l'enzyme du maïs. *Simkania negevensis* était la seule enzyme au fonctionnement plus ambigu : en apparence une spécificité analogue à la GlgX d'*E. coli* mais avec une meilleure aptitude à débrancher les chaines plus longues. Le mode de débranchement de type « isoamylase » semble plus adapté à la dépolymerisation rapide du glycogène cytosolique. Ceci pourrait corréler avec une fonction effectrice plus importante que celle requise par la synthèse de polysaccharide endogène aux bactéries. La similitude entre les enzymes du maïs et de *Protochlamydia amoebophila* nous a conduit à tester la

complémentation de mutants d'Arabidopsis substituant la synthèse d'amidon par celle de glycogène. Binquan Huang a montré que le phénotype était effectivement complémenté et que la synthèse d'amidon était restaurée dans les feuilles d'arabette.

2.2. Evolution convergente du métabolisme de l'amidon chez les cyanobactéries et les Archaeplastida : l'amidon synthase GlgA2 définit de nouvelles conditions pour l'agrégation sélective de l'amidon relativement au glycogène.

Cyanobacterium sp Clg1 définit une souche de cyanobactérie unicellulaire à l'origine diazotrophe qui aurait perdu ses capacités de fixation d'azote suite à son axénisation. *Cyanobacterium*sp Clg1 se distingue de tous les autres organismes répertoriés, à ce jour, par l'accumulation simultanée, par la souche sauvage, des deux types de polysaccharides de réserve : le glycogène et l'amidon [14]. Des mutagénèses à haut débit ont été entreprises sur une période de 4 années qui ont abouti à la sélection de près de 100 colonies soupçonnée mutantes. Une vingtaine d'entre elles se sont avérées déficientes pour la synthèse de l'amidon mais pas pour celles du glycogène. Dans la plupart des cas, la quantité normale de glycogène accumulée par la souche sauvage de Cyanobacterium augmentait notablement. Dans un seul cas cependant, celui de la souche 187G11 cette quantité stagnait au niveau d'accumulation de la souche sauvage. Par contre, l'amidon dans cette souche particulière avait totalement disparu alors que dans les autres cas, il subsistait une toute petite quantité d'amidon de structure anormale. Ugo Cenci, au sein de l'équipe de génétique microbienne de l'UGSF, avait montré que les mutants sur-accumulateurs de glycogène avaient été touchés dans un des deux gènes codant la protéine GlgX (GlgX2), l'enzyme débranchante bactérienne [14]. Depuis 1996, l'équipe de génétique microbienne a proposé que l'amidon des végétaux requiert, chez ces derniers, la présence d'une enzyme debranchante inhabituelle chez les eucaryotes : l'enzyme de débranchement directe, si typiquement bactérienne [15, 16]. Le rôle de cette DBE encore appelée isoamylase (Isa) dans ce cas était impliquée dans l'épissage de la structure de l'amylopectine au cours de la synthèse. L'enzyme débrancherait les chaines dont les branches seraient espacées et qui, de ce fait, interdiraient l'alignement des chaines voisines en cristaux, conduisant à l'agrégation de la macromolécule au sein d'un grain solide. Cette fonction supposée a été validée chez un grands nombre de végétaux et de microalgues. Les mutants déficients pour ces enzymes dans les feuilles Arabidopsis, et l'albumen des grains de céréales chez le maïs, le riz, l'orge ou chez les microalgues telles que *Chlamydomonas reinhardtii* substituent la synthèse d'amidon par celle de glycogène appelé pour cette raison phytoglycogène [15, 16]. Ugo Cenci a montré que tous les mutants de phénotype équivalents étaient affectés dans le gène codant GlgX2 et dans aucun autre gène du métabolisme des

polysaccharides de réserves présents dans le génome [14]. La présence d'une série allélique déficiente pour GlgX2 provoquant la suraccumulation de glycogène, prouve que cette enzyme joue le même rôle dans la synthèse de l'amidon cyanobactérien que l'isoamylase végétale. L'enzyme de Cyanobacterium appartient d'ailleurs bien à la même sous famille GH13 que l'activité végétale. Nous avons précisé, dans ce résumé, que la protéine GlgX des bactéries avait pour mission de débrancher les chaines laissées après l'action de la glycogène phosphorylase [17]. Ces dernières sont constituées exclusivement par des chaines de DP4 (maltotetraose) et l'enzyme bactérienne ne débranche d'ailleurs efficacement que ce type de chaines. L'isoamylase végétale se distingue de la protéine GlgX classique par une sélectivité de substrat très différente, particulièrement en accord avec sa fonction d'épissage dans la maturation de la pre-amylopectine. L'enzyme GlgX2 de cyanobacterium a été étudiée par Ugo Cenci [14]. Il a pu montrer que l'enzyme de la cyanobactérie avait très précisément les mêmes propriétés cinétiques que l'enzyme végétale. On serait donc tenté de croire que l'enzyme cyanobactérienne aurait évolué en premier par une duplication du gène GlgX et que ce gène aurait ensuite été transféré au cours de l'endosymbiose primaire du plaste à l'ancêtre commun des plantes, conduisant ce dernier à accumuler de l'amidon dans son cytosol. De fait cette explication simple et élégante peut être rejetée avec confiance. En effet, les phylogénies détaillées de GlgX montrent que l'enzyme végétale a été transférée aux végétaux par l'entremise des pathogènes intracellulaires de type Chlamydiales et non par le partenaire cyanobactérien de l'endosymbiose [7]. Les mutants de cyanobacterium suraccumulant du glycogène définissent un cas remarquable de convergence évolutive puisqu'à deux moments différents : lors de l'émergence des cyanobactéries voici plus de deux milliards d'années et lors de l'établissement de l'endosymbiose plastidiale, il y a 1 à 1,5 milliards d'années, la sélection naturelle a recruté deux enzymes appartenant à la même sous famille CAZy, à partir d'origines bactériennes distinctes, pour effectuer le même travail : l'épissage des glucanes. Les deux activités ont ainsi convergé à partir d'une activité de type GlgX vers une spécificité de clivage de type isoamylase. Au cours de mon stage de master 1, du travail expérimental de master 2 et de mes deux premières années de thèse, je me suis attachée à éclaircir l'origine de la déficience portée par le mutant 187G11 et à en trouver l'explication. J'ai travaillé dans un premier temps en aidant Mathieu Ducatez à établir l'origine de la déficience. Celui-ci avait montré que la mutation de 187G11 corrèle avec l'absence d'une activité visualisable sur zymogramme d'activités que nous avons appelée GlgA2. Mathieu Ducatez a montré que la souche 187G11 n'était déficiente pour aucune enzyme de synthèse du glycogène/amidon cyanobactérien et ne renfermait, de plus, aucune autre mutation dans les gènes correspondant.

J'ai, pour ma part, lors de mon master 1 contribué à montrer la position de l'altération génique supportant le gène de structure de l'activité GlgA2. Cette altération conduisait à un déplacement de cadre de lecture à l'extrémité C-terminale de la protéine suivis par la présence d'un arrêt de traduction prématuré. Un petit segment particulièrement conservé des glycogène/amidon synthases, perdu dans le mutant, explique particulièrement bien la baisse d'activité mesurée dans les extrait bruts mutants, ainsi que la disparition de la bande d'activité en zymogramme. Au cours de mon stage de master 2, j'ai étudié les conséquences de la disparition des activités correspondant respectivement à GlgA1 et GlgA2 dans un autre modèle cyanobactérien, *Synechocystis* sp. PCC6803.

Pour ce faire, j'ai cloné des séquences codant ces deux enzymes, ainsi que les enzymes correspondantes de la souche sauvage de *Cyanobacterium* sp Clg1, au cours de mon M2 et de mes premières années de thèse. Enfin, j'ai exprimé dans *E coli* outre ces 4 enzymes sauvages, la protéine altérée du mutant 187G11. J'ai étudié de manière détaillée les propriétés enzymatiques de ces glycogène/amidon synthases et leur capacité à complémenter une mutation inactivant le gène GlgA d'*E. coli*. J'ai pu déduire de ces expériences que GlgA1 tant chez Synechocystis que Cyanobacterium était une enzyme distributive tandis que GLgA2 chez les deux cyanobactéries était d'avantage processive. L'extinction de la fonction GlgA2 provoque une disparition de l'amidon, tandis que celle de Synechocystis conduit à l'accumulation d'un glycogène à chaines plus courtes, confirmant la nature distributive de GlgA1. Par ailleurs toutes les protéines y compris l'enzyme altérée de 187G11 se sont révélées capables de complémenter la mutation *glgA* d'*E. coli* à condition que la source de carbone soit constituée de maltose. Des sources différentes, requérant l'amorçage de la synthèse des polysaccharides, étaient par contre incapables d'engendrer une complémentation. Ce résultat suggère que les enzymes cyanobactériennes sont incapables d'amorcer l'élongation chez *E. coli*. J'ai pu montrer que les cyanobactéries renfermaient un facteur d'amorçage absent chez *E. coli* et que cet amorçage était défectueux chez 187G11. Ce résultat majeur sous-entend, que contrairement à ce qui est publié dans la littérature, la synthèse du glycogène chez les bactéries nécessite, au moins chez certaines espèces, la présence d'un mécanisme d'initiation de l'élongation particulier [13]. Le génome de Cyanobacterium ne code pas de glycogénine et nous sommes à ce jour ignorants de la nature biochimique du facteur d'initiation chez les cyanobactéries si ce n'est sa nature protéique que j'ai démontrée.

Nous pouvons également conclure, du fait du phénotype du mutant 187G11, que GlgA2 est spécifiquement nécessaire à la synthèse d'amidon chez les cyanobactéries et n'impacte pas du tout la synthèse de glycogène chez Cyanobacterium. Il existe donc bien,

chez cet organisme, deux voies de synthèse des polysaccharides de réserves. Le glycogène serait synthétisé de manière prédominante par GlgA1, et GlgA2 n'interviendrait de manière significative que si l'isoamylase GlgX2 est déficiente. L'amidon, quant à lui, ne serait synthétisé que par GlgA2 en présence de GlgX2. Il est donc plausible que ces deux voies soient régulées séparément. Jusqu'à mon travail, seule l'isoamylase semblait constituer le facteur distinguant synthèse d'amidon et de glycogène. La nature de l'enzyme d'élongation semble désormais, ici, aussi importante. Quelle pourrait donc être la propriété concernée ? Les propriétés d'allongement des glucanes diffèrent énormément entre GlgA1 et GlgA2. GlgA1 n'est capable que de faire de très courtes chaines et le produit n'est jamais visualisable par coloration à l'iode tandis que GlgA2 est capable de synthétiser des produits beaucoup plus longs. Nous pensons que les produits de GlgA1 ne sont pas suffisamment longs pour être pris en charge par les enzymes de branchement participant à la ramification de la pre-amylopectine. Le produit engendré par le branchement de ce glycogène n'est pas un substrat efficace pour l'isoamylase et les crystallites d'amylopectine sont incapables de se former. Il est remarquable de noter que sur le plan phylogénique GlgA2 est affilié à l'enzyme GlgA des chlamydiales et à la SSIII/IV des plantes. Toutefois, la SSIII qui joue un rôle analogue à GlgA2 dans la biosynthèse de l'amidon des plantes et algues vertes ne provient certainement pas des cyanobactéries par le truchement de l'endosymbiose plastidiale. En effet, l'origine chlamydienne de cette fonction ne fait plus de doutes pour nous [7]. Encore une fois, comme pour GlgX2 et ISA1/2, GlgA2 et la SSIII ont donc convergés indépendamment, voici respectivement plus de 2,2 milliards d'années et entre 1 à 1,5 milliards d'années pour remplir une fonction analogue dans la synthèse de l'amidon.

3. Conclusion générale

L'ensemble des résultats acquis au cours de ma thèse démontre que l'évolution de la spécificité d'une enzyme de débranchement directe au préalable sélective pour le malotetraose semble avoir conduit à plusieurs reprises la transformation du métabolisme du glycogène en celui de l'amidon. Néanmoins cette évolution ne semble pas définir la seule condition pour une transformation réussie et nous avons montré que chez les cyanobactéries des propriétés d'élongation minimales étaient également requises. L'examen détaillé des propriétés des enzymes du métabolisme du glycogène des Chlamydiales sont en accord avec ces conclusions. De plus cet examen nous a permis de révéler un impact majeur de ces organismes sur l'émergence des métabolismes des polysaccharides de réserve chez les animaux et les plantes.

Introduction

Introduction

The comparative biochemistry of glycogen metabolism in eukaryotes and bacteria:

Glycogen chains consisting of glucose linked by α-1,4 glycosidic linkages are linked through the presence of 8 -12% α-1,6 branches (for review see Buléon et al., 1998). Simplified models of glycogen structure consist of chains carrying mostly 2 branches. Mathematical modeling predicts that this would generate a progressively more crowded structure towards the periphery of the glycogen particle until a maximal theoretical diameter of 42 nm is reached (Melendez-Hevia et al., 1993). The density of the outer chains prevents further growth because there would be no space to accommodate the size of the glycogen metabolism active sites. This corresponds to a particle of 55000 glucose residues where 36% lies in the outer (unbranched) shell. The small size of the glycogen particles and its detailed structure renders the polymer hydrosoluble. It therefore defines a dynamic and osmotically inert form of storage where the outer chains are readily accessible to degradation by the glycogen catabolism enzymes (see below). It is thought that the eukaryote host at the time of plastid endosymbiosis was a glycogen accumulator and that the transition to starch occurred shortly thereafter (reviewed in (Ball et al., 2011). The primary symbiotic link that established the cyanobiont within the host was proposed to consist of the export of ADP-glucose from the cyanobacterium to the host cytosol where it was polymerized into glycogen (for detailed evidence and discussion see Ball et al., 2011). Hence to understand the generation of the symbiotic link that must have occurred immediately for effective selection of the endosymbiosis partnership it is necessary to have a clear understanding of the similarities and differences of storage polysaccharide metabolism that connected the three partners.

The pathways of glycogen metabolism in eukaryotes and Gram-negative bacteria including Cyanobacteria and Chlamydia are clearly related (see panels A and B) (for a review of glycogen metabolism see (Roach, 2002; Wilson et al., 2010). In both cases, an activated form of glucose is generated in the form of a glucosyl-nucleotide used for elongation. This nucleotide-sugar is synthesized from Glucose-1-P and a nucleotide triphosphate (either ATP (bacteria) or UTP (eukaryotes)) through the action of ADP-Glc or UDP-Glc synthetase (also known as ADP or UDP-glucose pyrophosphorylases) (GlgC in *E. coli*) to generate ADP-glc in bacteria and UDP-Glc in eukaryotes (see panels A and B step 1). Because ADP-glc is solely devoted to the synthesis of glycogen in bacteria its biosynthesis through ADP-glucose pyrophosphorylase defines the major regulatory step of bacterial glycogen metabolism (see panel A in bold).

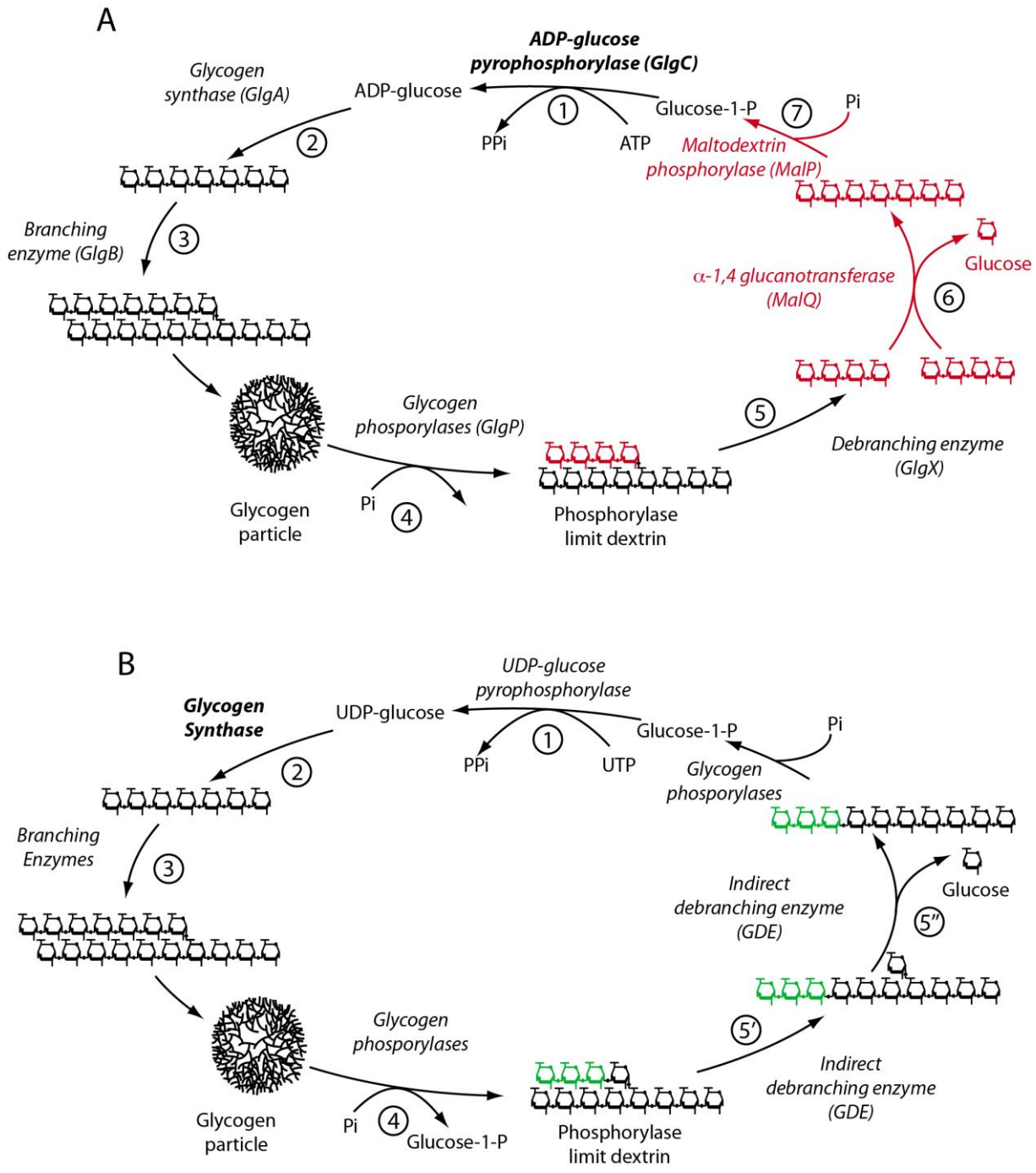


Figure1: Comparison of glycogen metabolism in bacteria and eukaryotes.
The chains present on the glycogen particle are displayed in black whereas those chains that are released by glycogen catabolism in the form of maltooligosaccharides (MOS) are drawn in red. The enzymes considered as *bona fide* glycogen metabolism enzymes are represented in black whereas those considered as part of maltooligosaccharide metabolism are drawn in red. The segments of chains transferred by indirect debranching enzyme are highlighted in green. Major rate controlling steps are highlighted in bold. Panel A: glycogen metabolism in bacteria. Panel B: glycogen metabolism in eukaryotes.

In contrast, the first committed step of glycogen synthesis in eukaryotes is defined by elongation (see panel B in bold) because UDP-glc being is used in many other pathways. Elongation proceeds through glycogen synthases (GlgA in *E. coli*) by transfer of the activated glucose of either ADP-Glc (bacteria) or UDP-glc (eukaryotes) on the non-reducing ends of a previously synthesized chain thereby generating a novel α-1,4 linkage (see panels A and B step 2) . Branching, in all cases, involves branching enzymes (GlgB in *E. coli*) which catalyze the hydrolytic cleavage of previously synthesized α-1,4 linked chain and the transfer of a cleaved segment in α-1,6 position (see panels A and B step 3) of an acceptor chain. Glycogen degradation in all cases proceeds by glycogen phosphorylases (GlgP in *E. coli*) which in the presence of orthophosphate will release the glucose at the non-reducing end in the form of Glc-1-P and will always stop 4 residues away from an α-1,6 branch (see panels A and B step 4).

The remaining glycogen particle (called phosphorylase limit dextrin) constitutes a highly specific substrate for both direct (bacteria) and indirect (eukaryotes) debranching enzymes. In bacteria, the direct debranching enzymes (GlgX in *E. coli*) hydrolyze the short external branched chains and release maltotetraose from the phosphorylase limit dextrin (see panel A step 5) (Dauvillée et al., 2005; Alonso-Casajus et al., 2006). This maltotetraose will be used both as an acceptor and donor in a series of different transfer reactions catalyzed by amylomaltase or disproportionnating enzyme (MalQ in *E. coli*) resulting among others in the production of maloheptaose and glucose (see panel A step 6) (for review of malto-oligosaccharide metabolism in Bacteria, see (Boos and Shuman, 1998). The maltoheptaose can be further recessed to maltotetraose by the action of maltodextrin phosphorylase (MalP in *E. coli*) (see panel A step 7). In eukaryotes the short external chains of 4 glucose residues left over by glycogen phosphorylase will be attacked by "indirect debranching enzyme". This enzyme with two catalytic sites will first catalyze the transfer of the 3 outer glucose residues to the neighboring glycogen chain generating a longer external chain that can be further recessed by glycogen phosphorylase (see panel B step 5'). The left-over glucose at the branch is released by the second catalytic site of the indirect debranching enzyme (see panel B step 5''). Eukaryotes such as Amoebozoa or parabasalids often contain an additional pathway of glycogen mobilization not present in fungi and animals (Ball et al., 2011) which branches off at the level of step 4 (see panel B step 4). This includes β-amylase and transglucosidase. (Dpe2) β-amylase produces maltose (2 Glc residues linked by one α-1,4 glucosidic linkage) from the outer chains of glycogen and stops 2 to 3 residues away from an α-1,6 branch depending on the presence of an even or uneven number of glucose residues. This leaves a

polysaccharide residue known as β-amylase limit dextrin. Transglucosidase belongs to the same family of α-1,4 glucanotransferases as the bacterial disproportionnating enzymes and amylomaltase (GH77). It will transfer one glucose residue from β-maltose to the outer chains of various polysaccharides (reviewed in Ball et al., 2011). It is expected that indirect DBE will also be able to process all the outer chains generated on glycogen after β-amylase action.

To summarize, eukaryotes cannot metabolize maltooligosaccharides other than maltose nor can they use the bacterial specific glycosyl-nucleotide ADP-glucose. It is believed that eukaryotes regulate glycogen metabolism (e.g., animals and fungi) by activating or inhibiting glycogen synthase and starch phosphorylase through protein kinases and phosphatases. In contrast, bacteria regulate flux by allosteric activation and inhibition of ADP-glucose pyrophosphorylase and tightly couple glycogen to malto-oligosaccharide metabolism. These similarities and differences between glycogen metabolism in Amoebozoa and bacteria were used by the three partners of the primary plastid endosymbiosis to establish the symbiotic carbon flux.

PART I: Investigating storage polysaccharide metabolism evolution in Chlamydiales

PART I: Investigating storage polysaccharide metabolism evolution in Chlamydiales

Foreword:

Results presented in this section were performed by myself, Mathieu Ducatez, Binquan Huang, Ugo Cenci, Maria-Cecilia Arias, Sylvain Laurent (Master 2 internship) Christophe Colleoni. Compiling and writing this section of my thesis resulted from intensive discussions between Christophe Colleoni, Steven Ball and myself. As I am not native English-speaking, Steven Ball considerably edited the first draft of this manuscript

Preliminary characterization of the recombinant proteins GlgC, GlgA, GlgP and GlgX were performed by Binquan Huang, Mathieu Ducatez, Sylvain Laurent/Christophe Colleoni and myself respectively. Ugo Cenci and Maria-Cecilia Ariasperformed the phylogenetic analysis. I carried out both preliminary and final detailed characterization on all GlgX proteinsAll experiments on glycogen metabolism enzymes of Chlamydiales secretion by the heterologous system of *Shigella flexnerii* were done by myself in close collaboration with Agathe Subtil's lab in Paris.

I. INTRODUCTION

Glycogen and starch define the most widely distributed types of storage polysaccharide found in Archaea, Bacteria, and Eukaryotes. Both are composed solely of glucose residues linked by α-1,4 glycosidic linkages with occasional α-1,6 branches. Glycogen can be defined as small hydrosoluble particles of 50 nm maximal diameter, with one third of the glucose remaining readily accessible on the outer chains to the cytosolic hydrosoluble glycogen degradation enzymes. It thereby defines a very dynamic, yet osmotically inert, form of glucose storage (Melendez et al., 1997; Melendez et al., 1999). Starch is a mixture of amylose and amylopectin in semi-crystalline solid granules of unlimited size, which largely escapes degradation by hydrosoluble enzymes (Buleon et al., 1998).

The most common pathway of glycogen/starch metabolism is the so-called glucosyl-nucleotide pathway.In bacteria, the synthesis pathway includes 3 steps consisting of ADP-glucose pyrophosphorylase (GlgC), glycogen synthase (GlgA), and glycogen branching enzyme (GlgB) while the degradation pathway involves glycogen and maltodextrin

phosphorylases (GlgP and eventually MalP), debranching enzyme (GlgX) and amylomaltase (MalQ). Glycogen metabolism differs in bacteria and eukaryotes through 3 aspects (Wilson et al., 2010). First bacteria synthesize glycogen exclusively through ADP-glc a bacterial specific metabolite devoted to glycogen synthesis, while eukaryotes always use UDP-glc. Second, glycogen debranching in bacteria releases maltooligosaccharides (maltotetraose) which requires the coupling of glycogen metabolism to the maltooligosaccharide degradation enzymes MalQ and MalP, while only glucose is released by the eukaryotic enzyme, thus obviating this coupling. Third, ADP-glucose pyrophosphorylase defines the first committed step of glycogen synthesis in bacteria and as such, is subjected to tight allosteric controls by activators and inhibitors. In eukaryotes, because UDP-glc is used in many different pathways, it is the next step (glycogen synthase) which gates the flux of carbon to storage. Hence, this enzyme, together with the eukaryotic phosphorylase, is subjected to many distinct allosteric controls, and is switched on and off through complex phosphorylation-dephosphorylation events, modulating the enzyme sensitivity to allosteric activators and inhibitors.

Starch metabolism has been shown in all cases to have evolved through modifications of a pre-existing glycogen synthesis machinery (Ball et al., 1996). In both cyanobacteria and Archaeplastida, a bacterial type of debranching enzyme named isoamylase has been recruited to trim off chains within glycogen-like structures, which prevent formation of the amylopectin chain clusters and the ensuing polysaccharide crystallization (Mouille et al., 1996; Cenci et al., 2014). Remarkably, this happened at least twice, through recruitment of similar enzymes from different bacterial sources, within either a bacterial glycogen metabolism enzyme network in cyanobacteria or a cytosolic eukaryotic network in Archaeplastida. Isoamylases, while having evolved in both cases from preexisting GlgX-DBEs, do not show their high substrate selectivity and are characterized by novel substrate specificities in agreement with their trimming function.

Research on glycogen metabolism have led to a number of important findings in biological sciences. The first gene analyzed by Gregor Mendel proved to encode starch branching enzyme while the animal glycogen phosphorylase, was used by Carl and Gerty Cori for the first *in vitro* synthesis of a biological polymer (Brown et al., 1961; Illingworth et al., 1961; Bhattacharyya et al., 1990). Moreover, the study of animal glycogen phosphorylase and synthases have led to the discovery of protein phosphorylation, of protein kinases and phosphatases as well as AMPc and signal transduction(Brushia and Walsh, 1999).

Mitochondrial and plastidial endosymbioses have recently been proposed to have required the participation of obligate intracellular bacterial pathogens of the Rickettsiales and

Chlamydiales order (Ball et al., 2016). In both cases, these organisms have been suggested to offer the very high level of preadaptation required for establishment of the symbiont in a hostile intracellular environment.

Mitochondria were directly derived from an ancestral Rickettsiales pathogen, while cyanobacteria were probably sheltered within the inclusion of intracellular Chlamydiales. The Chlamydiales provided protection from antibacterial defenses as well as key transporters and cytosolic effectors that were required to establish symbiosis (Ball et al.2013, Facchineli et al.2013).

Storage polysaccharide metabolism has been shown to be selectively lost from all obligate intracellular bacterial symbionts and pathogens, with the noticeable exception of Chlamydiales (Henrissat et al., 2002). Quite interestingly, Chlamydiales have been recently proven to use the previously thought house-keeping enzymes of glycogen metabolism as important protein effectors, diverting host storage polysaccharide metabolism to their benefit (Gehre et al., 2016). Most importantly, the chlamydial effector glycogen synthase (GlgA) was proposed to define the enzyme at the core of the plastid endosymbiosis symbiotic flux (Ball et al., 2015). Indeed, this ADP-glc specific enzyme was required to polymerize photosynthate exported in the form of ADP-glc into the host glycogen pools (Colleoni et al., 2010). In this respect, it is worth stressing that the only two prokaryotic additions to the otherwise eukaryotic cytosolic starch metabolism network of Archaeplastida are two chlamydial effectors of glycogen metabolism: GlgA and GlgX (glycogen debranching enzyme).

The essential role of intracellular pathogens in both mitochondrial and plastid endosymbiosis has been largely overlooked. This is because these organisms were initially believed to have evolved comparatively recently, as specific pathogens of the animal lineage. In both cases however, pathogens with larger genomes and a much wider host spectrum have now been reported (Bertelli et al., 2010; Bertelli et al., 2015). These display the basic features of the obligate intracellular life style of the animal infecting Rickettsiae or Chlamydiae. The finding of such pathogens pulls back the dating of these organisms over a billion years, possibly back to the time of the first eukaryotes. Because Chlamydiales are the only intracellular pathogens known to manipulate storage polysaccharide metabolism, we now explore how these organisms may have impacted evolution of eukaryotic glycogen metabolism in general. We have thus selected well studied representatives of the major Chlamydiales clades, and have explored their genome content with respect to glycogen metabolism enzymes and their phylogeny, as well as the sites of glycogen synthesis upon infection in *Acanthamoeba castellani* or human cells. In addition, we have tested the effector

or resident nature of all these proteins in the Shigella semi-*in vitro* TTS system (Subtil et al., 2001). Finally, we have examined their detailed biochemical properties, their substrate preferences, and their kinetic behavior. From all these studies, we conclude that the chlamydial glycogen metabolism has had a surprising major impact in the evolution of eukaryotic storage polysaccharide metabolism. We show that the glycogen phosphorylase enzyme gene was donated to the common ancestor of choanoflagellates and animals by the common ancestor of the animal infecting Chlamydiaceae. We further show that the eukaryotic MalQ-Dpe2 α-1,4 glucanotransferase, although closely related to the chlamydial enzymes did not originate from the pathogens but possibly from spirochaetes very early on during eukaryote evolution. This facilitated the evolution of the maltose producing eukaryotic β-amylase. The enzymes substrate preferences and effector nature of Parachlamydiaceae and Simkaniaceae are further demonstrated to be in perfect agreement with their inferred functions as the molecular triggers of the symbiotic carbon flux at the onset of plastid endosymbiosis, thereby considerably strengthening the "Ménage à trois" (MATH) hypothesis. Finally, the biochemical properties of the GlgX effectors of Chlamydiaceae and Parachlamydiaceae are shown to be those of an archaeplastidal starch-producing isoamylase and not of a bacterial GlgX. In addition, we were able to complement the glycogen producing isoamylase defective mutants of Arabidopsis with these chlamydial isoamylases which restored starch production. We therefore propose that the presence of this isoamylase in the cytosol of the common ancestor of Archaeplastida greatly facilitated the switch from glycogen to starch metabolism and discuss how this switch would have restored host rather than chlamydial control of glycogen stores breakdown.

II. Materials and Methods

Strains and culture conditions:

Escherichia coli Rosetta^{TM} strain was purchased from Promega. Complementation experiments were carried out using single knockout ΔglgA, ΔglgC, ΔglgP and ΔglgX mutantsderivative of wild type strain K-12 (BW25113) purchased from the *E.coli* stock center (http://cgsc.biology.yale.edu). The inducible T7 RNA polymerase gene was introduced into knockout mutants using ΔDE3 lysogenization kit (Novagen). Professor Agathe Subtil (Institut Pasteur Paris) has kindly given both IpaB (T3SS+) and MixD (T3SS-) *Shigella flexenerii* strains.

LB (Luria-Bertani) medium supplemented with kanamycin (50µg/ml) or ampicillin (100µg/ml) was used to grow *Escherichia coli* strains at 37°C.

Gene cloning and sequencing:

glgC, glgA, glgB, glgP, glgX, malQ genes (for biochemical characterizations) or the first 90 nucleotides (for the secretion tests) were amplified from genomic DNAs of *Chlamydia trachomatis, Simkania negevensis, Protochlamydia amoebophila, Parachlamydia acanthamoeba, Waddlia chondrophila* and *Estrella lausanensis* using primers designed to clone the PCR products either into pyc19cya vector (for the secretion tests) or into pET15b expression vectors (for the biochemical characterizations). The PCR products were cloned into Pyc19cya or pET15b expression vectors using the T7 ligase (Thermo Scientific), or Gateway Technology with Clonase II (Invitrogen), respectively. The constructions were then transferred into the chemical competent *E. coli*TOP10 Mach1™-TR, and plated on LB agar with ampicillin. Plasmids were purified using the Kit NucleoSpin Plasmid then sequenced by GATC Biotech Company according to Sanger methods.

glgC-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAATACATTAACAATGTT R: AGTCTCTAGAACTAGCCACTTGACGCATATCA
glgC-PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGTCTTTGTTAACCACACCCC R: AGTCTCTAGAGGATGCAACCCGATCTGTG
glgC-CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGCTGGTAGACGGACG R: AGTCTCTAGATCCGCATAAGACAATAACTCCTAC
glgC-SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGTCAATACGAGAAAAAAAC R : AGTCTCTAGAAGGATAGAGACGACTTCCTT
glgA-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGCACATCATTCATATAGC R: AGTCTCTAGATAACTCACGGCAAAGACC
glgA-PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGCATATTATTCATATCGC R: AGTCTCTAGATAGTTCACGAGAGAGTCC
glgA-SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGTATATCGTTCATGTGA R: AGTCTCTAGAAAGCGCTTTCCCTAAA
glgA-CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAAAATTATTCACACAGCTATCG R: AGTCTCTAGATGCTAGTCCGTATAGCGCG
glgA-WC	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGACCGTTTCCGCCATT R: AGTCTCTAGATTTCATTTGCAAATCCATTTCG
glgA-EL	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAAAATATTTAATATTGA R: AGTCTCTAGAGCCTAGATCTTCATTT
glgB-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGATGACTATGCAACAGACTGA R: AGTCTCTAGAAGCATGAGGTTGGTGATGAAC
glgB-PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGAAACGCACACATCT R: AGTCTCTAGATAGTCCTAAAAATTGGTGGG
glgB-SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGCAAAAAATATTGATGAAAG R: AGTCTCTAGATTCAGGCCGCCAAAGA
glgB-CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGATCCTTTTTTCTTA R: AGTCTCTAGAAGAAACAATTCCCAAAAG
glgP-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTTTGGTAACTAGTTTTGAA R: AGTCTCTAGAGCGAATTAATTCTTCAGC
glgP-PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTTTAATATATCCGACACCG

	R: AGTCTCTAGATGGGGGCTAGCAC
glgP-SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGAATTTCAGGCT R: AGTCTCTAGATTCAAGATTTGCTTC
glgP-CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGTATTTCGATCGGACA R: AGTCTCTAGATTGAGGAGTCTGGACTA
glgP-WC	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAACCAGCCGAGCA R: AGTCTCTAGAGAGCCCTCCGCTATAG
glgP-EL	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAATCACGATACCGATCGCA R: AGTCTCTAGACCCGAGGCCCCCGCTATAG
glgX-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAATATACAAATAA R: AGTCTCTAGAATTTTTAGCATAAAGA
glgX-PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGCACGTGAGAGGA R: AGTCTCTAGATAAATGAGGAATTTTTTC
glgX-SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAACCCATCCTTTA R: AGTCTCTAGAATGGTGGGAATAGACAGC
glgX-CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGGAATCTTTGTCTG R: AGTCTCTAGAAGAAAATAGAGAAAAAC
glgX-WC	F: AGTCAAGCTTGTAATAGTTTTGTTTTTGCATTGAAGCAATGC R: AGTCTCTAGACCTGATGACTTCCGGGAT
glgX-EL	F: AGTCAAGCTTGTAATAGTTTTGTTTTT ATGAGCTACAAATTCAAAACT R: AGTCTCTAGA TGAGTAGAGCGTAAAGTTTA
MalQ-CPA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTTTAAATAATAAGTTGTTGC R: AGTCTCTAGAAGAAATTGGAGCTATCG
MalQ -PA	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGACAAACTTTTTTCTGC R: AGTCTCTAGAGGAGTGTAAAGAGAAAATGG
MalQ -SN	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAGTCTTGAGTCTAAGC R: AGTCTCTAGAGGGAATATTGATTCCATGG
MalQ -CT	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAAATAGAGCCTCAAG R: AGTCTCTAGAAAAAATAATTGGGTGTATCG
MalQ -WC	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGAATAATCTAAGATTTTTCCTATC R: AGTCTCTAGAAACGCACACCCCGTGA
MalQ -EL	F: AGTCAAGCTTGTAATAGTTTTGTTTTTATGACACTGCCCAAAGACCTTGC R: AGTCTCTAGAGGAAAAGAGCGGCACGGC

Table: Primers used in PCR to clone the recombinant proteins of Chlamydiales glycogen metabolism in the expression vectors (for the secretion tests).

Protein expressions in*Escherichiacoli*:

His-tagged recombinant proteins (pET15b vectors) are produced in the *E.coli*ΔCAP, ΔglgC, ΔglgA or ΔglgX mutant strains. A 200 mL LB medium supplemented with ampicillin at 37°C is inoculated with an overnight culture. At the middle of exponential growth, IPTG is added to reach 0.5mM final concentration. The cells were harvested by centrifugation at 16000g for 10 min, the pellets were resuspended in Tris-Acetate buffer (Tris/acetate25 mM pH 7.5,10 mM DTT), the cells were lysate by sonication. The crude extracts were purified for further enzymatical characterization.

Zymogram analysis:

The crude extracts of *Escherichia coli* strains transformed with the recombinant proteins GlgX and GlgP were separated by non-denaturing polyacrylamide gel electrophoresis (PAGE) containing 0.6% of Beta limit dextrin (Megazyme) or rabbit glycogen (Sigma

Aldrich) respectivly. After migration the gel was incubated overnight at room temperature in Tris-Acetate Buffer (Tris/acetate25 mM pH7.5, 10 mM DTT). The gel was stained with iodine solution (0.5 g I_2, 10 g KI) to reveal the activities.

Purification of the recombinant proteins:

Recombinant proteins were purified by affinity onto Ni-NTA Agarose (Qiagen). One ml of crude extract was incubated with 1ml of the NI-NTA Agarose for 1hour at 4°C, then washed with the washing buffer (50 mM of $NaH_2PO_4.2H_20$, 300mM of NaCl and 30mM of Imidazole), the proteins were eluted with the elution buffer (50 mM of $NaH_2PO_4.2H_20$, 300mM of NaCl and 250mM of Imidazole). The purified enzymes were used for the characterization.

Biochemical characterization of the glycogen debranching activities:

Purified extracts of glycogen debranching activities were incubated in buffer containing 120 mM Tris, 30 mM imidazole, 30 mM acetic acid, pH 7.5, 0.5% (w/v) of polysaccharide (glycogen, amylopectin, glycogen limit phosphorylase, amylopectin limit phosphorylase and beta limit dextrin) at 30°C for 1hour. The reducing ends release were measured using the Dinitrosalicylic acid (DNS) method, after 10 min at 99°C the OD was measured at 540 nm, a glucose solution (1mg/ml) was used as standard. The released chains were analyzed by FACE (Fluophore Asisted Carbohydrate Electrophoresis) using APTS (Amino PyreneTrisulfonic Acid)as fluophore.

Heterologous secretion assay in *Shigella flexneri*:

The chimiocompetente IpaB and MixD strains of *Shigella flexneri* were transformed with Pyc19cya constructions. 30mL of LB medium were inoculated with 1 mL of an overnight culture and incubated at 37°C for 4 hours. One mL of exponential culture was centrifuged at 13 200 rpm for 5min, the pellet was resuspended in Laemmli buffer for analysis. The cells were harvested by centrifugation at 4000g for 20min, the supernatant was filtered by a Millipore filter of 0.2µm, Trichloroacetic Acid TCA was added to precipitate the proteins, then the supernatant was centrifuged at 10 000rpm for 15min and resuspended in Laemmli buffer. Fused proteins composed of first 30 amino acid of target protein and a reporter protein (cya) are detected in the pellet and the supernatant by western blot analysis.

Western blot analysis of the secreted proteins:

The pellet and the supernatant were loaded onto 12 % SDS-PAGE. After separation the proteins were electrotransferred on PVDF (Polyvinylidene Difluoride) membrane (BioRad). The membrane was first incubated in blocking buffer for 1h (5% of milk powder in Tris Buffer Saline TBS) and then incubated in a fresh blocking buffer containing the primary antibodies raised against adenylate cyclase (mousse), IpaD and CRP (Rabbit) diluted 1:1000, 1:500, 1:500 in blocking buffer respectively (Institute Pasteur Paris). The membrane was washed by the TTBS buffer (Tween-Tris Buffer Saline), then incubated with the secondary antibodies anti-mousse and anti-rabbit diluted at 1:10000 in blocking buffer (GE healthcare). The membrane was revealed by ECL^{TM} prime western blotting reagent kit (GE healthcare).

Complementation of the *Arabidopsis thaliana* isoamylase mutants:

The complementation was performed according the protocol in (Facon et al., 2013). The leaves were incubated in ethanol 80% at 70 ° C and then rinsed and stained with iodine solution.

III. Results:

Storage polysaccharide metabolism gene content

A survey of glycogen metabolism enzymes was carried out in available completed genome sequences of Chlamydiales (table 1). Among this diversity we have selected the following clades for more detailed studies: the major human pathogen *Chlamydia trachomatis* (Chlamydiaceae), the deep branching *Simkania negevensis* (Simkaniaceae), *Protochlamydia amoebophila* and *Parachlamydia acanthamoeba* (Parachlamydiaceae), *Waddlia chondrophila* (Waddliaceae) and *Estrella lausanensis* (Criblamydiaceae). The phylogeny of these taxa within the order Chlamydiales is presented in Figure 1.

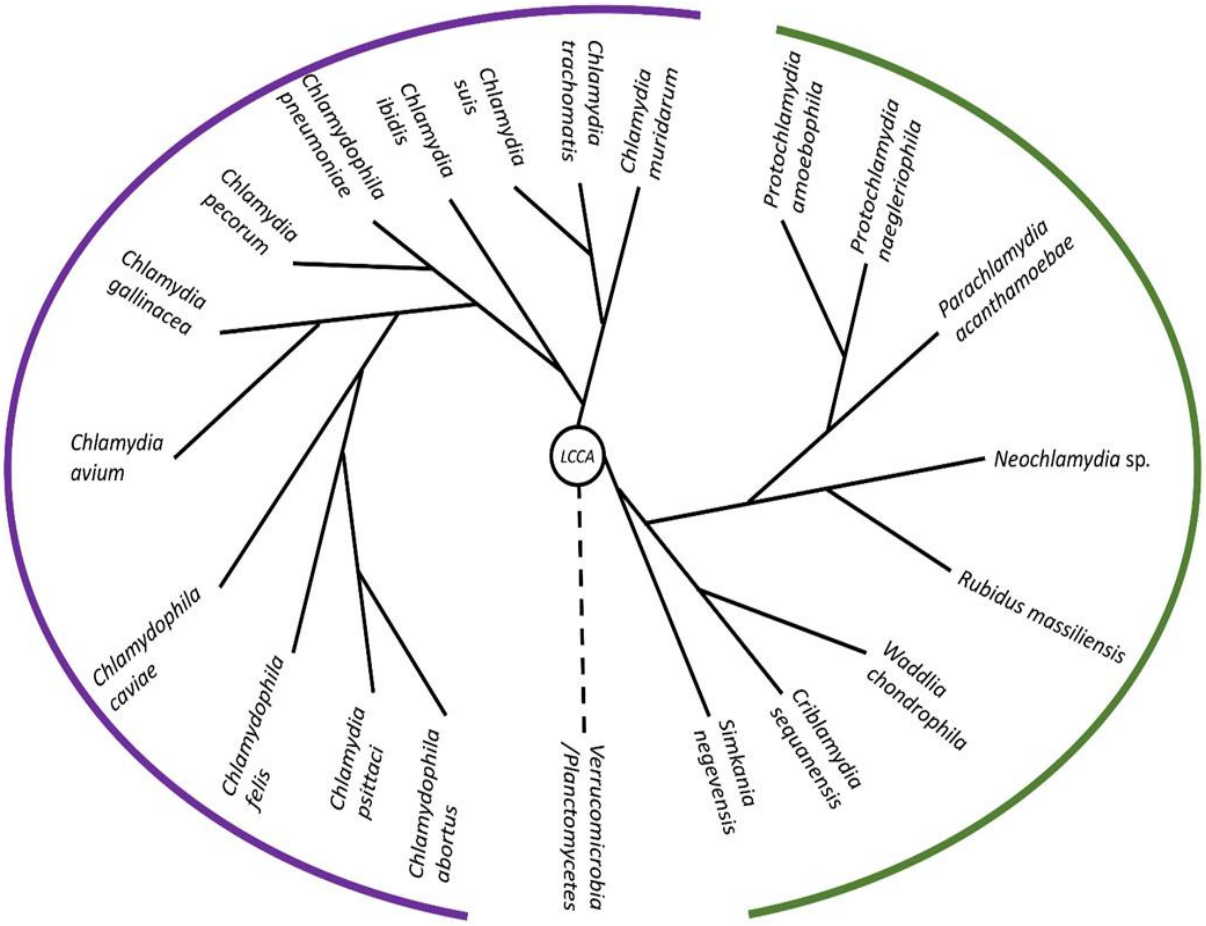


Figure 1: Schematic phylogenetic tree representing the different families of the Chlamydiales order. Chlamydiales are separated in two groups: The Chlamydiaceae family (in purple), and the so-called "environmental"Chlamydiales that include the others families (Criblamydiaceae, Simkaniaceae, Waddliaceae and Parachlamydiaceae) (in green).

All Chlamydiales contain a minimum of 4 (GlgP, MalQ, GlgX, GlgB) out of the 6 enzymes of the classical ADP-glc pathway of bacterial glycogen metabolism. Waddliaceae and Criblamydiaceae have lost the GlgC (ADP-glucose pyrophosphorylase) gene. Consistent with this loss, *Estrella lausanensis* contains no active GlgA gene and therefore apparently does not support glycogen synthesis through ADP-glc. *Waddlia chondrophila* has substituted its GlgA gene by a GlgA-GlgB gene fusion that is apparently unique within bacteria. Concomitant with evolution of the fused gene both Waddliaceae and Criblamydiaceae have acquired a novel functional GlgB gene. Hence the GlgB domain within the fusion is probably not active for the synthesis of glycogen within the pathogens. The *Estrella lausanensis* fusion contains an upstream confirmed stop codon, unlikely to code for an active protein and at first glance no GlgB and probably no GlgA activity are expected. The gene fusion is therefore likely to qualify as a pseudogene. The results listed in table 1, together with the Chlamydiales phylogeny, lead us to propose that an ADP-glc based glycogen metabolism existed in the last common ancestor of Chlamydiales as it diverged from the Planctomycetes-Verrucomicrobia and became an obligate intracellular bacterium. It consisted of the 6 aforementioned enzymes of bacterial glycogen metabolism (GlgC, GlgA, GlgB, GlgP, GlgX, MalQ). GlgC was lost in the common ancestor of Waddliaceae and Criblamydiaceae simultaneously with the appearance of the GlgA-GlgB fusion and that of a novel GlgB gene. The fusion was inactivated in *Estrella lausanensis* suggesting a non-essential function of this protein for chlamydial glycogen metabolism. We will discuss below the significance of these results in the light of a novel MOS-based (malto-oligosaccharides) pathway of bacterial glycogen synthesis. In addition to these basic features, some genes (mostly GlgX and GlgP) have experienced occasional duplications and more frequently gains through LGTs from other diverse bacterial sources (see phylogenetic analysis). Notably the GlgP additional gene copy is carried by the *Simkania negevensis* TRA megaplasmid as cargo.

	GlgC	GlgA (GT5)	GlgB	GlgX	GlgP	MalQ (Denzyme)
Waddlia chondrophila 2032/99	0	1 fus glgB	1 +1fus glgA	1	1	1
Waddlia chondrophila wsu 86-1044	0	1 fus glgB	1 +1fus glgA	1	1	1
Criblamydia sequanensis	0	1 fus glgB	1 +1fus glgA	2	2	1
Estrella lausannensis	0	1 fus glgB (stop)	1 +1fus glgA	2	1	1
Simkania negevensis z	1	1	1	1	2	2
Chlamydia trachomatis a/har-13	1	1	1	1	1	1
Chlamydia trachomatis d/uw-3/cx	1	1	1	1	1	1
Chlamydia trachomatis e/150	1	1	1	1	1	1
Chlamydia trachomatis l2b/uch-1/proctitis	1	1	1	1	1	1
Chlamydia muridarum nigg	1	1	1	1	1	1
Parachlamydia acanthamoebae str. Hall's coccus	1	1	1	2	2	1
Parachlamydia acanthamoebae uv-7	1	1	1	1	2	1
Protochlamydia amoebophila	1	1	1	1	1	1
Protochlamydia naegleriophila	1	1	1	4	1	1
Chlamydophila abortus s26/3	1	1	1	1	1	1
Chlamydophila caviae gpic	1	1	1	1	1	1
Chlamydophila felis fe/c-56	1	1	1	1	1	1
Chlamydophila pecorum e58	1	1	1	1	1	1
Chlamydophila pneumoniae ar39	1	1	1	1	1	1
Chlamydophila pneumoniae lpcoln	1	1	1	1	1	1
Chlamydophila psittaci 6bc	1	1	1	1	1	1

Table 1: Glycogen metabolism enzymes found in complete genome sequences of Chlamydiales. GlgC: ADP-glucose pyrophosphorylase, GlgA: glycogen synthase, GlgB: branching enzyme, GlgP: glycogen phosphorylase, GlgX: debranching enzyme, MalQ: amylomaltase. Fus GlgB stands for fusion between GlgA and the former GlgB gene

Induction of glycogen synthesis and sites of storage polysaccharide accumulation in infected *Acanthamoeba castellanii.*

Among all Chlamydiales species selected for detailed studies, *Chlamydia trachomatis* remains the only specialized pathogen with a restricted host spectrum (humans). All other Chlamydiales are able to replicate in diverse eukaryotic hosts including the single cell amoeba *Acanthamoeba castellanii*. Host glycogen accumulation has been documented to occur in human cells infected by Chlamydiaceae (Gehre et al, 2016). A recent study demonstrated that in addition to cytosolic host glycogen accumulation, *Chlamydia trachomatis* synthesized significant amounts of glycogen within the inclusion through the action of chlamydial effector enzymes. Few studies report the presence of glycogen particles within the bacteria. Nevertheless, significant glycogen accumulation was observed within the pathogens but restricted to *C. trachomatis* EBs.

We applied a specific stain of polysaccharides (PATAg) to localize glycogen particles. Among Parachlamydiaceae, *Parachlamydia acanthamoeba* bacteria contained very abundant glycogen granules while in both *Protochlamydia amoebophila* and of Simkaniaceae, we were unable to detect granules within the inclusion or bacteria (supplementary Figure 1). In contrast, *Estrella lausanensis* and *Waddlia chondrophila* accumulated very significant amounts of glycogen in the bacteria at late stages of infection possibly selectively in EBs (Figure 2).

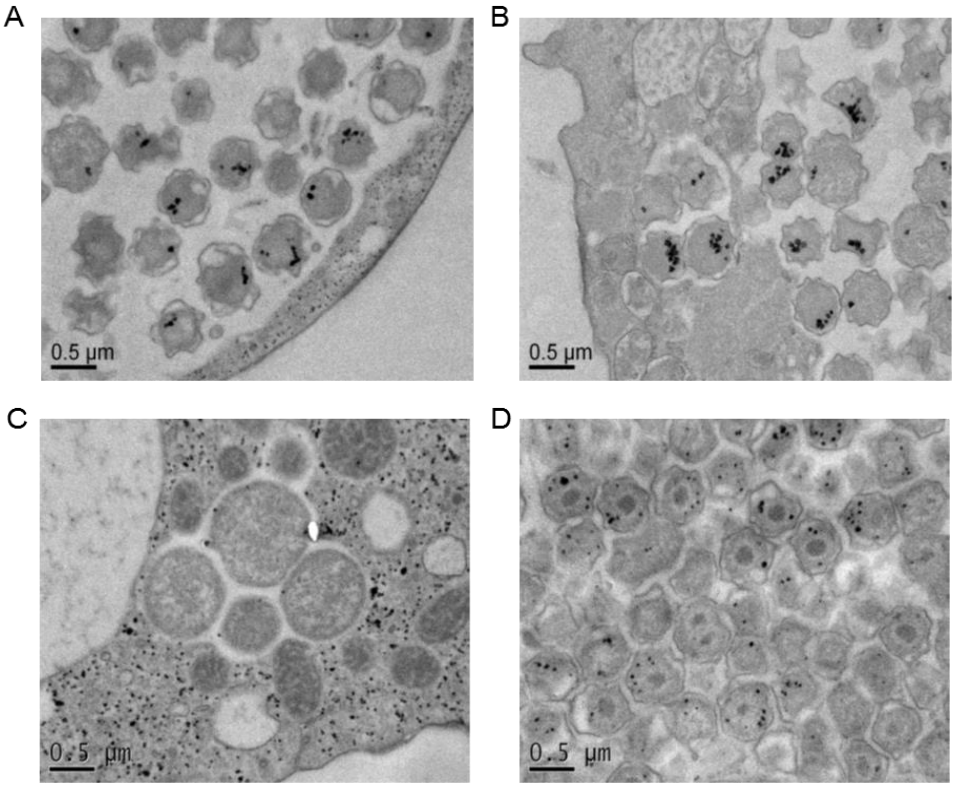


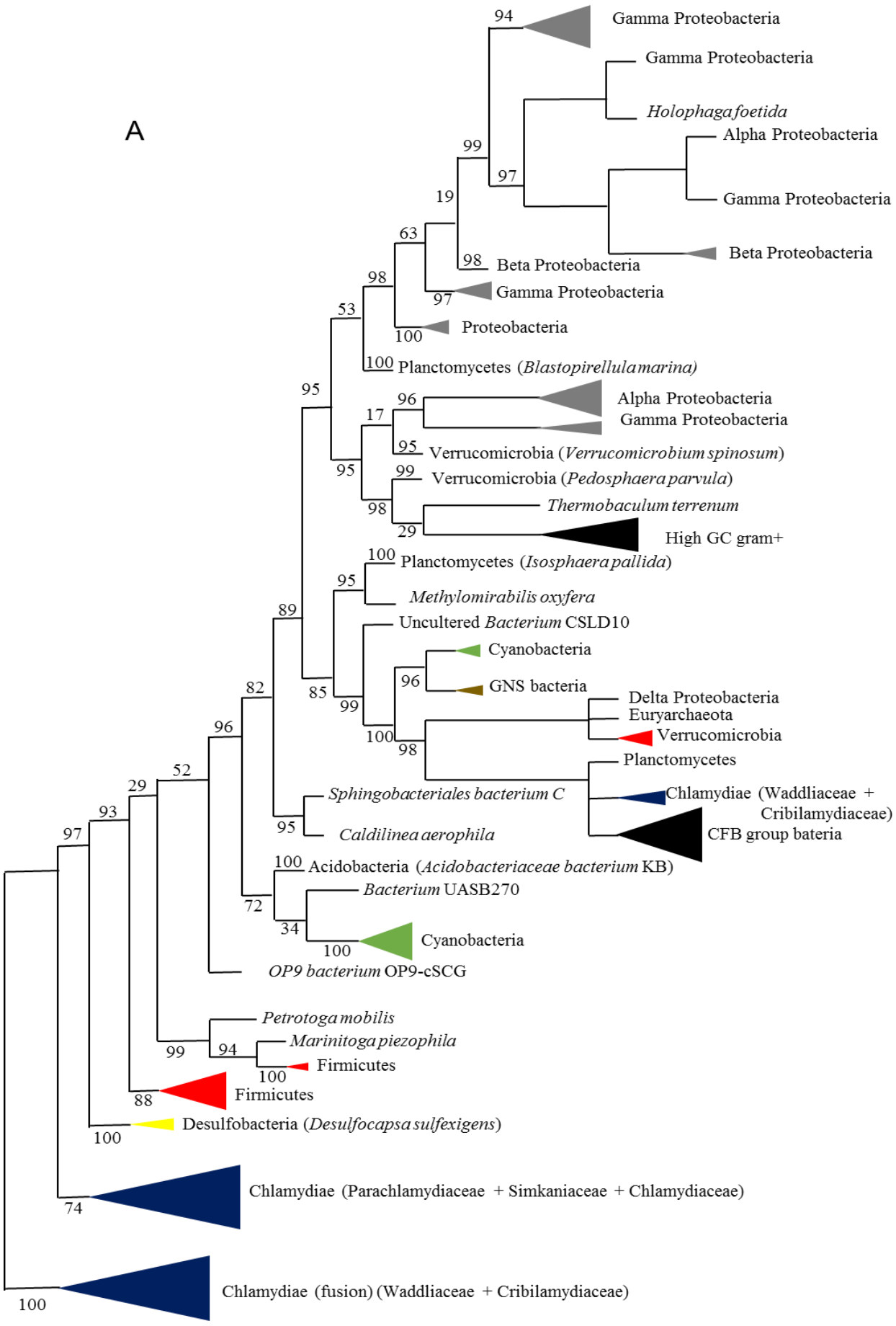
Figure 2: Accumulation of glycogen in infected *Acanthamoeba castellaniin* glycogen was observed by TEM (Transmission Electron Microscopy) after PATAg coloration in *Estrella lausanensis* after 16h (A) and 24h of infection (B), and in *Waddlia chondrophila* after 16h(C) and 24h (D) of infection. Glycogen (black dots) is accumulated within the bacteria for both strains.

Phylogenetic analysis of Chlamydiales glycogen metabolism genes

In order to understand the evolutionary history of glycogen metabolism genes in the Chlamydiales order, we subjected GlgA, GlgB, GlgC, GlgP, GlgX and MalQ to detailed phylogenetic analyses. Phylogenies of GlgX and GlgA have been extensively discussed with respect to their involvement in plastid endosymbiosis. We recapitulate this effort in supplementary figures 2 (GlgA) and 3 (GlgX). The phylogenies of GlgB (Fig3A), GlgC (Fig3B), GlgP (Fig4A) and MalQ (Fig4B) are presented. Chlamydiales behave as a monophyletic group in the phylogenies of all genes with the noticeable exception of GlgP. In the case of GlgB, the additional gene copy gained by the Waddliaceae-Criblamydiaceae upon appearance of the GlgA-GlgB fusion was gained by LGT from a foreign common bacterial source. The diverse planctomycetes-verrucomicrobia found at the root of the Chlamydiales GlgC tree strongly suggests the presence of this sequence in the last common ancestor of

Chlamydiales. This result gives strong support to the presence of the glycogen metabolism ADP-Glc pathway in last common ancestor of Chlamydiales. The GlgA phylogeny however does not show any support for the presence of PVC members at the root of the Chlamydiales. We believe this suggests that either extensive gene sharing has erased the evidence for a PVC origin or, more likely, that the last common Chlamydiales ancestor has received GlgA from a proteobacterial source as it diverged from the PVC ancestors (see supplementary figure 2). Evidence for a more restricted number of PVC sequences has been found for MalQ (Fig4B) and for one particular group of Chlamydiales glycogen phosphorylases (Fig4A) among the numerous subgroups found in this typically polyphyletic phylogeny. Among the 6 genes of Chlamydiales involved in glycogen metabolism only three have not experienced important LGTs with eukaryotes: GlgC, MalQ and GlgB. The GlgA and GlgX LGTs to the common ancestor of Archaeplastida and their important function in the establishment of plastid endosymbiosis has been abundantly discussed previously. To clarify the monophyly of Chlamydiales in these GlgX and GlgA phylogenies we have also constructed trees without archaeplastidal sequences (supplementary figures 2 and 3).

The chlamydial MalQ encodes a GH77 α-1,4 glucanotransferase related to Dpe2 an enzyme known in plants to selectively hydrolyse maltose produced by β-amylase. α-1,4 glucanotransferases are very commonly found in bacteria. Thousands of bacterial sequences are evidenced presently in the data bases while the enzyme distributes only to those eukaryote clades that accumulate starch or glycogen and contain β-amylase (archaeplastida, cryptomonads, amoebozoa and a few glycogen accumulating excavates). No fungus nor animal contain such sequences. Assuming therefore a prokaryotic origin for these enzymes, the phylogeny displayed in Figure 4B suggests a bacterial LGT from Spirochetes to ancestral eukaryotes. The implications of this LGT for the evolution of β-amylase and eukaryotic glycogen metabolism are discussed. The eukaryotic Dpe2 gene has in addition been transferred to gut bacteria as previously reported (Arias et al., 2012).



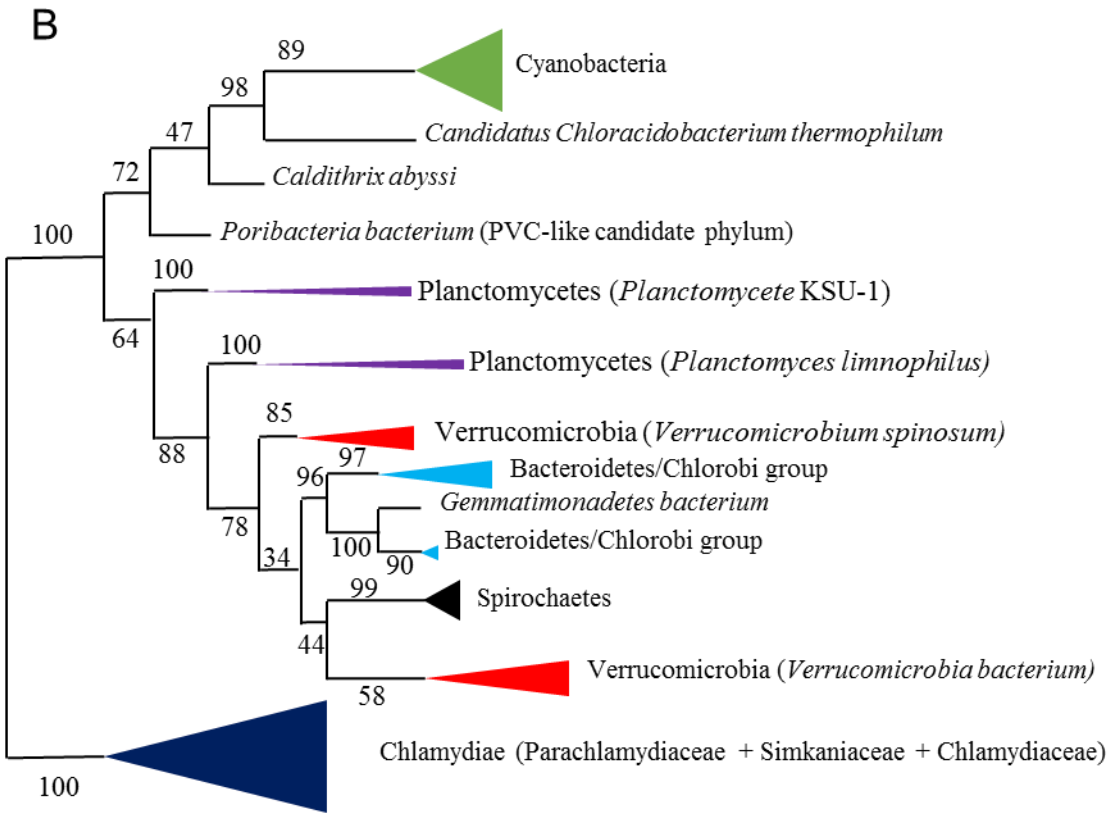
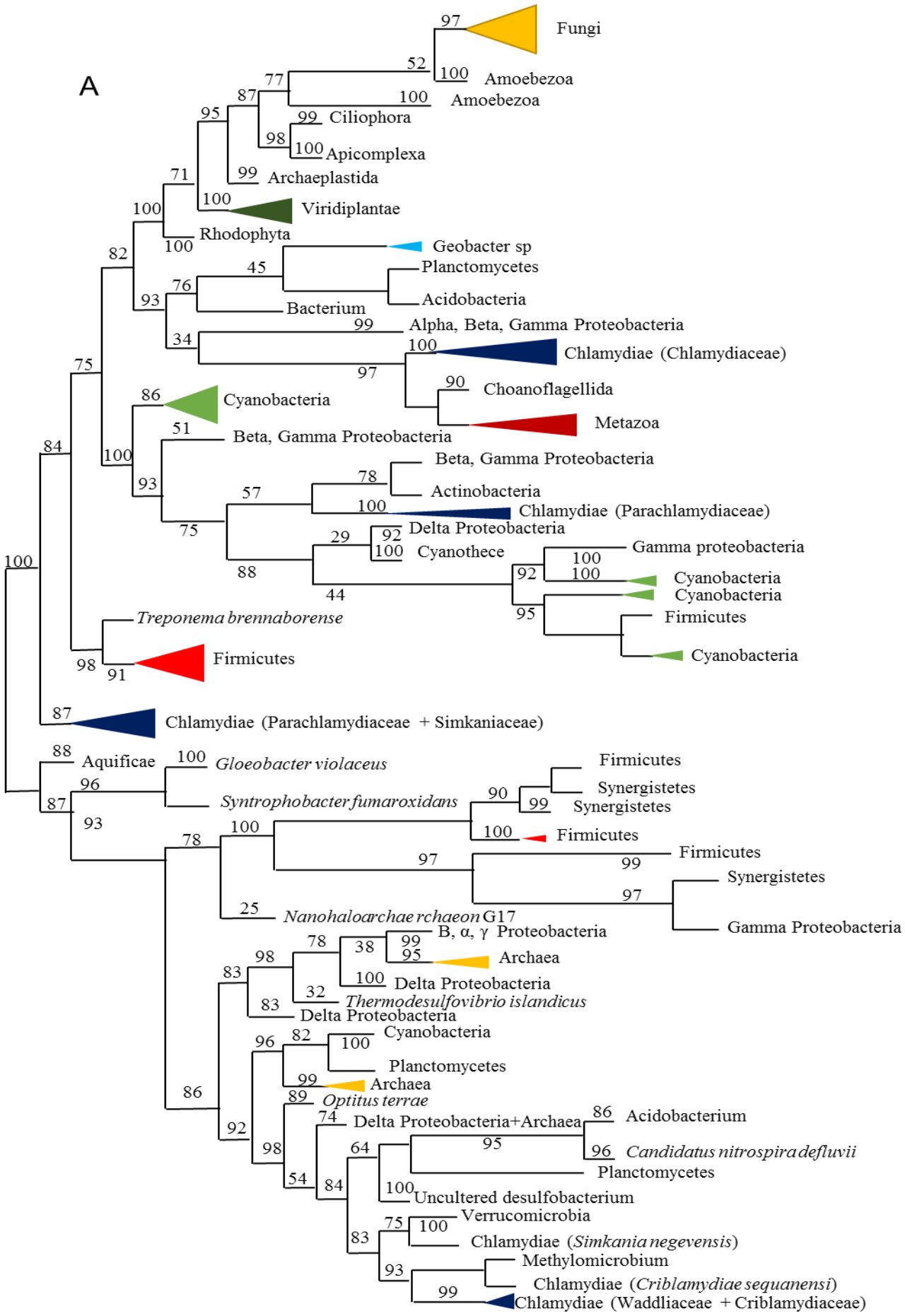


Figure 3: Schematic representation of maximum likehood phylogeny with bootstraps values mapped into the branches for glycogen branching enzyme (A) and ADP-glucose Pyrophosphorylase enzyme (B).

The phylogeny of Chlamydiales GlgP is typically polyphyletic (Figure 4 A). The sequences of Waddliaceae, one of the several Simkaniaceae GlgP and Criblamydiaceae cluster with other PVC sequences suggesting that this group contains an ancestral GlgP gene.

Parachlamydiaceae-Simkaniaceae have experienced LGT from other bacterial sources and replacement of its ancestral gene in at least two separate occasions. Quite unexpectedly Fig4A shows that Chlamydiaceae (the animal pathogens) and all animals, Ichthyosporea and choanoflagellates are united by LGT with good support. At first glance the topology supports an LGT from the Chlamydiaceae to their hosts. However the number of bacterial clades separating the animal-and choanoflagellates from the other eukaryotic clades is restricted, calling for caution as additional LGTs among bacteria might obscure the eukaryotic affiliation of the animal glycogen phosphorylase. A reasonable level of congruence was observed for this enzyme between the basal choanoflagellates, Ichthyosporea and animals. We thus reasoned that, if the glycogen phosphorylase was of ancient eukaryotic origin we should detect sisterhood with the fungal enzyme if we eliminate bacteria from the analysis. Our preliminary data suggests that this is clearly not the case and we can thus safely propose that

the ancestor of animals and choanoflagellates received this gene from the common ancestor of Chlamydiaceae. In addition the closest prokaryotic relatives of this GlgP phosphorylase contain several Spirochaete, Bacteroides-chlorobi and Planctomytes, which define sisters to Chlamydiales in 16S rRNA phylogenies, thereby suggesting that this sequence may define a second ancestral phosphorylase sequence among Chlamydiales. We therefore conclude that the animal glycogen phosphorylase defines an enzyme of chlamydial origin. Replacement of the ancestral eukaryotic phosphorylase occurred at the earliest stage of holozoa diversification thereby proving that Chlamydiaceae co-evolved as pathogens with the animal lineage.



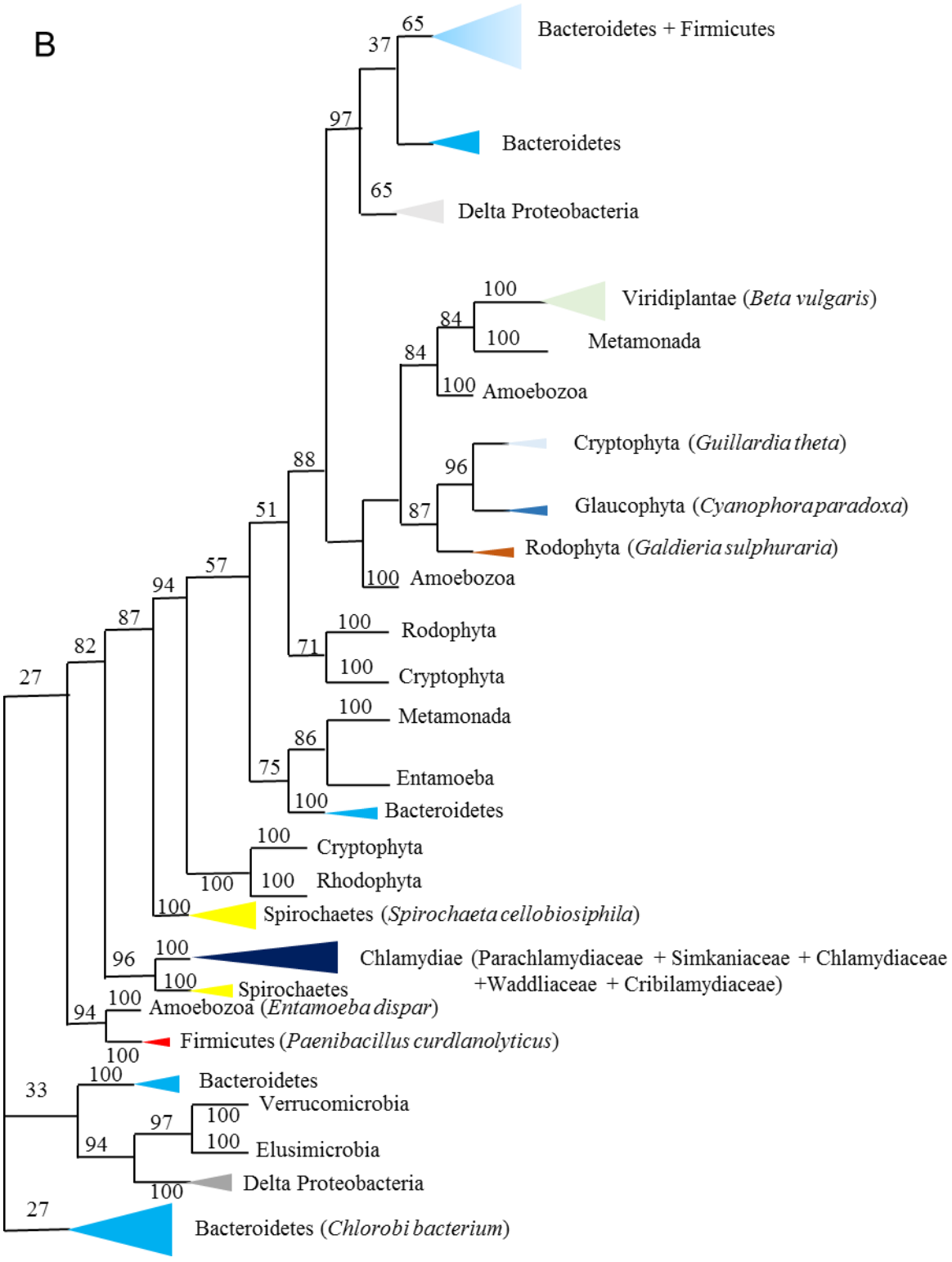


Figure 4: Schematic representation of maximum likehood phylogeny with bootstraps values mapped into the branches for glycogen phosphorylase enzyme GlgP (A) and amylomaltase MalQ (B).

Investigating the effector nature of glycogen metabolism enzymes in a semi *in vitro* system from *Shigella flexenerii*

In Chlamydiales, protein effectors synthesized within the bacterium are secreted through or within the inclusion by the type three secretion system. An*in vivo* assay in *Shigella flexneri* has been set up to identify TTS secretion sequences in chlamydial proteins. The assay is based on conservation of structural determinants of the TTS within bacteria. *Shigella flexneri* was chosen as a practical experimental system to probe 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 was validated in a genome-wide screen for TTS in *Chlamydia pneumoniae*with a rate of false positives that fell under 5%. To test whether proteins of glycogen metabolism of our selected Chlamydiales contained a functional TTS signal, we made fusions between the first 30 codons from the genes and the reporter gene, and tested their secretion by *S. flexneri*. Previous studies report similar experiments performed in *C. trachomatis* and Parachlamydiaceae. However only a minor part of the enzymes of glycogen metabolism were analyzed and no data had been produced for Simkaniaceae, Waddliaceae and Criblamydiaceae. The results displayed in Table 2 and Figure 5 show that upstream sequences derived from all, but one, enzymes of glycogen metabolism in all tested Chlamydiales enabled the TTS-mediated secretion by Shigella of the reporter protein. The important exception consisted of *C. trachomatis* GlgC as previously reported.

Chlamydiales	GlgC	GlgA	GlgB	GlgP	GlgX	MalQ
Protochlamydia amoebophila	+	+	+	+	+	ND
Chlamydia trachomatis D/UW-3/CX	-	+	+	+	+	ND
Estrella lausannensis	ND	+	ND	+	ND	+
Parachlamydia acanthamoebae str. Hall's coccus	+	+	+	ND	ND	+
Simkania negevensis Z	+	+	+	+	ND	ND
Waddlia chondrophila WSU 86-1044	+	ND	ND	ND	ND	+

Table 2 : Summary table showing the secretion of the glycogen metabolism enzymes of Chlamydiales by the heterologous system of *Shigella flexnerii*. GlgC: ADP-glucose pyrophosphorylase, GlgA: glycogen synthase, GlgB: branching enzyme, GlgP: glycogen phosphorylase, GlgX: debranching enzyme, MalQ: amylomaltase. ND: presently not determined

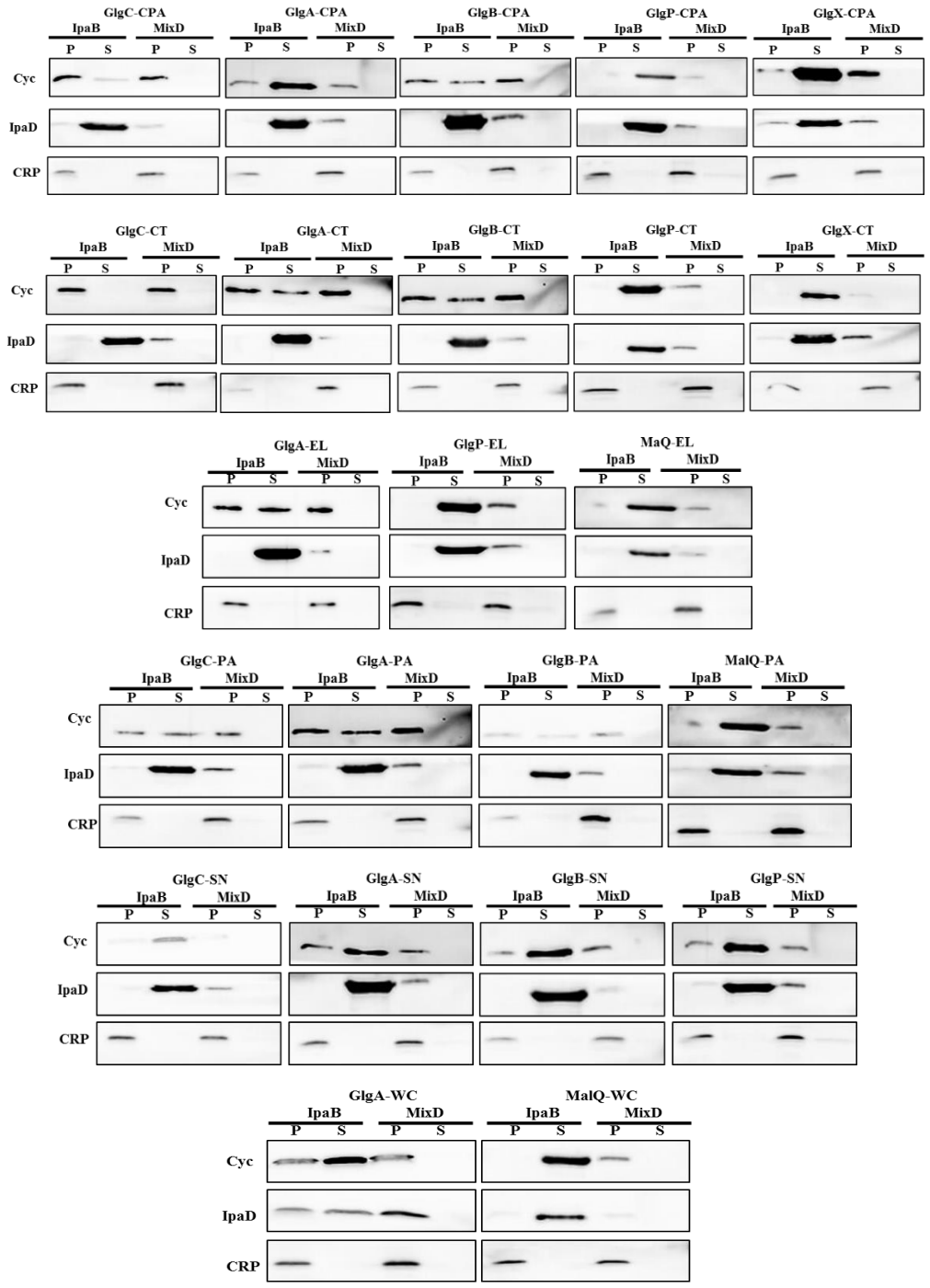
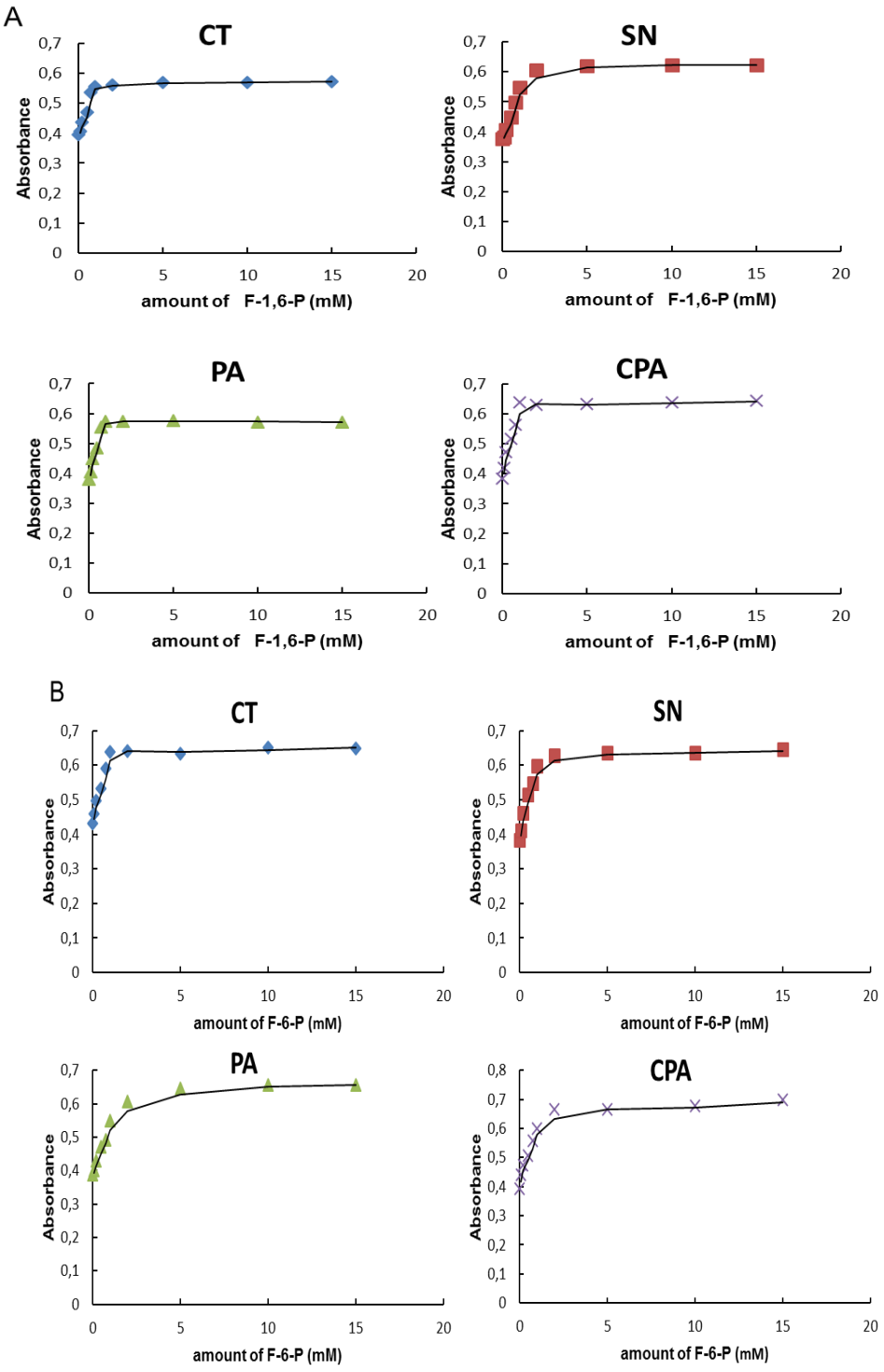


Figure 5: Secretion test of the recombinant proteins of the Chlamydiales glycogen metabolism in a heterologous system *Shigella flexnerii*. The 30 amino acids of the N-terminal extremity of each protein were fused to a reporter gene (adenylate cyclase). The recombinant proteins were expressed in the *Shigella flexenerii* strains IpaB (T3SS+) and MixD (T3SS-). Supernatant (S) and pellet (P) were analyzed by western blot using the antibody anti-adenylate cyclase. IpaD is a positive control secreted via the T3SS and the negative control CRP is non secreted via the T3SS. All the enzymes are secreted by the T3SS except GlgC of *Chlamydiea trachomatis.* CPA: *Protochlamydia amoebophila*, CT: *Chlamydiea trachomatis*, EL: *Estrella lausanensis*, PA: *Parachlamydia acanthamoeba*,SN: *Simkania negevensis*, WC: *Waddlia chondrophila*

Characterization of the GlgC and GlgA recombinant proteins from Chlamydiales

We purified the tagged recombinant GlgC from Chlamydiaceae, Parachlamydiaceae and Simkaniaceae. All enzymes behaved in a similar fashion and were activated by fructose-1,6-diphosphate, fructose-6-phosphate, pyruvate and being inhibited by Adenosine Monophosphate (AMP)(Figure 6), This regulation is reminiscent of the regulation of bacteria relying on both glycolysis and Entner-Doudoroff pathways which is different from enterobacteria such as *E. coli* or *Salmonella* whose enzyme is activated solely by fructose-1,6-diphosphate and is very different from that operating in cyanobacteria and green algae and plants which are activated by 3-PGA.



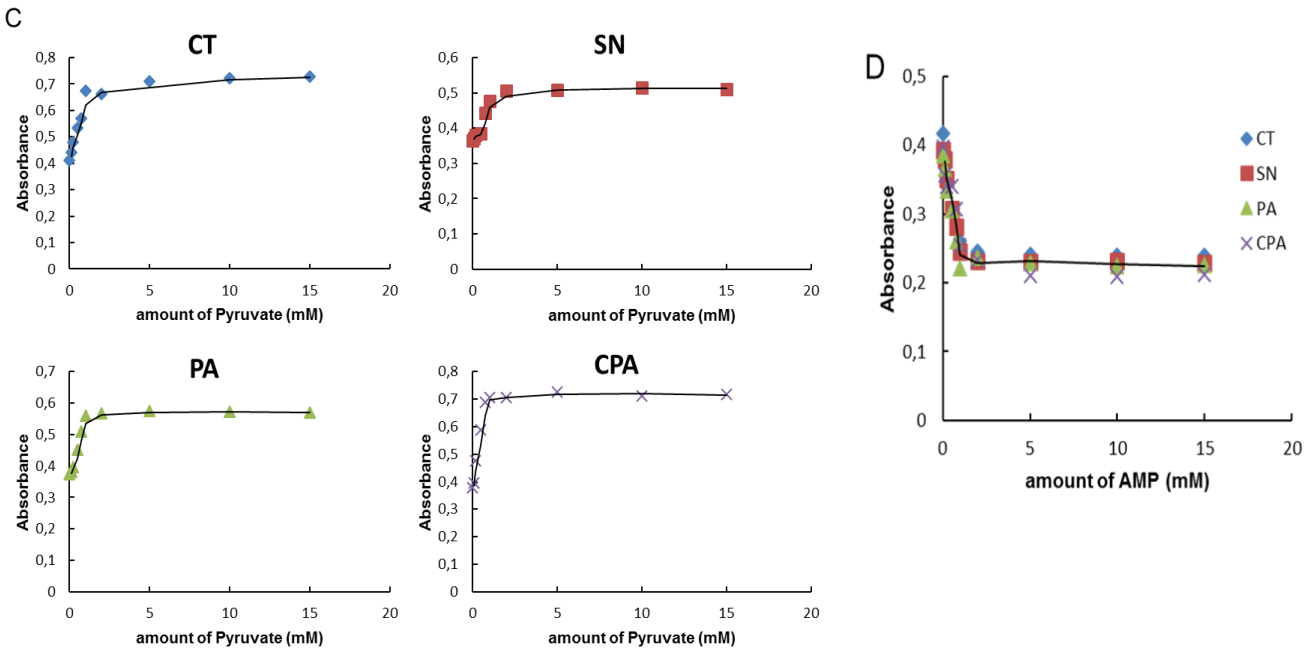


Figure 6: Characterization of the ADP-Pyrophosphorylase of Chlamydiales. The recombinant proteins of*Chlamydiae trachomatis*(CT), *Simkania negevensis* (SN), *Parachlamydia acanthamoeba* (PA) and *Protochlamydia amoebophila* (CPA) were expressed in *Escherichia coli* strain. Purified recombinant enzyme activity is measured in the direction of ADP-Glc degradation. Hence the production of glucose-1-P was measured by a spectrophotometer after incubation with phosphoglucomutase and using the glucose-6-P dehydrogenase assay (NADPH production). The enzymes were incubated with Fructose 1,6 diphosphate (F-1,6-P) or Fructose 6 phosphate (F-6-P) or Pyruvate or Adenosine Monophosphate (AMP). All the enzymes were activated by F-1,6-P, F-6-P and Pyruvate and inhibited by AMP.

We purified partially the untagged recombinant GlgA from *Chlamydia trachomatis*, *Parachlamydia acanthamoeba*, *Protochlamydia amoebophila*, *Simkania negevensis*, *Waddlia chondrophila* and *Estrella lausanensis*. As expected we recovered no activity from the Estrella defective gene fusion. and tested their activity with both ADP-Glc and UDP-Glc as well as their activity using either glycogen, amylopectin or maltoheptaose. We also checked the processive or distributive nature of the activities in zymogram gels in the presence of glycogen primer added to the gel. Results listed in Table 3 and Figure 7 showed that most enzymes that showed significant radiolabel incorporation into polysaccharide from nucleotide sugars also gave a significant iodine stain in our zymogram gels containing the same substrate. Because the iodine stain starts developing when the polymerization reaches a minimum of 12 glucose residues, and because the extinction coefficient of the iodine-polysaccharides increases significantly upon further increasing the chain length, this result implies that the GlgA proteins displayed a significant level of processivity in the polymerization reaction. A noticeable exception to this otherwise general property of

Chlamydiales GlgA proteins was found in *Simkania negevensis*. In this case while incorporation of label into glycogen was routinely found, we were unable to detect any iodine stain in our numerous zymogram assays. This proves that *Simkania negevensis* GlgA is a highly distributive rather than processive enzyme. Only *Chlamydia trachomatis* was able to use more efficiently UDP-Glc than ADP-Glc, while maintaining high levels of activity with the less preferred ancestral substrate (ADP-Glc). All other Chlamydiales displayed the expected preference for ADP-glucose. *Parachlamydia acanthamoebae* and *Waddlia chondrophila* displayed very low albeit significant levels of activity with UDP-Glc while *Simkania negevensis* and *Protochlamydia amoebophila* could not use this substrate. These results correlate with the effector nature of ADP-glucose pyrophosphorylase. We correlated these results to our quantified complementation assays of the *E. coliΔglgA* mutant (results taken from Mathieu Ducatez thesis are represented for convenience in Figure 8). All chlamydial enzymes with the noticeable and expected exception of *Estrella lausanensis* were able to complement the *glgA* mutation. However despite the highest of all levels of complementation displayed by the *Waddlia chondrophila* enzyme (110% of the measured wild-type reference), we were unable to recover any iodine staining of glycogen in the complemented strain, nor were we able to recover staining with the Simkania sequence which nevertheless gave over one third of the wild-type glycogen levels. We believe this to be due to the accumulation of shorter chain glycogen in both of these complemented bacteria. Such structures possibly due to activity of the GlgB domain of the Waddlia fusion or to the distributive nature of the Simkania enzyme would fall below the level of detection with iodine.

	ADP-glucose	UDP-glucose
Escherichia coli	0,81±0,23	-*
Parachlamydia acanthamoeba	0,53±0,24	ND**
Chlamydia trachomatis	ND	1,70±0,65
Protochlamydia amoebophila	0,72±0,25	-
Waddlia chondrophila	0,79±0,17	-
Simkania negevensis	2,8±0,45	-
Estrella lausanensis	-	-

**: ND Not Determined (linear slope) *:- No activity

Table 3: Kinetic experiments of the recombinant proteins glycogen synthase of the *Escherichia.coli* and the Chlamydiales order (*Parachlamydia acanthamoeba, Chlamydia trachomatis, Protochlamydia amoebophila, Waddlia chondrophila, Simkania negevensis and Estrella lausanensis)*. Activities were measured by following the incorporation of the radiolabeled C^{14} from the ADP-Glc or the UDP-Glc. Data represented in this table are apparent Km values that were calculated using GraphPad sofware. We believe these data to be overestimates because of underappreciated product inhibition of the measured enzyme activites

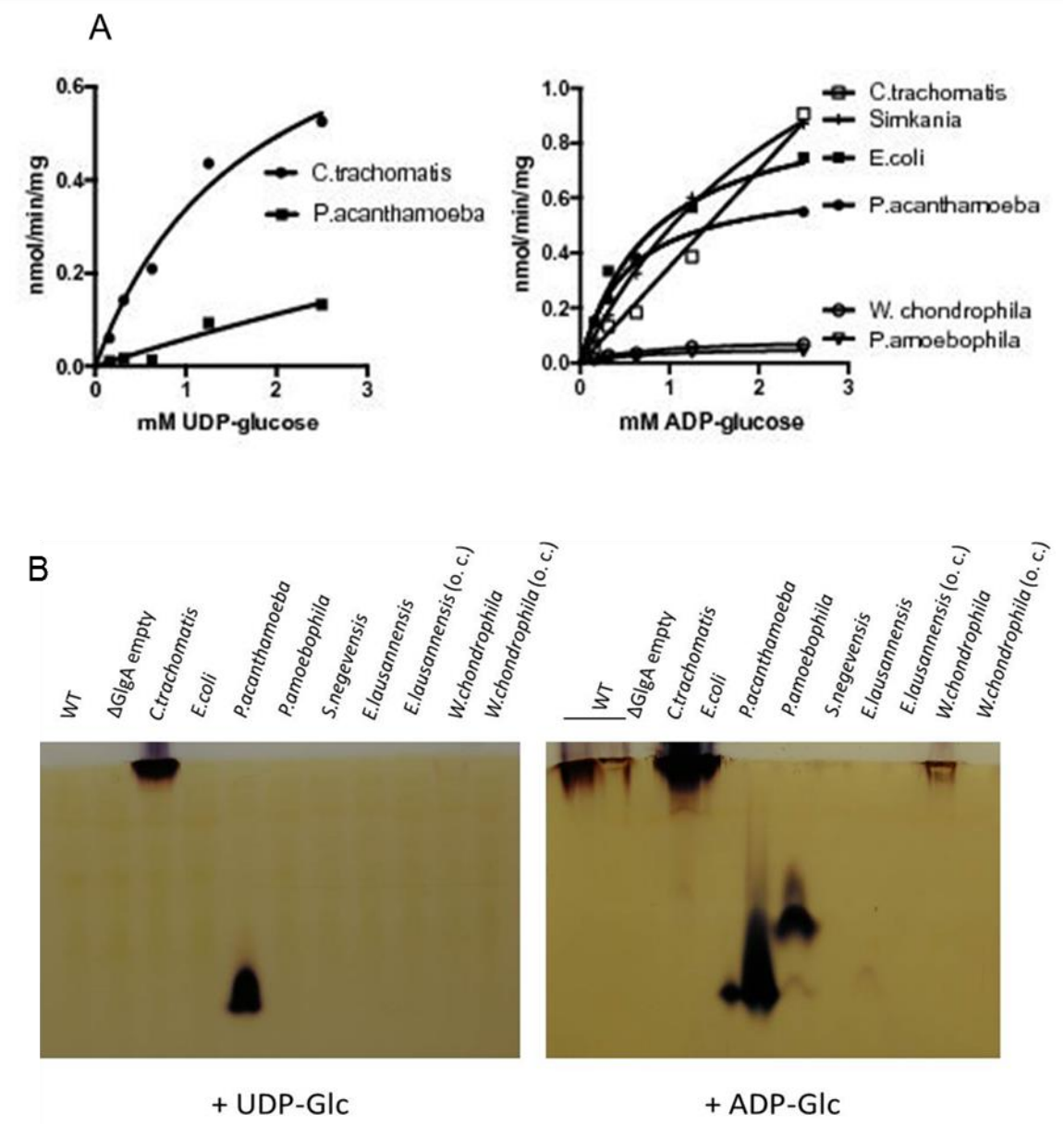


Figure 7:A. Kinetic experiments of *Parachlamydia acanthamoeba, Chlamydia trachomatis, Protochlamydia amoebophila, Waddlia chondrophila, Simkania negevensis and E.coli*recombinant glycogen synthase. The activities were measured by following incorporation of C^{14} radiolabeled ADP-Glc or UDP-Glc. Km were calculated according the incorporation of C^{14}. **B.** Zymogram gels of recombinant glycogen synthase of Chlamydiales: the recombinant proteins were expressed in the ΔglgA *E.coli* strain, the total proteins of crude extract were separated on a polyacrylamide gel containing glycogen as substrate, after migration the gels are incubated in a buffer with ADP-Glc or UDP-Glc (1.2mM) overnight. The glycogen synthase activities are revealed after ioding staining. Thes results are taken from Mathieu Ducatez thesisand used for convenience (Ducatez, 2016).

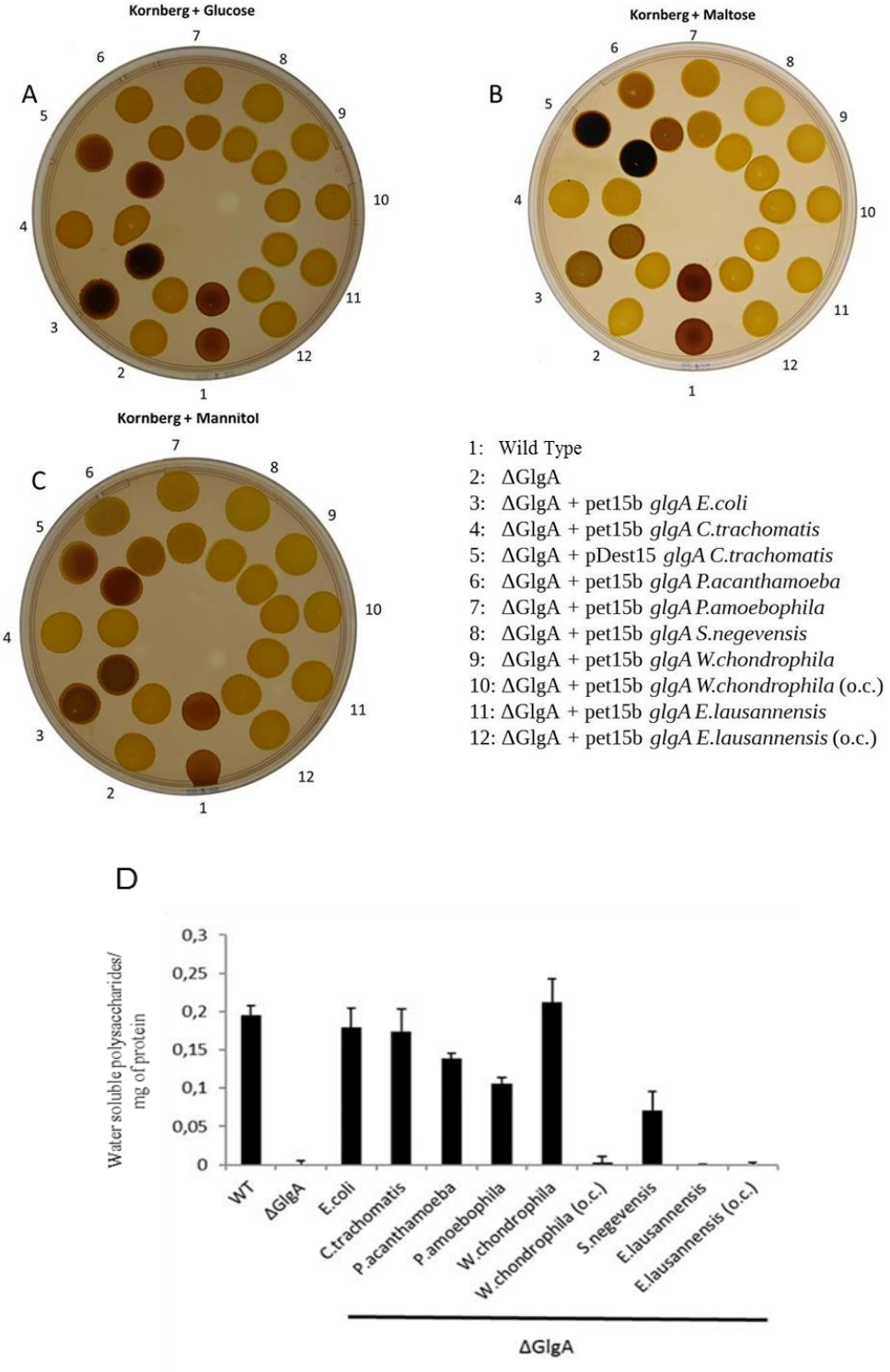


Figure 8: Complementation of ΔglgA *E.coli* strain by the recombinant proteins of Chlamydiales. A. Iodine staining of the ΔglgA *E.coli* strain expressing the glycogen synthase of Chlamydiales on Korenberg medium complemented with glucose (**A**), maltose (**B**) and mannitol (**C**). After an overnight incubation at 37°C the plates were stained with iodine. **D.** Diagram representation of the complementation in liquid M9 medium supplemented with maltose, results consist of total glycogen assays present in bacterial pellets.Thes results are taken from Mathieu Ducatez thesis and used for convenience (Ducatez, 2016).

Characterization of the GlgP and GlgX recombinant proteins of Chlamydiales

Recombinant proteins from Chlamydiacae, Waddliaceae and Criblamydiaceae were successfully produced in *E. coli* ΔCAP mutants defective for GlgP (Figure 9A and B). These results were obtained by Sylvain Laurent and previously reported in Mathieu Ducatez's thesis. We purified the recombinant GlgP proteins and assayed their activity in both, the presumed "physiological" (release of glucose-1-P) direction, and the reverse (polymerization) direction of synthesis, in the presence and absence of respectively glucose-1-P and orthophosphate. In both, our zymogram and assays, we used rabbit liver as well as maltohexaose and maltotetraose as primers. We produced in addition recombinant *E. coli*GlgP and MalP as respectively glycogen and maltodextrin (MOS) phosphorylase references. Results are displayed in Figure 9. Despite the detection of recombinant protein from *Chlamydia trachomatis* on western blot (Figure 9 B), the enzyme gave no stain in our zymogram gels suggesting the presence either of an inactive enzyme or of a highly distributive mode of action of the Chlamydiaceae protein. The detection of radiolabel glucose into polymerized product should clarify the situation.

Unlike the GlgP of *E coli*, our results show no inhibition of the Waddlia and Estrella phosphophorylases by glycosyl-nucleotides (Figure 9C) and higher affinity for both MOS (Km 1.1±0,5 $mg.mL^{-1}$ and 1.6 ±0,4 $mg.mL^{-1}$, respectively) and glycogen (not shown) unlike the MalP (3.3±0,5 $mg.mL^{-1}$) or GlgP (16.4±4,8 $mg.mL^{-1}$)which preferred respectively maltooligosaccharides and polysaccharides (Figure 9 D). We further tested the distribution of the products generated by the chlamydial phosphorylases of *Waddlia chondrophila* and *Estrella lausanensis*. In both cases the enzyme produced a substantial proportion of maltriose, which was never detected with GlgP nor MalP (Figure 10) and is reported as a possible phosphorylase product.

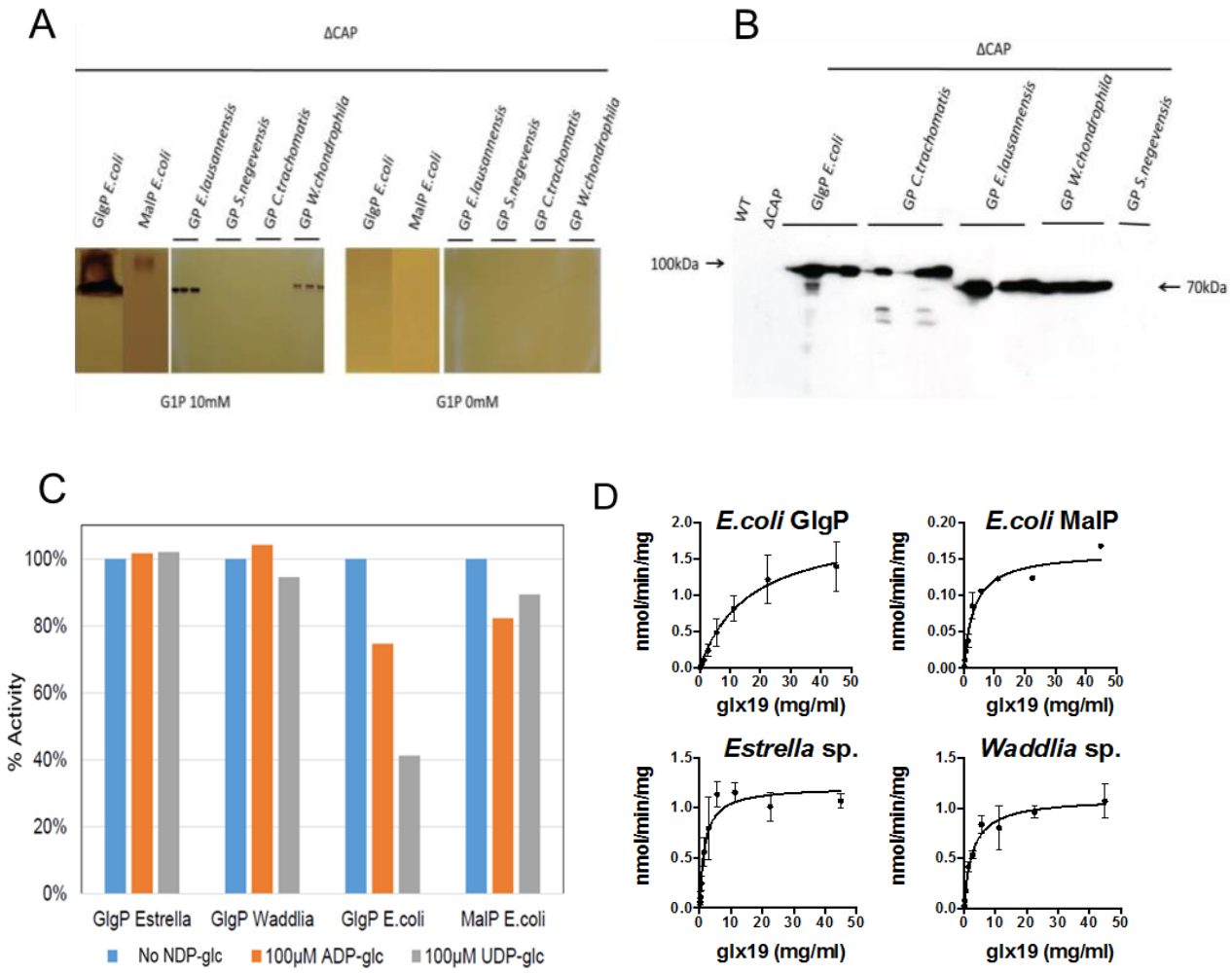


Figure 9: Characterization of the Chlamydiales glycogen phosphorylase activities. **A.** Zymogram analysis of the glycogen phosphorylase activities of *Estrella lausanensis,Simkania negevensis, Chlamydiae trachomatis, Waddlia chondrophila* and GlgP, MalP of *Escherichia coli.* The recombinant proteins were expressed in ΔCAP *E.coli* strains (defective for GlgC, GlgA and GlgP), the crude extracts were separated on a native polyacrylamide gel containing glycogen, after migration the gel was incubated in a buffer containing or not 10mM of G1P. The gel was stained with iodine solution to reveal the glycogen Phosphorylase activities. **B.** Western blot analysis of the glycogen phosphorylase expression using an antibody directed against the His tag of the recombinant proteins. **C.** Impact of the nucleotide sugar on the glycogen phosphorylase activities. The activity was measured after 15min of reaction at 30°C using the Glucidex 19 as a substrate, in addition or not of the nucleotide sugar ADP-glc or UDP-glc (100µM). **D.** Km determination of phosphorylase activities for malto-oligosaccharide (Glx19). Kinetic experiments were carried out in the phosphorolysis direction by measuring the release of glucose-1-phosphate). Glx19 (MOS) are composed of a population of linear glucan chains from 6 to 12 residues of glucose. Phosphorylase activities are expressed as nmol of G-1-P released/min/mg of protein.

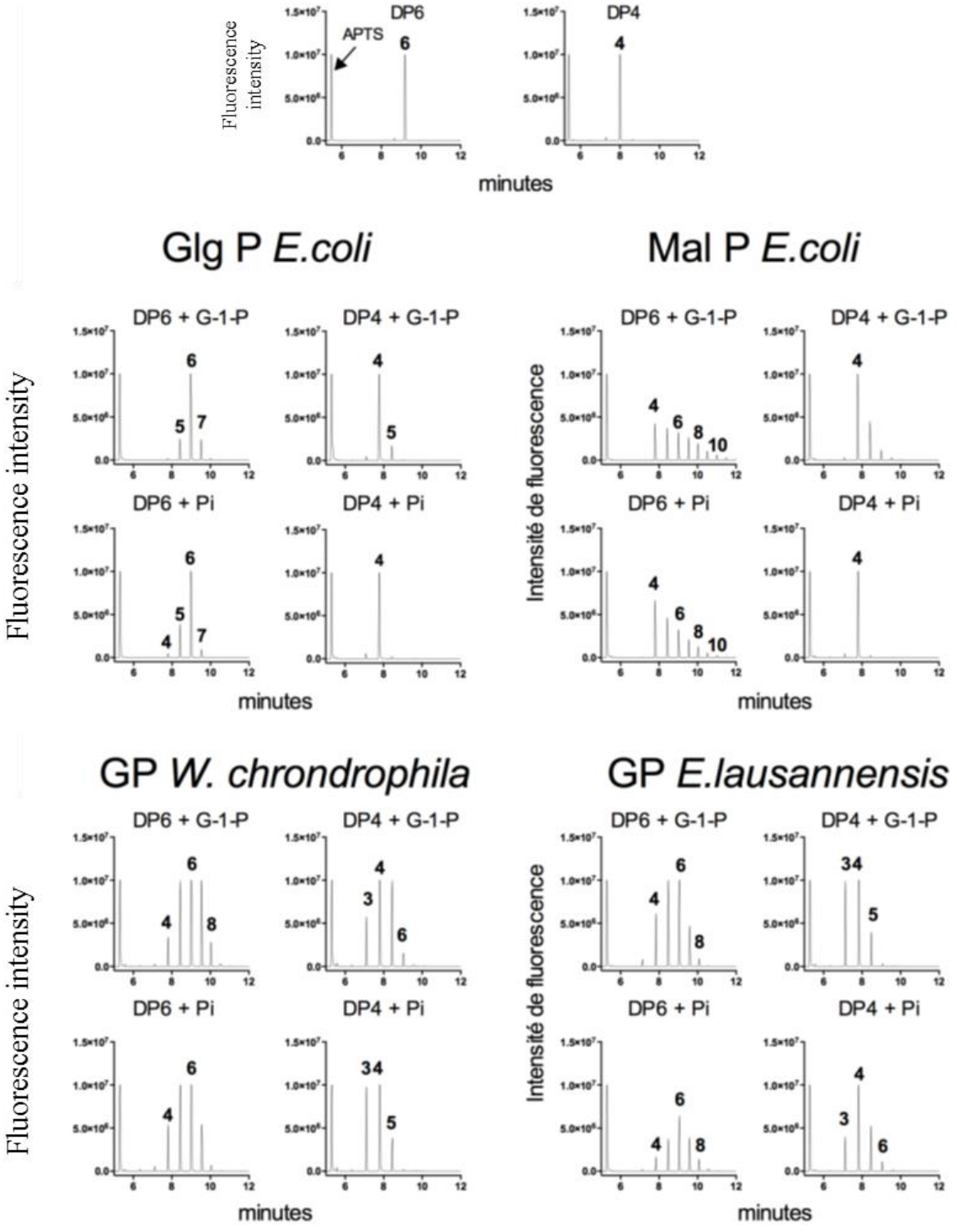


Figure 10: Face (Fluorophore Assisted Carbohydrate Electrophoresis) analysis of the chains produced by the glycogen phosphorylase of*Estrella lausanensis, Waddlia chondrophila,* GlgP and MalP of *Escherichia coli* using DP4 (maltotetraose) or DP6 (maltohexaose) as a substrate.The chains were separated according to degree of polymerization (DP) by capillary elctrophoresis FACE.

We then turned our attention to the purified glycogen debranching enzyme (GlgX) of Chlamydiaceae, Parachlamydiaceae, Simkaniaceae, Waddliaceae and Criblamydiaceae (Figure 11A). These recombinant enzymes were compared in the same experiments with the behavior of recombinant *E. coli* GlgX, maize Isa1 (isoamylase 1), and commercial *Pseudomonas* sp isoamylase (Figure 11B). We tested the chlamydial activities with respect to glycogen and amylopectin predigested or not by rabbit liver glycogen phosphorylase. Because all enzymes offered high activities with respect to both intact glycogen and amylopectin, we chose these 2 substrates for our more detailed characterizations (Figure 12A and 12B). GlgX and isoamylases belong to the same subfamily of CAZy enzymes (GH13 subfamily 11) but display very different kinetics with respect to the chain lengths preferentially released during the debranching reaction (Stam et al., 2006). In agreement with its function the bacterial GlgX enzymes are exquisitely specific for the release of the very short external chains produced by the glycogen phosphorylases (mostly maltotetraose) (Figure 11B) (Dauvillée et al., 2004). These enzymes displayed significantly slower debranching of longer chains. Because glycogen branching enzyme produces chains that are longer than 6 glucose residues during synthesis, the restrained activity of GlgX toward chains longer than 6 prevents the presence of a futile cycle during glycogen synthesis. Some bacteria secrete analogous enzymes to digest most often starch-derived amylopectin-like molecules in the external medium, which are among the most abundant polysaccharides found in the biosphere. In agreement with this function, the bacterial "isoamylases" digest chains above 6 glucose residues very fast and display significantly slower debranching of shorter chains (Figure 12A and 12B). Plant isoamylases, such as the maize sugary-1 gene product Isa1, have acquired a specific function for the production of semi-crystalline structures that aggregate into starch granules that thereby escape further enzymatic modifications (James et al., 1995). This function consists of trimming off those chains that prevent this aggregation (Ball et al., 1996). Hence, these chains are debranched more as a function of the spacing of the branches and their accessibility during the branching-debranching cycles than of a selective chain length. The futile cycle energy costs that are entailed by this mechanism are limited thanks to the action of α-1,4 glucanotransferases. In agreement with this function, starch-producing plant isoamylases digest both long and short chains present in the maturing polysaccharides. The kinetics of the chains released by the maize enzyme is in perfect agreement with this function (figure 12A and B). Chlamydiales have been reported to accumulate glycogen both internally and in the inclusion. In addition, induction of cytosolic glycogen accumulation may be at least in part

triggered by the secretion of chlamydial effectors in the cytosol. If for a particular Chlamydiales species internal accumulation is privileged (as in *Waddlia chondrophila* or *Estrella lausanensis* (this work)) we would expect the enzyme to maintain the GlgX-type of debranching kinetics. In those cases, where little or no glycogen is evidenced in the bacteria these constraints would not apply and a GlgX or isoamylase type of specificity will be preferred as function of the speed of cytosolic host glycogen mobilization potentially induced by the pathogen effectors (fast for isoamylases and moderate for GlgX because they respectively do not or do require prior degradation by phosphorylase). GlgX type of enzymes specificities are evidenced clearly in *Waddlia chondrophila* (Figure 12A and 12B) and *Estrella lausanensis* (Figure 12A and 12B). Interestingly the *Waddlia chondrophila* enzyme is surprisingly effective on external chains of three residues in glycogen which is the structure produced selectively by its own glycogen phosphorylase (see above). An intermediate type of kinetics is evidenced in *Simkania negevensis* (Figure 12A and 12B) which still qualifies as a true GlgX but with the ability to debranch slightly longer chains. The debranching enzymes of *Protochlamydia amoebophila* (Figure 12A and 12B) and *Chlamydia trachomatis* (Figure 12A and 12B) display clear-cut isoamylase type of debranching kinetics. These kinetics are so surprisingly close to those of the maize starch-producing enzyme Isa1, that we tested the ability of these Chlamydiales to induce starch production in plants (see below).

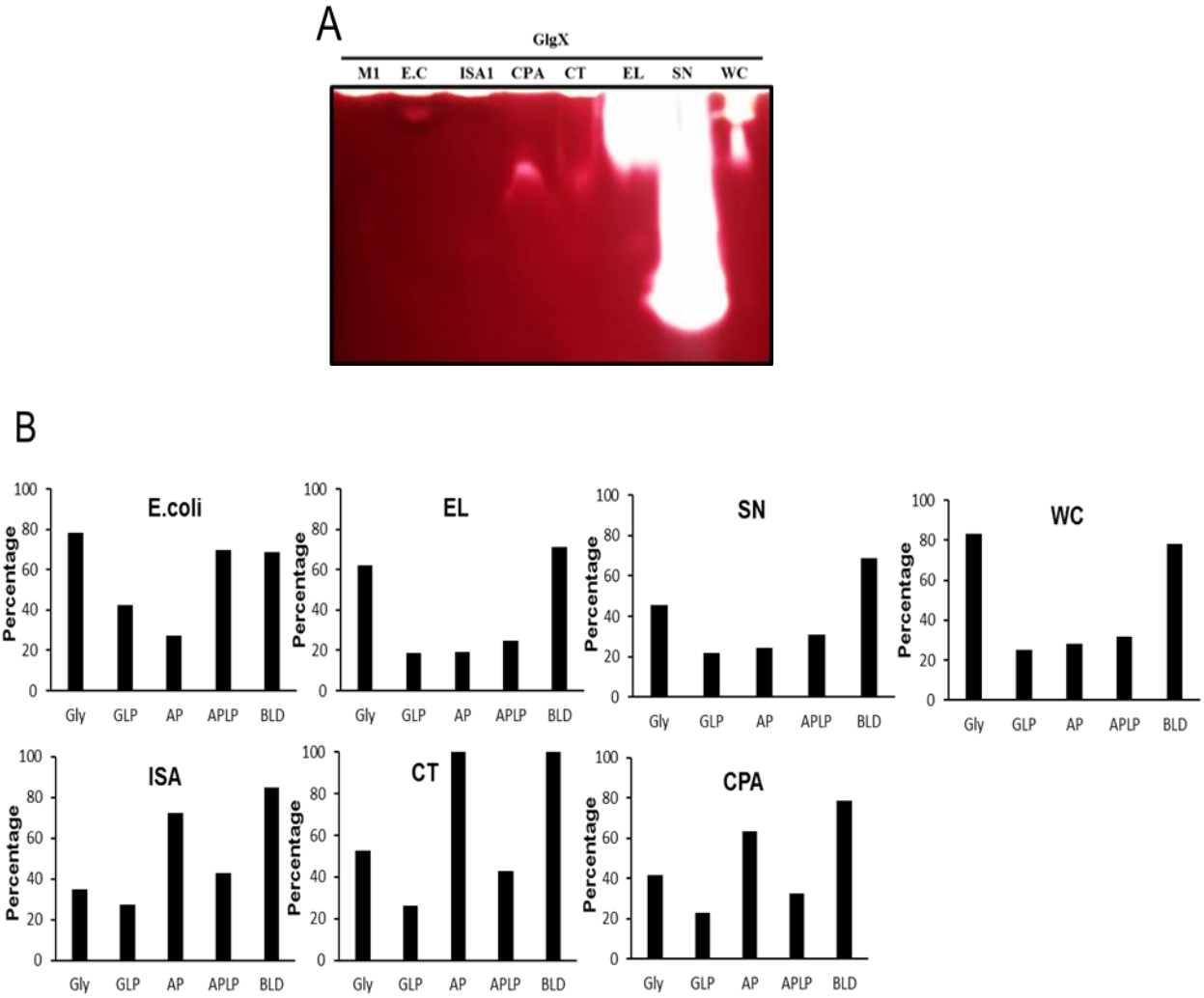


Figure 11: Characterization of the Chlamydiales order glycogen debranching enzymes. **A.** Zymogram analysis of the glycogen debranching enzyme activities of *Protochlamydia amoebophila* (CPA), *Chlamydiae trachomatis* (CT), *Estrella lausanensis* (EL), *Simkania negevensis* (SN), *Waddlia chondrophila* (WC),isoamylase1 of maize (ISA1) and GlgX of *Escherichia coli* (E.coli). The recombinant proteins were expressed in *E.coli* strains, the crude extracts were separated on a native polyacrylamide gel containing beta limit dextrin, after migration the gel was incubated in buffer A. The gel was stained with iodine solution to reveal the glycogen debranching activities. **B.** The purified enzymes were incubated in a buffer containing glycogen (Gly), amylopectin (AP), glycogen limit phosphorylase dextrin (GLP), amylopectin limit phosphorylase dextrin (APLP) and beta-amylase limit dextrin (BLD). The results expressed as the percentage of reducing ends released represent the ratio of the amounts of reducing ends released for each recombinant debranching enzyme over the total amounts of reducing ends released for each polysaccharide using a commercial isoamylase type of debranching enzyme (Pseudomonas, Megazyme) and Dinitrosalicylic acid (DNS) assay.

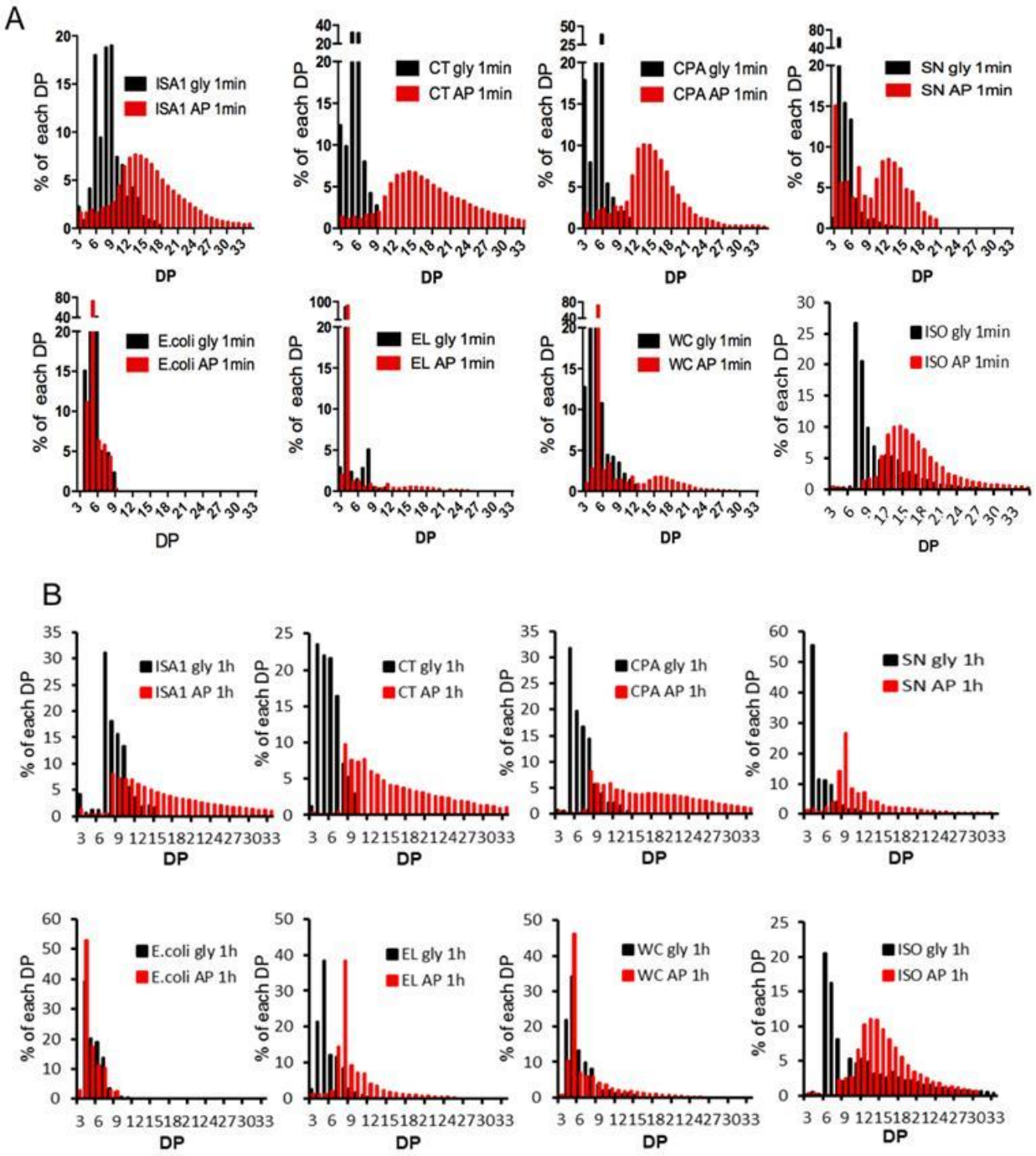


Figure 12:FACE analysis of the glucan chains released by the debranching enzymes of Chlamydiales order, *E.coli* and ISA1 of maize. The chains were separated according to degree of polymerization (DP) by capillary elctrophoresis FACE (Fluorophore Assisted Carbohydrate Electrophoresis) after 1min **(A)** and 1hour **(B)** of incubation. The y axis represents the percentages of chains corresponding to each DP.

The *Protochlamydia amoebophila* isoamylase restores leaf starch synthesis in glycogen accumulating mutants of Arabidopsis thaliana

Because the GlgX proteins of Parachlamydiaceae and Chlamydiaceae display identical properties to the maize Isa1 protein responsible for the production of starch in plants, we assayed the ability of the chlamydial proteins to trigger starch accumulation from preexisting storage polysaccharide metabolism networks. These mutants substitute starch by glycogen synthesis. Expression of the Protochlamydia enzyme restored starch synthesis to over half of the wild-type amount (Figure 13).

From these experiments we can deduce that expression of the chlamydial GlgX enzyme in a preexisting glycogen network does not by itself trigger starch synthesis. However, expression of the same enzyme in a starch metabolism network that lacks this activity, and therefore accumulates glycogen, does.

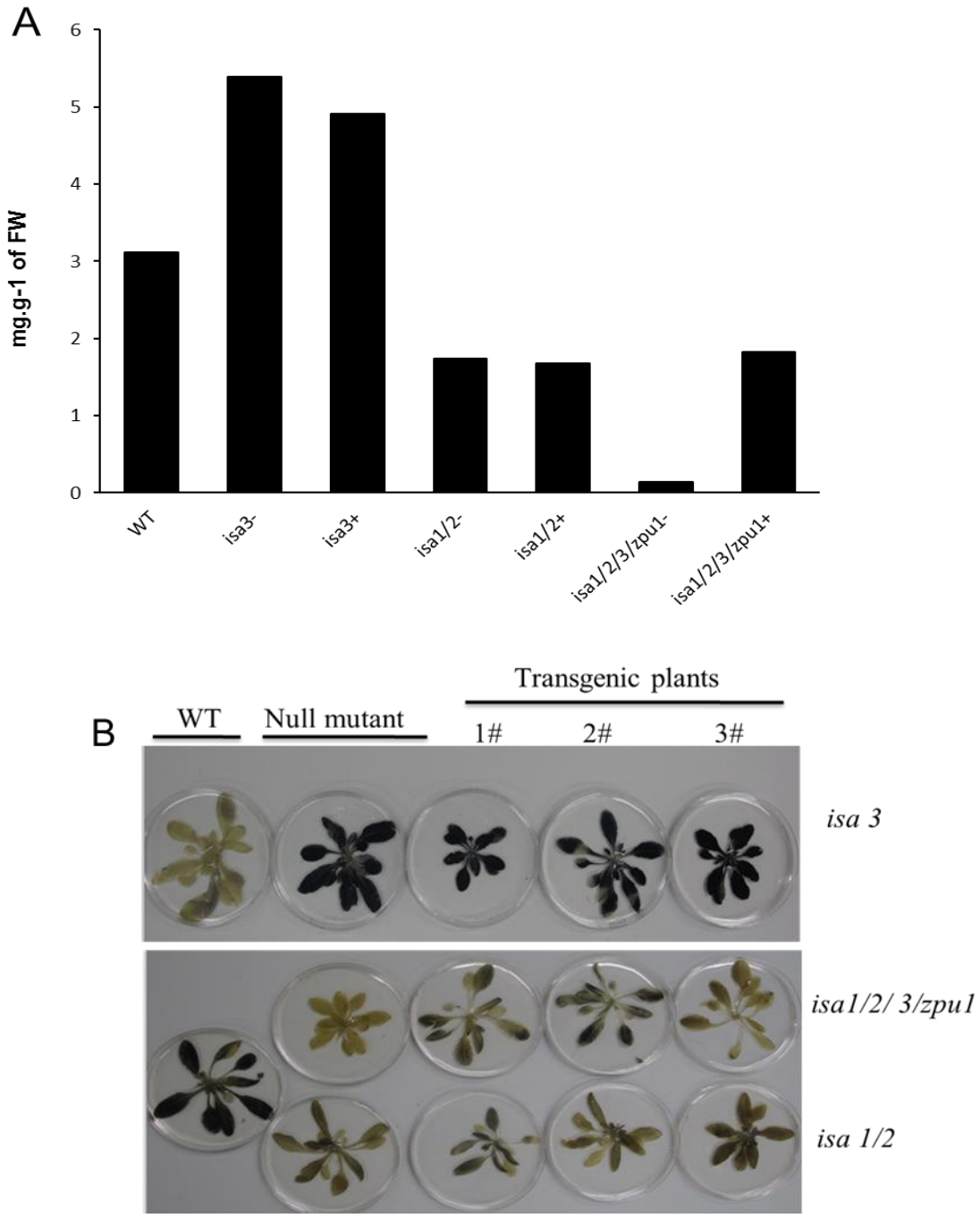


Figure 13: Complementation of the *Arabidopsis thaliana* isoamylase mutants with the debranching enzyme of *Protochlamydia amoebophila.* **A.** Diagramm showing the quantities of glycogen produced by the wild type (WT) and the *Arabidopsis thaliana* isoamylase mutants transformed by the debranching enzyme of *Protochlamydia amoebophila.* **B.** Iodine staining of the *Arabidopsis thaliana* mutants leaves expressing the debranching enzyme of *Protochlamydia amoebophila.*

IV. Discussion

The prokaryotic ancestry of the animal glycogen phosphorylases may explain its sensitivity to AMP allosteric activation

The study of mammalian glycogen phosphorylase has probably impacted the history of biochemistry more than that of any other single enzyme. It has led to the discovery by Gerty and Carl Cori of the first allosteric enzyme through its activation by AMP(Cori et al., 1943). At the same time, they achieved the first complete *in vitro* synthesis of a biological polymer (amylose and glycogen) which earned them the Nobel prize(Cori and Cori, 1943). Several students of the Cori lab pursued on solving the mystery of conversion between the active AMP independent (a form) and inactive AMP dependent (b form) forms of glycogen phosphorylase. This led Edwin Krebs and Edmond Fisher to the discovery of phosphorylase kinase and of enzyme regulation by protein kinases(Fischer and Krebs, 1955; Krebs and Fischer, 1955). Earl Sutherland, another Cori lab student of glycogen phosphorylase discovered cAMP and the first protein phosphorylation cascade in response to hormone stimulation that impacted the switch from the AMP requiring and AMP independent glycogen phosphorylase forms(Rall et al., 1956; Fischer, 2013).

Edwin Krebs underlined that the potato phosphorylase did not respond to AMP. Several researchers discovered more recently that this was equally true for yeast, *Neurospora crassa* (Shepherd and Segel, 1969), *Dictyostelium discoideum* (Jones and Wright, 1970), and probably also for *Trichomonas vaginalis*(Nielsen et al., 2012), *Giardia lambia*, *Mastigamoeba balamuthi* and *Entamoeba histolytica*(Wu and Muller, 2003). Hence neither plants, amoebas, excavata or fungi seem to display a glycogen phosphorylase enzyme sensitive to AMP allosteric regulation(Rathore et al., 2009). The mechanisms of protein phosphorylation and the impact on enzyme activity have been studied in great detail in both yeast and mammals(Wilson et al., 2010). Both fungi and animals display analogous sophisticated cascades leading to enzyme activation or deactivation. However, the detailed molecular mechanisms are strikingly different. To be specific, the nature of the protein directly responsible for phosphorylation of the fungal enzyme is unknown to this day but it is certainly not a phosphorylase kinase type of enzyme. In addition, the site of protein phosphorylation and the consequences on enzyme structure are also distinct in the animal and yeast enzyme. By contrast, bacterial glycogen phosphorylases, unlike bacterial maltodextrin phosphorylases or eukaryoticphosphorylases often respond to activation by AMP

(Schinzel and Nidetzky, 1999). We have demonstrated that the holozoan and choanoflagellate glycogen phosphorylase gene was donated by the ancestor of the animal specific Chlamydiaceae pathogens. As many bacterial glycogen phosphorylases are sensitive to AMP activation, we believe this property was transferred to the ancestor of holozoa by Chlamydiales. The reasons explaining the switch from the ancestral eukaryotic enzyme to the pathogen's phosphorylase are unclear. We can at least speculate that acquisition of this gene led to subsequent evolution of a novel kinase to integrate the new AMP sensitive protein in the complex pattern of kinases and phosphatases regulation present in eukaryotic metabolism. In turn, this may have been favored by natural selection as it may have tamed the chlamydial effector enzyme into a less virulent and more controlled relationship with host metabolism. The finding that the chlamydial LGT happened before holozoa diversification but after separation from the fungi proves that the important mammalian pathogen Chlamydiaceae have co-evolved with the choanoflagellates and animals over 700 millions years ago a sobering thought when considering that they witnessed all animal specific defense mechanisms including their diverse immune systems.

Lateral gene transfer of GH77 α-1,4 glucanotransferase from Spirochetes may explain the evolution of cytosolic β-amylase in the eukaryotic domain.

"Carly, Carly, its free glucose!" When Joseph Larner made a vivid account of his glycogen research years in the Cori lab, he told how Gerti Cori ran down to her husband office informing him that Joseph had just found that the final product of glycogen debranching. This product was not glucose-1-P (the Cori ester) nor small oligosaccharides but free glucose. Joseph Larner had just discovered how eukaryotes debranch the final product of glycogen particle digestion by glycogen phosphorylase(Cori and Larner, 1951). What they did not know, at the time, is that the complex mechanism of glycogen degradation witnessed in all eukaryotes did not apply to prokaryotes. The glucose produced in the eukaryote cytosol could be immediately recycled after phosphorylation. In the case of bacteria however, direct debranching of the external maltotetraose left over after particle digestion by glycogen phosphorylase, releases maltotetraose and not glucose in the bacterial cytosol. This requires the presence of malto-oligosaccharide metabolism and the further processing by GH77 α-1,4 glucanotransferases. These enzymes are never found in animals or fungi but can be occasionally found in other eukaryotic clades such as glycogen accumulating amoebozoa and excavates and Archaeplastida and some of their starch accumulating secondary endosymbiosis derivatives. The eukaryotic α-1,4 glucanotransferases are restricted to one

particular subtype called Dpe2, originally reported in *Arabidopsis thaliana*. By comparison, thousands of α-1,4 glucanotransferases are detected throughout the prokaryotic tree of life. Because the mechanism of glycogen debranching does not require the presence of malto-oligosaccharide metabolism in eukaryotes, we can safely conclude that the root of the GH77 tree must lie within prokaryotes. With this in mind, the topology that we now report in the MalQ phylogeny, demonstrates that an ancestral eukaryote received Dpe2 possibly from Spirochetes ancestors. Despite the relatedness between the chlamydial and eukaryotic enzymes, the phylogeny clearly rejects a possible chlamydial origin. Depending on where this ancestral protein was originally expressed, different scenarios can be proposed. If the protein was initially expressed on its own in lysosomes to digest MOS from preys, then, transfer of such sequences would have been favored irrespective of the nature of the bacterial donor. These enzymes would indeed have facilitated the digestion of oligosaccharides during phagocytosis of bacteria. Nevertheless, if this was the case we would expect the presence of eukaryotic Dpe2 in diverse phagotrophic lineages that do not synthesize glycogen or starch as storage material. This is presently not the case, and the enzyme only distributes to Archaeplastida, diplomonads, parabasalids, amoebozoa, cryptomonads and not to β-glucan storing eukaryotes. In fact, Dpe2 is found systemically in the presence of β-amylase in these clades with the possible exceptions of *Dictyostelium discoideum* and *Giardia lambia* where it is apparently found on its own. The selective role of Dpe2 in the assimilation of the disaccharide maltose, generated by β-amylase, is well established in plants. Because Dpe2 remains to be found in the absence of β-amylase in phagotrophs that do not accumulate starch or glycogen, we surmise that this LGT happened after evolution of the novel amylase in protists. This amylase may indeed have evolved first for lysosomal degradation in the presence of lysosomal glucosidases able to degrade maltose. Subsequent Dpe2-MalQ LGT from bacteria (Spirochetes) for expression of MalQ in the cytosol would have allowed the redirection of β-amylase in this compartment, thereby creating a novel hydrolytic pathway of cytosolic glycogen mobilization. This new pathway is the predominant one used by extant Archaeplastida to mobilize starch. In support of this hypothesis, the β-glucan accumulating *Naegleria gruberi*, an excavate phagotroph, contains β-amylase genes in the absence of dpe2. What is presently unclear, are the reasons favoring the selection of this pathway in the cytosol of glycogen accumulating eukaryotes. Hydrolysis is clearly redundant and less energy efficient than classical glycogen phosphorolysis. One possibility would be that this pathway specialized in cytosolic degradation of abnormal glycogen, which consists of hyperphosphorylated starch-like material that escapes standard phosphorolytic digestion (see

below). Clearly, the status of this pathway in amoebas and Naegleria and the precise localization of the enzymes to the cytosol or lysosome require further investigations to test these speculations.

Lateral gene transfer of chlamydial GH13 isoamylase may explain the evolution of starch in the cytosol of the last common ancestor of Archaeplastida.

It is now well established that starch evolved several times from a pre-existing glycogen metabolism machinery. Detailed functional analysis in plants and green algae have led to the discovery of glucan trimming, a process by which isoamylases debranch those chains that prevent polysaccharide aggregation into insoluble starch granules (Ball *et al*., 1996). Cyanobacteria and plants have been demonstrated through genetic and biochemical analysis to both use a GlgX-derived direct debranching enzyme that has acquired an isoamylase-type of substrate selectivity. Surprisingly, the source of the plant enzyme is chlamydial and not cyanobacterial. In this respect, we believe that finding extant Chlamydiales that produce a GlgX protein with a substrate specificity of cleavage identical to the maize starch producing Isa1 isoamylase is significant. Expression by itself of such an enzyme in a glycogen metabolism network, will not necessarily immediately trigger starch accumulation, but a particular balance of branching enzyme and elongation activities in the presence of the chlamydial isoamylase such as the one observed in Archaeplastida or single cell diazotrophic cyanobacteria will. Complementation of plant mutants defective for isoamylase proves that the presence of the chlamydial isoamylase effector preadapted the cytosolic glycogen metabolism machinery to a switch to starch. It appears that all glycogen synthesizing cells, if given sufficient time will accidently lead to the accumulation of abnormal insoluble polysaccharides known as Lafora bodies. Lafora bodies consist of hyperphosphorylated glycogen with decreased branching and a tendency to aggregate and become insoluble(Roach, 2015). Although these structures are reminiscent of starch there are important differences: starch is not hyperphosphorylated, does not form by similar mechanisms and does not consist of aggregates of glycogen particles. In lafora bodies, it is generally assumed that it is the excess phosphate that both reduces the branching and leads to aggregation. The source of glycogen phosphorylation is presently unclear and affords for a lively debate between those advocating the presence of a glycogen kinase and those sustaining a side reaction of glycogen synthase. Anyhow, Lafora bodies escape degradation by glycogen phosphorylase and it is therefore important to prevent their accumulation with time. Animals contain laforin, a dual specificity phosphatase that prevents hyperphosphorylation of glycogen by directly cleaving

out the phosphate present on the abnormal particles. In its absence, humans develop a fatal progressive myoclonic epilepsy due to accumulation of abnormal glycogen in the brain. It is known that extant plants mobilize starch by first phosphorylating it by a dikinase enzyme and then hydrolyzing it through β-amylases. Glycogen/starch phosphorylase is unable to participate in this process. β-amylases are thus more adapted to digest aggregates of phosphorylated polysaccharides. It is entirely possible, that this property of β-amylase is shared by all protist enzymes and that the cytosolic hydrolytic pathway of glycogen degradation is, in fact, specialized in removing abnormal hyperphosphorylated glycogen. The laforin pathway of abnormal glycogen clearing is present in animals, in the ancestors of Archaeplastida, in cryptomonads and alveolates while the β-amylase pathway that we now propose is present in amoebozoa excavates Archaeplastida and cryptomonads. To date fungi seem to define the only glycogen/starch accumulating eukaryotes deprived of both of these pathways. How and why fungi escape the formation of abnormal glycogen is not known.

Evidence is accumulating suggesting that, at the time of endosymbiosis, the cyanobacterial ancestor of plastids was sheltered in the parasitophorous vacuole (the inclusion) of Chlamydiales, thereby escaping the antibacterial defense systems of the host. In addition, it is believed that Chlamydiales donated through conjugation the genes encoding critical transporters allowing efflux of photosynthetic carbon as well as operons encoding genes that were overexpressed in the cyanobiont. In addition to this, the pathogens secreted protein effectors in the cytosol that were essential for polymerizing glucose from photosynthesis into the host glycogen stores thereby linking host and symbiont metabolic networks. These effectors consisted of the GlgA and GlgX gene products. This polymerization of photosynthate from the cyanobiont to the host cytosolic stores defines the symbiotic link uniting the future plastid and its host. The three organisms were therefore engaged in a tripartite symbiosis known as the MATH (Ménage à Trois Hypothesis)(Facchinelli et al., 2013). The presence in the cytosol of both the chlamydial isoamylase effector and of the host β-amylase allowed for the synthesis of starch-like material. The presence of β-amylase ensured that these polysaccharide aggregates that resulted from the trimming of glycogen chains by the chlamydial effectors could be nevertheless mobilized. In order to ensure optimal activity of β-amylase, the aggregates needed to be phosphorylated. This resulted from a single innovation which consisted of a gene fusion involving two pre-existing domains: a carbohydrate binding module and a dikinase (Ball et al., 2015). The resulting glucan water dikinase ensured phosphorylation of the polysaccharides processed by the chlamydial isoamylase. The presence of laforin ensured that the phosphoglucans resulting from digestion

could be further metabolized. This finalized the switch from glycogen to starch metabolism in the cytosol of the Archaeplastida ancestor. In the context of chlamydial infection, it is reasonable to speculate that this switch was selected because it restored host control of catabolism of its carbohydrate stores. Indeed the GlgP chlamydial effectors of polysaccharide catabolism had lost its direct access to glycogen and could not interact with semi-crystalline starch.

Biochemical properties of recombinant GlgA of Parachlamydiaceae and Simkaniaceae supports the Ménage à Trois hypothesis.

The Ménage à Trois Hypothesis (MATH) hypothesis is presently sustained by a plethora of phylogenetic and molecular evidence giving a very detailed view of the early events of plastid endosymbiosis. All this evidence agrees with a scenario where the cyanobiont exported carbon in the form of G6P in the inclusion lumen where Chlamydiales controlled glycogen synthesis occurred through ADP-glucose (Facchinelli et al., 2013; Facchinelli et al., 2013). The excess ADP-glucose would then be exported by an inclusion membrane associated ADP-glc transporter (NST: nucleotide sugar translocator)(Colleoni et al., 2010). In turn, the exported glycosyl nucleotide would be incorporated into the host glycogen stores thanks to the GlgA chlamydial effector. This requires the presence of an ADP-glucose pathway in the ancestral chlamydiales. It also requires that all glycogen metabolism enzymes including ADP-glucose pyrophosphorylase be TTS effectors to be secreted both in the inclusion lumen and the host cytosol. Very recently, experimental evidence was produced in Chlamydiaceae, demonstrating that in *C. trachomatis* glycogen synthesis was indeed sustained in the inclusion lumen by glycogen metabolism enzyme effectors from the pathogen (Gehre et al., 2016). In this system however, glycogen synthesis operated through UDP-glucose in the inclusion lumen and host cytosol. In line with this observation, ADP-glucose pyrophosphorylase was not an effector but remained within the bacteria to sustain bacterial glycogen synthesis through ADP-glucose. Inclusion glycogen synthesis in *C. trachomatis* required the presence of a UDP-glucose translocator on the inclusion membrane. Experimental inhibition of synthesis of the major human UDP-Glc NST yielded a strong decrease of glycogen synthesis within the inclusion, proving the importance of this NST (Gehre et al., 2016). Our work now proves that Parachlamydiaceae and Simkaniaceae the closest Chlamydiales taxa in trees supporting chlamydial LGTs sustain glycogen metabolism solely through ADP-glucose. Unlike *C. trachomatis* their ADP-glucose pyropshophorylase (GlgC) is shown to be a candidate effector in the Shigella test system. This observation further strengthens the MATH

by showing that extant Simkaniaceae and Parachlamydiaceae do behave as predicted by this hypothesis.

A novel pathway of glycogen synthesis operates in Waddliaceae and Criblamydiaceae.

GlgC was lost in the common ancestor of Waddliaceae and Criblamydiaceae. GlgA was demonstrated to be inactive in *Estrella lausanensis*. However abundant accumulation of glycogen particles is evidenced in both *Waddlia chondrophila* and *Estrella lausanensis.* The diversity of glycogen metabolism networks has been recently reported in gamma proteobacteria (Almagro et al., 2015). In this elegant study the authors outline the frequent loss of GlgC in a surprisingly high number of proteobacterial clades. Relevant to this study, is the reported presence of GlgB GlgX and MalQ in a single operon sometimes accompanied by TreZ in several β-proteobacterial clades. MalQ encodes a diversity of α-1,4 glucanotransferases. These enzymes, including the Chlamydiales MalQ reported in this work, are able to use MOS both as acceptors and donors for transfers reaction involving 2 or more glucose result. The net result of enzyme activity is the production of progressively longer MOS at the expense of shortening of MOS donors and often at the final expense of glucose production. When MOS length reaches the minimal size for branching enzyme, the latter will incorporate it in a growing glycogen particle. Provided MOS may be supplied by the external medium, which is probably the case for Chlamydiales infecting glycogen accumulating eukaryotes, this pathway bypasses the requirement of glucose activation by ATP into glycosyl-nucleotides and is thus very energy efficient. The presence of maltodextrin phosphorylase, because it recesses the MOS back to maltotetraose may prevent this pathway for operating. Phosphorylase in *Waddlia chondrophila* and *Estrella lausanensis* behaves as both a glycogen and maltodextrin type of phosphorylase. This provides the dual advantage of providing breakdown products of glycogen in the form of hexose-P (rather than free glucose) and of enabling the synthesis of short MOS in the presence of low orthophosphate conditions due to the readily reversible mode of action of glycogen/maltodextrin phosphorylases. Nevertheless, this is expected to severely limit the length of the chains incorporated into glycogen. While no experimental study exists that confirms that glycogen is being synthesized through this pathway in the aforementioned β-proteobacteria, we believe that mutant work performed in *E. coli* strongly suggests that this is indeed the case. In a recent study, *E. coli* mutants defective for both GlgC and GlgA have been produced that also lack MalP but still contain MalQ and glgB (Park et al., 2011; Kwak et al., 2016). In the presence of maltose in the external medium, these mutants sustained abundant glycogen synthesis. If MalP was

reintroduced in these strains, no glycogen accumulation was observed. We thus believe that *Waddlia chondrophila* and *Estrella lausanensis* both use this pathway for glycogen accumulation within the bacteria, the GlgA-GlgB fusion of Waddlia being only used as an effector enzyme.

References

Almagro G, Viale AM, Montero M, Rahimpour M, Munoz FJ, Baroja-Fernandez E, Bahaji A, Zuniga M, Gonzalez-Candelas F, Pozueta-Romero J (2015) Comparative genomic and phylogenetic analyses of Gammaproteobacterial glg genes traced the origin of the *Escherichia coli* glycogen glgBXCAP operon to the last common ancestor of the sister orders Enterobacteriales and Pasteurellales. PLoS One **10:** e0115516

Arias MC, Danchin EGJ, Coutinho P, Henrissat B, Ball S. (2012) Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants. Mobile genetic elements **2**:81-87

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

Ball SG, Bhattacharya D, Weber AP (2016) EVOLUTION. Pathogen to powerhouse. Science **351:** 659-660

Ball SG, Colleoni C, Arias MC (2015) The transition from glycogen to starch metabolism in cyanobacteria and eukaryotes. *In* Y Nakamura, ed, Starch: metabolism and structure. Springer, pp 93-158

Ball SG, Colleoni C, Kadouche D, Ducatez M, Arias MC, Tirtiaux C (2015) Toward an understanding of the function of Chlamydiales in plastid endosymbiosis. Biochim Biophys Acta

Bertelli C, Aeby S, Chassot B, Clulow J, Hilfiker O, Rappo S, Ritzmann S, Schumacher P, Terrettaz C, Benaglio P, Falquet L, Farinelli L, Gharib WH, Goesmann A, Harshman K, Linke B, Miyazaki R, Rivolta C, Robinson-Rechavi M, van der Meer JR, Greub G (2015) Sequencing and characterizing the genome of Estrella lausannensis as an undergraduate project: training students and biological insights. Front Microbiol **6:** 101

Bertelli C, Collyn F, Croxatto A, Ruckert C, Polkinghorne A, Kebbi-Beghdadi C, Goesmann A, Vaughan L, Greub G (2010) The Waddlia genome: a window into chlamydial biology. PLoS One **5:** e10890

Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell **60:** 115-122

Brown DH, Illingworth B, Cori CF (1961) The Mechanism of the De Novo Synthesis of Polysaccharide by Phosphorylase. Proc Natl Acad Sci U S A **47:** 479-485

Brushia RJ, Walsh DA (1999) Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure. Front Biosci **4:** D618-641

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

Cenci U, Nitschke F, Steup M, Minassian BA, Colleoni C, Ball SG (2014) Transition from glycogen to starch metabolism in Archaeplastida. Trends Plant Sci **19:** 18-28

Colleoni C, Linka M, Deschamps P, Handford MG, Dupree P, Weber AP, Ball SG (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

Cori CF, Cori GT, Green AA (1943) Crystalline muscle phosphorylase III Kinetics. J. Biol. Chem.**:** 33-55

Cori GT, Cori CF (1943) Crystalline muscle phosphorylase IV. Formation of glycogen. J. Biol. Chem.**:** 57-63

Cori GT, Larner J (1951) Action of amylo-1,6-glucosidase and phosphorylase on glycogen and amylopectin. J Biol Chem **188:** 17-29

Facchinelli F, Colleoni C, Ball SG, Weber AP (2013) Chlamydia, cyanobiont, or host: who was on top in the menage a trois? Trends Plant Sci **18:** 673-679

Facchinelli F, Pribil M, Oster U, Ebert NJ, Bhattacharya D, Leister D, Weber AP (2013) Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis. Planta **237:** 637-651

Facon M, Lin Q, Azzaz AM, Hennen-Bierwagen TA, Myers AM, Putaux JL, Roussel X, D'Hulst C, Wattebled F (2013) Distinct functional properties of isoamylase-type starch debranching enzymes in monocot and dicot leaves. Plant Physiol **163:** 1363-1375

Fischer EH (2013) Cellular regulation by protein phosphorylation. Biochem Biophys Res Commun **430:** 865-867

Fischer EH, Krebs EG (1955) Conversion of phosphorylase b to phosphorylase a in muscle extracts. J Biol Chem **216:** 121-132

Gehre L, Gorgette O, Perrinet S, Prevost MC, Ducatez M, Giebel AM, Nelson DE, Ball SG, Subtil A (2016) Sequestration of host metabolism by an intracellular pathogen. Elife **5**

Henrissat B, Deleury E, Coutinho PM (2002) Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? Trends Genet **18:** 437-440

Illingworth B, Brown DH, Cori CF (1961) The de novo synthesis of polysaccharide by phosphorylase. Proc Natl Acad Sci U S A **47:** 469-478

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

Jones TH, Wright BE (1970) Partial purification and characterization of glycogen phosphorylase from Dictyostelium discoideum. J Bacteriol **104:** 754-761

Krebs EG, Fischer EH (1955) Phosphorylase activity of skeletal muscle extracts. J Biol Chem **216:** 113-120

Kwak JY, Kim MG, Kim YW, Ban HS, Won MS, Park JT, Park KH (2016) Properties of a glycogen like polysaccharide produced by a mutant of *Escherichia coli* lacking glycogen synthase and maltodextrin phosphorylase. Carbohydr Polym **136:** 649-655

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

Melendez R, Melendez-Hevia E, Cascante M (1997) How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. J Mol Evol **45:** 446-455

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. Plant Cell **8:** 1353-1366

Nielsen TJ, Pradhan P, Brittingham A, Wilson WA (2012) Glycogen accumulation and degradation by the trichomonads Trichomonas vaginalis and Trichomonas tenax. J Eukaryot Microbiol **59:** 359-366

Park JT, Shim JH, Tran PL, Hong IH, Yong HU, Oktavina EF, Nguyen HD, Kim JW, Lee TS, Park SH, Boos W, Park KH (2011) Role of maltose enzymes in glycogen synthesis by *Escherichia coli*. J Bacteriol **193:** 2517-2526

Rall TW, Sutherland EW, Wosilait WD (1956) The relationship of epinephrine and glucagon to liver phosphorylase. III. Reactivation of liver phosphorylase in slices and in extracts. J Biol Chem **218:** 483-495

Rathore RS, Garg N, Garg S, Kumar A (2009) Starch phosphorylase: role in starch metabolism and biotechnological applications. Crit Rev Biotechnol **29:** 214-224

Roach PJ (2015) Glycogen phosphorylation and Lafora disease. Mol Aspects Med **46:** 78-84

Schinzel R, Nidetzky B (1999) Bacterial alpha-glucan phosphorylases. FEMS microbiology letters **171:** 73-79

Shepherd D, Segel IH (1969) Glycogen phosphorylase of Neurospora crassa. Arch Biochem Biophys **131:** 609-620

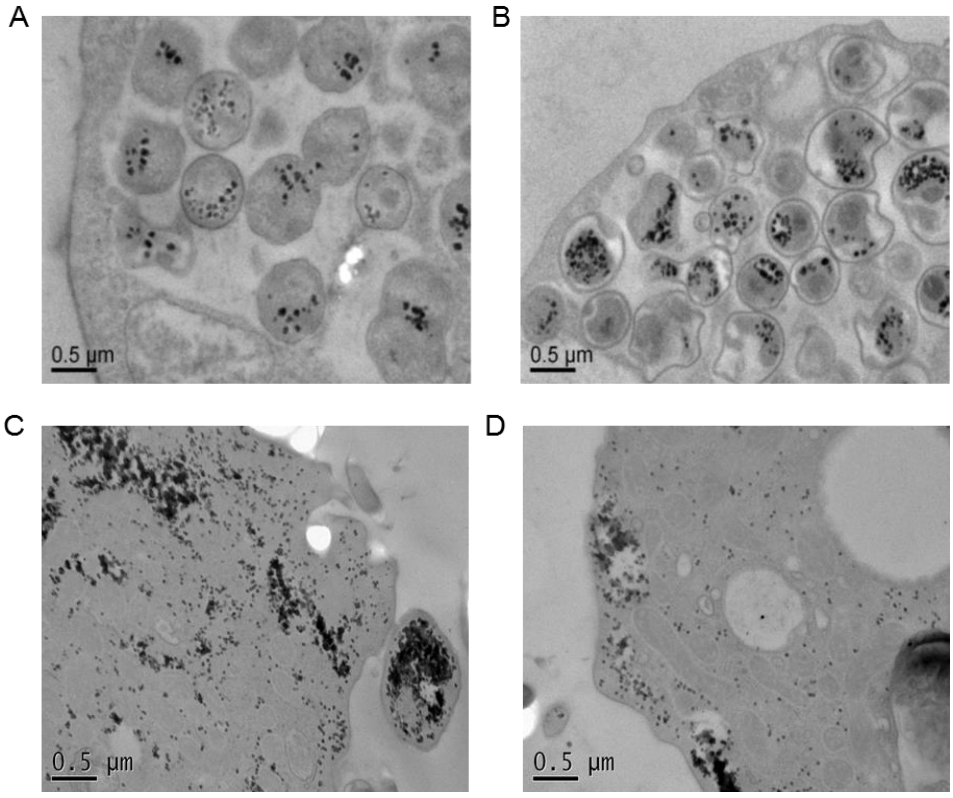
Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng Des Sel **19:** 555-562

Subtil A, Parsot C, Dautry-Varsat A (2001) Secretion of predicted Inc proteins of Chlamydia pneumoniae by a heterologous type III machinery. Mol Microbiol **39:** 792-800

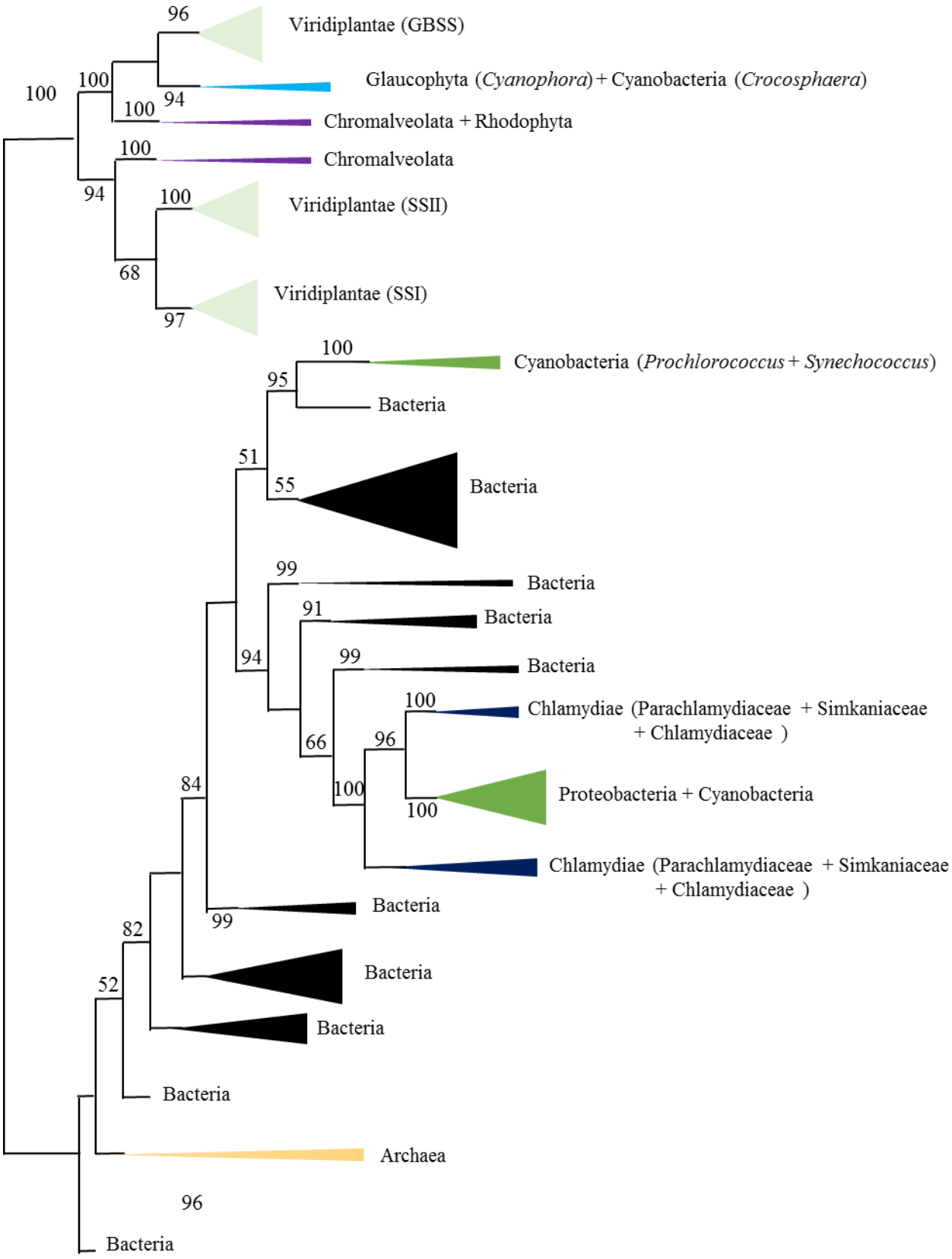
Wilson WA, Roach PJ, Montero M, Baroja-Fernandez E, Munoz FJ, Eydallin G, Viale AM, Pozueta-Romero J (2010) Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev **34:** 952-985

Wu G, Muller M (2003) Glycogen phosphorylase sequences from the amitochondriate protists, Trichomonas vaginalis, Mastigamoeba balamuthi, Entamoeba histolytica and Giardia intestinalis. J Eukaryot Microbiol **50:** 366-372

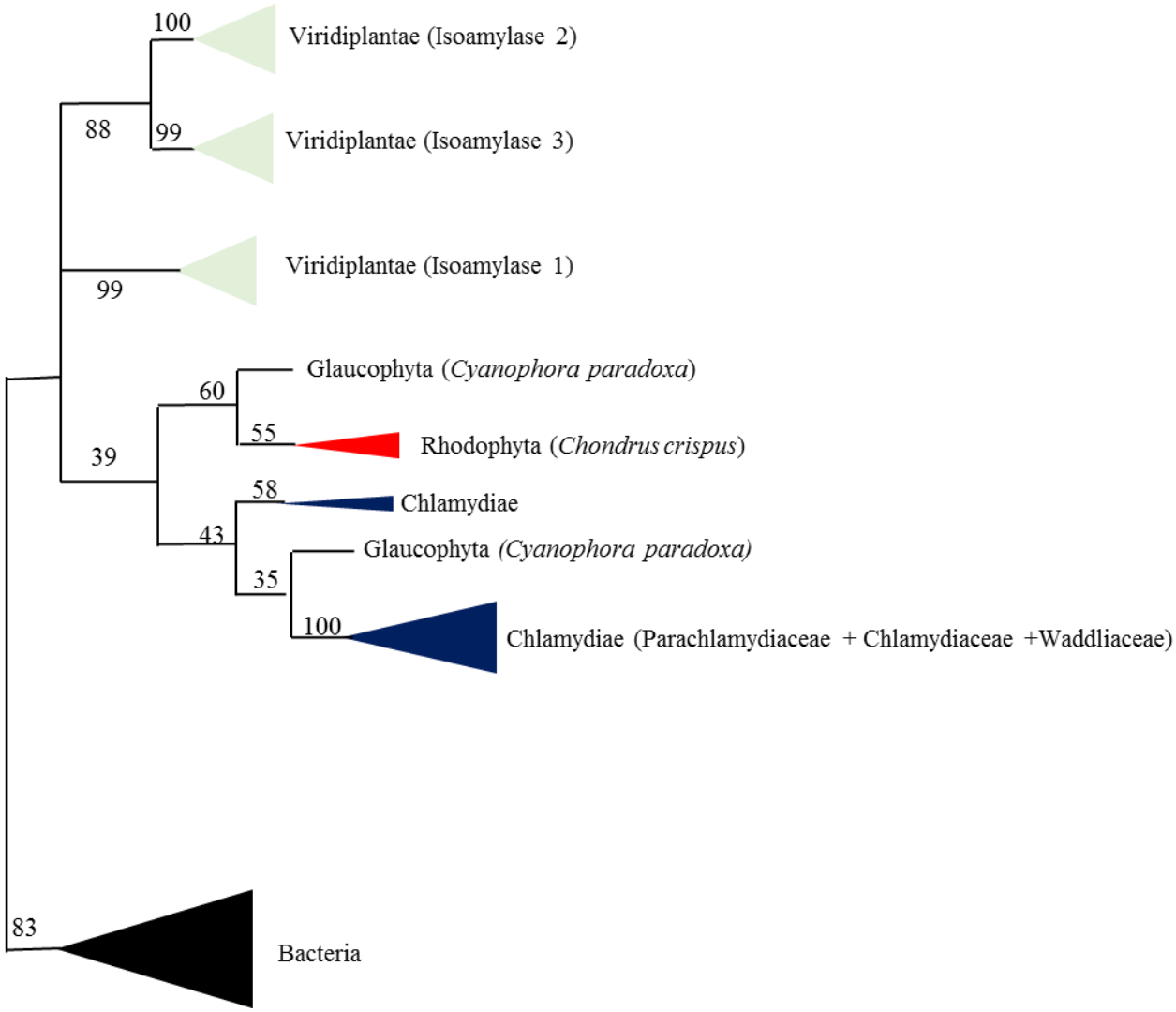
SUPPLEMENTARY FIGURES



Supplementary Figure 1: Accumulation of glycogen in infected *Acanthamoeba castellaniin* glycogen was observed by TEM (Transmission Electron Microscopy) after PATAg staining in *Parachlamydia acanthamoebae* after 16h (A) and 24h of infection (B), *Simkania negevensis* (C) and*Protochlamydia amoebophila* (D).Glycogen particles (balck dots) are accumulated within the bacteria for *Parachlamydia acanthamoebae* and no glycogen particles were detected within the bacteria or inclusion for *Simkania negevensis* and*Protochlamydia amoebophila*.



Supplementary Figure 2: Schematic representation of maximum likehood phylogeny with bootstraps values mapped into the branches for glycogen/starch synthase based on Ball et al, 2013.



Supplementary figure 3: Schematic representation of maximum likehood phylogeny with bootstraps values mapped into the branches for glycogen/starch debranching enzyme based on Ball et al, 2013.

PART II: Investigating storage polysaccharide metabolism evolution in Cyanobacteria

PART II: Investigating storage polysaccharide metabolism evolution in Cyanobacteria

Foreword:

In order to understand the evolution of glycogen and starch synthesis my lab became interested in cyanobacteria producing both glycogen and starch. When I started my internship in this lab I worked on *Cyanobacterium.*sp CLg1 which had been axenized in our laboratory, I worked on the 187G11 mutant, this mutant was generated after a UV mutagenesis campaign performed by Catherine Tirtiaux. During four years, she led mutagenesis campaigns and sub-cloning and screening cycles to select a collection of 90 mutants by iodine vapor selection results. Results on glycogen overptoducing mutants had been published in the plant cell by Ugo Cenci, who showed that they all carried mutations in the *GlgX2* gene (Ugo et al 2013). GlgX2 is an isoamylase that was shown to play an important role in starch synthesis analogous to that of plant Isa1 isoamylase. Nevertheless Isa1 did not evolve from GlgX2 after plastid endosymbiosis, thereby proving that starch evolved independently through convergent evolution in cyanobacteria and plants.

Results presented in this section concern the 187G11 mutant of *Cyanobacterium*. sp CLg1 selected by Catherine Tirtiaux and which displayed the same level of water soluble polysaccharides (WSP) than the wild type but completely lacks starch. Hence this strain laclked starch but did not overproduce glycogen.

The 187G11 mutant was initially selected by Catherine Tirtiaux. Dr. Ugo Cenci initially studied the purified mutant through activity gels and the first characterization of the WSP produced by the mutant. The transmission electronic microscopy was performed by Jean Luc Putaux. The phylogenetic trees were performed by Ugo Cenci and Maria Cecilia Arias.

The characterization of the 187G11 mutant was done by myself and Mathieu Ducatez. The activity gels for all the starch metabolism enzymes and the sequencing of the starch metabolism enzyme genes within the mutant were done by Mathieu.

The sequencing of the glgA2 gene of the mutant, recombinant protein characterization, complementation experiments, all the western blot analysis, the glgX2 activity gels, the *Synechosystis* mutant strain characterization, the priming activity zymogram gels, growth experiments of the wild type and the 187G11 mutant, mutagenesis of the conserved motif

YxxxY in *E.coli* and the characterization of the mutants and the structural analysis of the glycogen accumulated by *E.coli* were done by myself.

Results:

Characterization of function of the GlgA2 glycogen/starch synthase in *Cyanobacterium* sp. Clg1 highlights convergent evolution of glycogen metabolism into starch granule aggregation.

Derifa Kadouche[a2], Mathieu Ducatez[a2], Ugo Cenci[a], Catherine Tirtiaux[a], Eiji Suzuki[b], Yasunori Nakamura[b], Jean-Luc Putaux[c], Amandine Durand Terrasson[c], Sandra Diaz-Troya[d], Francisco JavierFlorencio[d], Maria Cecilia Arias[a], Alexander Striebeck[e] , Monica Palcic[f], Steven G. Ball[a,1] and Christophe Colleoni[a,1]

[a] Univ. Lille, CNRS, INRA, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France
[b] Department of Biological Production, Akita Prefectural University, 241-438, Shimoshinjo Nakano aza Kaidohata Nishi, Akita, 010-0195 Japan
[c] Centre de Recherches sur Les Macromolécules Végétales (CERMAV-CNRS), Université Grenoble Alpes, BP 53, F-38041 Grenoble Cedex 9, France
[d]Instituto de Bioquímica Vegetal y Fotosíntesis cicCartuja, CSIC-Universidad de Sevilla, Avda. Américo Vespucio, 49 41092 Sevilla, Spain
[e] Raw Materials Group, Carlsberg Laboratory, Gamle Carlsberg Vej 10, 1799 Copenhagen V, Denmark
[f]The Department of Biochemistry and Microbiology, University of Victoria, Victoria BC, Canada V8W 3P6
Running title: Evolution of starch metabolism.
[1] Co-corresponding authors. Christophe Colleoni and Steven G. Ball and, UMR8576 CNRS-USTL, Bâtiment C9, Université des Sciences et Technologies de Lille, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France. Phone: 33 3 20436543, Fax : 33 3 20436555, e-mail : steven.ball@univ-lille1.fr ; christophe.colleoni@univ-lille1.fr

[2]These authors contributed equally to this work

Short title: starch synthesis in cyanobacteria

One sentence summary: In starch accumulating cyanobacteria, the GlgA2/SSIII/SSIV enzyme is mandatory to obtain polysaccharide aggregation into amylopectin.

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors is: Christophe Colleoni (christophe.colleoni@univ-lille1.fr).

At variance with the starch accumulating plants and most of the glycogen-accumulating cyanobacteria, *Cyanobacterium* sp. CLg1 synthesizes both glycogen and starch. We now report the selection of a starchless mutant of this cyanobacterium that retains wild-type amounts of glycogen. Unlike other mutants of this type found in plants and cyanobacteria, the mutant proved to be selectively defective for one of the two types of glycogen/starch synthase: GlgA2. This enzyme is phylogenetically related to the previously reported SSIII/SSIV starch synthase that is thought to be involved in starch granule-seeding in plants. This suggests that in addition to the selective polysaccharide debranching demonstrated to be responsible for starch rather than glycogen synthesis, the nature and properties of the elongation enzyme defines a novel determinant of starch versus glycogen accumulation. We show that the phylogenies of GlgA2 and of 16S rRNA display significant congruence. This suggests that this enzyme evolved together with cyanobacteria when they diversified over 2 billion years ago. . However cyanobacteria can be ruled out as direct progenitors of the SSIII-IV ancestral gene found in Archaeplastida. Hence both cyanobacteria and plants recruited similar enzymes independently to perform analogous tasks, further emphasizing the importance of convergent evolution in the appearance of starch from a pre-existing glycogen metabolism network.

Introduction

Soluble glycogen/starch synthases of the GT5 (CAZy family 5 glycosyl- transferases) family transfer glucose from a nucleotide sugar to the non-reducing end of a growing α-1,4-linked glucan. Among the very large family of prokaryotic GT5 enzymes, the soluble starch synthases III/IV (SSIII/IV) found in the green plant or alga plastid and in the glaucophyte cytosol are united into a highly supported monophyletic group together with glycogen/starch synthases found in all chlamydiales intracellular pathogens, in a restricted number of proteobacteria and a large number of cyanobacteria (Ball et al., 2013). However, erosion of phylogenetic signal did not enable a clear determination of the root position of this SSIII/IV/GlgA2 sub-family within the large size GT5 GS phylogenetic tree (Ball et al, 2013). SSIII/IV in green plants and algae are known to be essential for starch synthesis and play both a role in building the large-size chains within amylopectin and a function in starch particle seeding and (or) polysaccharide synthesis priming (for reviews see D'Hulst *et al.*, 2015; Nakamura, 2015). Little is known about the function of the corresponding enzymes in bacteria. Cyanobacteria represent one of the most ancient groups of prokaryotes and the founders of oxygenic photosynthesis (Summons et al., 1999, Crowe et al., 2013). Like in plants, photosynthetic carbon is temporarily assimilated via the Calvin cycle in the form of homopolymers of D-glucose, such as glycogen or starch that both consist of glucan chains made of glucose residues linked in α-1,4 and branched by α-1,6 linkages. In spite of sharing the same chemical linkages, both polymers widely differ in physicochemical properties. Glycogen particles are highly branched polysaccharides (8 to 10 % of α-1,6 branches) resulting in the storage of small hydrosoluble particles with 30-50 nm maximal diameter in the cytosol of numerous organisms (Archaea, Bacteria and eukaryotes). One third of a maximum total 55000 glucose residues within a single particle is readily accessible to glycogen catabolism in the outer chains without cleaving off α-1,6 branches (Melendez-Hevia et al., 1993). Glycogen is thus a homogeneous structure and a very dynamic form of glucose

storage that combines low osmotic activity and accessibility to hydrosoluble enzymes. Starch granules are usually made up of two α-glucan polymers, namely amylopectin and amylose. The minor fraction, amylose, is composed of linear weakly-branched glucan chains (<1% of α-1,6 branches) while the major fraction, amylopectin, harbors an ordered branch pattern of α-1,6 linkages leading to the cluster organization responsible for starch crystallinity (Hizukuri, 1986, Laohaphatanaleart et al., 2010, Bertoft et al., 2010). The synthesis of starch granules was initially believed to be a hallmark of three sister lineages -plants/green algae, red algae and Glaucophytes (i.e Archaeplastida) stemming from primary plastid endosymbiosis and some of their secondary endosymbiosis derivatives (i.e. Alveolates and Cryptophytes) (Cenci et al., 2014; Ball et al., 2015). Several lines of evidence suggest that starch metabolism has evolved shortly after plastid endosymbiosis from a pre-existing cytosolic eukaryotic glycogen metabolism enzyme network. In line with this hypothesis, an overview of gene origin in Archaeplastida lineages points out that most of the starch metabolism enzymes display a common host phylogeny. Only 4 genes of the inferred ancestral Archaeplastida network display a clear-cut bacterial origin with two originating from cyanobacteria (GBSS (granule-bound starch synthase), and ADP-glucose pyrophosphorylase (GlgC)) and the remainder two from chlamydial intracellular pathogens (GlgA and GlgX respectively soluble glycogen/starch synthase and glycogen/starch debranching enzyme) (Ball et al., 2013). Interestingly, extant unicellular diazotrophic cyanobacteria were recently reported to synthesize starch-like polysaccharides with an enzyme network mostly unrelated to the one at work in Archaeplastida (Cenci et al., 2013). The presence of GBSS in chroococcales unicellular diazotrophic cyanobacteria may suggest that the plastid ancestor could have been an ancient starch accumulator related to such organisms (Cenci et al., 2013). Indeed GBSS is an enzyme responsible for amylose synthesis within starch and requires the binding to semi-crystalline polysaccharides to be active. We thus proposed that an ancestor of this group of diazotrophic unicellular cyanobacteria may define the plastid donor (Deschamps et al., 2008).

Because of the fastidious growth of many chroococcales unicellular diazotrophic cyanobacteria and because this group, like many other cyanobacteria, has resisted all attempts at genetic transformation, we applied a classical genetic approach to the dissection of starch metabolism in *Cyanobacterium* sp. CLg1. This strain initially reported as diazotrophic by Falcon et al. (2004) was axenized by us, but has lost the ability to fix nitrogen under laboratory conditions. *Cyanobacterium* sp. CLg1, has been reported to accumulate both a major starch fraction and a minor yet significant glycogen pool (Falcon et al., 2004; Cenci et al., 2013). We now report the selection of a starchless mutant of Cyanobacterium that synthesizes wild-type amounts of glycogen. This mutant proved to be selectively defective for the GlgA2 glycogen/starch synthase. This suggests that starch and glycogen are synthesized by at least partly distinct pathways in *Cyanobacterium* sp. CLg1. To our knowledge this is the first report for a requirement other than those previously assigned to starch debranching enzymes for the selective accumulation of starch rather than glycogen in living cells. The evolutionary implications of this novel function are discussed in the light of the origin and possible role of the SSIII-IV-GlgA family of glucan elongation enzymes within cyanobacteria.

Results

Selection of 187G11 a starchless mutant of *Cyanobacterium sp.* CLg1

A collection of 2.10^4 mutants was generated after UV mutagenesis followed by a minimum of 4 rounds of subcloning as previously detailed in Cenci et al. (2013). Following the first round of screening, the selected mutants were further subcloned to check for complete segregation of the mutant phenotype. After 3 years of segregation and phenotype screening we selected 7 strains defining the class C mutants, which contained water-soluble polysaccharides in amounts close to those of the wild-type reference but with significantly lower amounts of starch. 6 of these 7 mutants were previously reported in Cenci et al. (2013) but failed to reveal the biochemical explanation for the mutant phenotype. The seventh strain (187G11) displayed a very severe phenotype defined by the absence of iodine stain displayed after spraying cell patches with iodine vapors. This was correlated to a complete disappearance of starch, which fell below detection level (<0.5% of wild-type level) (Fig. 1). However the mutant remained able to accumulate normal amount of water soluble polysaccharide (0.34 ± 0.04 mg of WSP.mg^{-1} of protein) in comparison to the wild-type strain (0.26 ± 0.04 mg of WSP.mg^{-1} of protein). This phenotype is more severe than that exhibited by class A mutants which overproduced glycogen and retained very low levels (2-5% of wild type) of starch with modified structure (Cenci et al., 2013). Nevertheless unlike glycogen-less mutants of Synechocystis PCC6803 the mutant grew under 12h light/12 h darkness growth conditions albeit with a twofold increase in generation time (from 60 to 120 h) (Grundel et al., 2012) (Supplemental Fig. S1) in liquid medium. On solid media we did not observe significant delays in the appearance of single colonies.

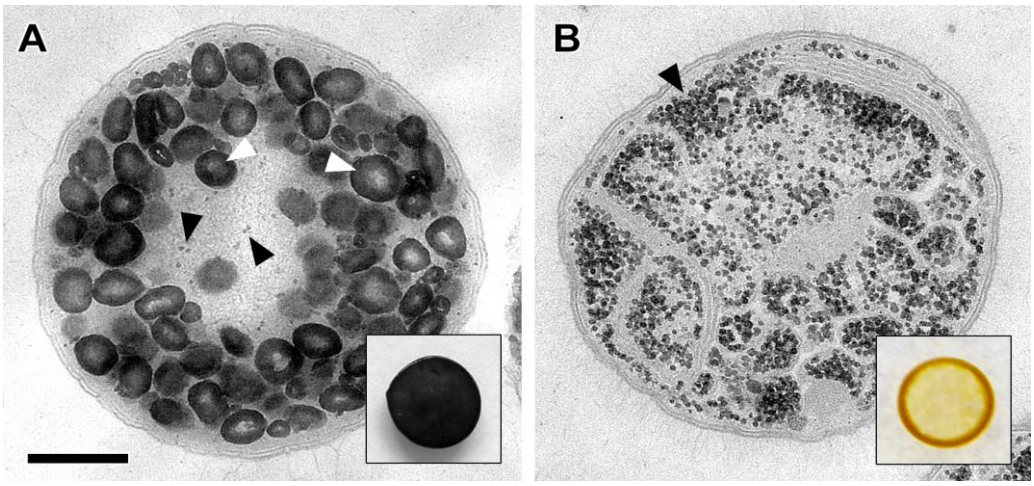


Figure 1. TEM images of ultrathin sections (70 nm) of wild type and 187G11 mutant strains. Polysaccharides in the wild-type CLg1 **(A)** and 187G11 mutant **(B)** are positively stained with PATAg. Both starch-like granules (white arrows) and glycogen particles (dots pointed by black arrows) were observed in the wild-type strain. Starch granules are absent in the 187G11 mutant and substituted by glycogen-like water-soluble polysaccharides (black arrow). The dark-blue iodine stain from a cell patch of the wild-type strain is shown in inset. The absence of starch granules in the 187G11 mutant yields a yellow-orange stain after spraying iodine vapors (inset in panel B). Bar: 500 nm

187G11 displays normal glycogen levels of slightly modified structure.

To characterize the water-soluble polysaccharide fraction accumulated in the 187G11 mutant, the latter was purified, sized by gel permeation chromatography and compared to wild type (Fig. 2A and 2B). Both mutant and wild type soluble polysaccharides are composed of high molecular weight polysaccharides (fractions 35 to 50) and short malto oligosaccharides (fractions 60 to 100). The formers were then examined by transmission electron microscopy (TEM) and further subjected to enzymatic debranching, followed by separation of chains by HPAEC-PAD (High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection). Chain length distribution analysis (Fig. 2C and 2D) and TEM observation of negatively stained preparations (Fig. 2E and 2F) suggest that WSP of 187G11 is composed of highly branched glucan chains capable to exclude the uranyl acetate molecules in a fashion similar to wild-type soluble polysaccharides (Fig. 2E). Altogether, these results (Fig.1 and Fig. 2 panels A to D) suggest that 187G11 contains normal amounts of of a similar although not identical branched polysaccharide with a chain-length distribution slightly enriched in small chains by comparison to the wild-type. Hence 187G11 synthesizes glycogen as efficiently as wild-type cells but selectively lacks starch.

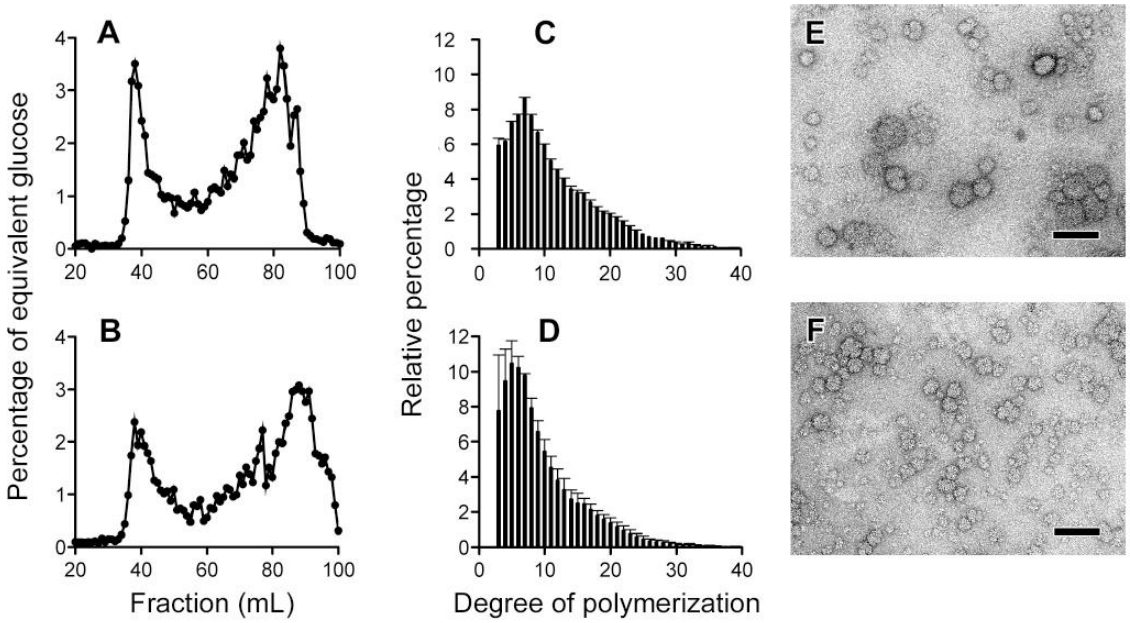


Figure 2. Structural analysis of water-soluble polysaccharide accumulated by the wild type and 187G11 mutant strains. Water-soluble polysaccharides purified from wild-type **(A)** and 187G11 mutant **(B)** strains were subjected to size exclusion chromatography analysis (TSKHW55 Toyopearl). The amount of total glucose was determined for each fraction by the phenol-sulfuric method (see methods). Results are expressed as weight percentages of equivalent glucose (black line). After complete digestion with commercial isoamylase, glucan chains were separated according to their degree of polymerization (DP) by HPAEC-PAD. The relative abundance for each DP (black bars) was determined for the wild type **(C)** and 187G11 mutant **(D)** from the mean of three independent extractions. TEM images of negatively stained preparations suggest that WSP of the wild type **(E)** and 187G11 mutant **(F)** are highly branched polysaccharides with a diameter below 50 nm similar to glycogen particles of rabbit liver (bars: 100 nm)

187G11 is specifically defective for the major starch /glycogen synthase

We undertook a large survey of starch metabolism enzymes through crude extract assays (ADP-glucose pyrophosphorylase and glycogen/starch synthase) and previously adapted zymogram procedures (phosphorylases, glycosyl hydrolases, and transferases including BE (branching enzyme), α-1,4 glucanotransferase, debranching enzymes, amylases, and glycogen/starch synthases) (Supplemental Fig. S2). We found a very large decrease in total glycogen primed glycogen/starch synthase activity (80% decrease with respect to wild-type (492 $nmol.min^{-1}.mg^{-1}$ of protein)) that correlated with the disappearance of the major glycogen/starch synthase. A second minor slow migrating glycogen/starch synthase was also witnessed selectively in the mutant 187G11 strain (Fig. 3). Unfortunately the unstable nature and low activity of this slow migrating band which we suspect to represent the GlgA2 mutant enzyme (see below) only allowed us to partially purify it from *Cyanobacterium* sp. CLg1 mutant crude extracts. We have previously published that mutants defective for a debranching enzyme (GlgX2) over-accumulated glycogen and witnessed a dramatic decrease in starch amounts. To make sure that the phenotype displayed in 187G11 could not result from a combination of a direct effect on the glycogen/starch synthase and an indirect effect on the GlgX2 debranching enzyme, we semi-quantified GlgX2 by zymogram analysis through the procedures detailed in Cenci et al. (2013) and found the activity to be normal qualitatively and quantitatively (Supplemental Fig. S3).

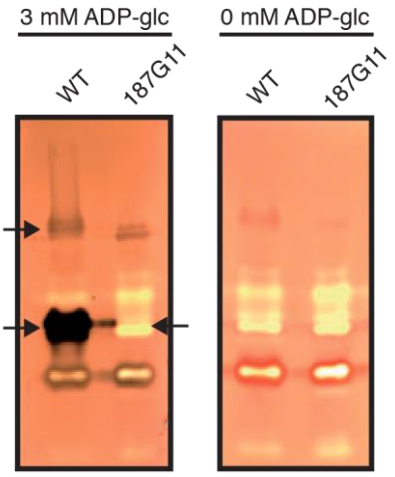


Figure 3. Zymogram analysis of glycogen/starch synthase activities from wild type and 187G11 mutants. Total protein of semi-purified crude extracts of both WT and 187G11 mutant strains were separated by native PAGE, then electro-transferred onto native PAGE containing 0.6% (p/v) of glycogen. The native gels were then incubated with or without 3 mM ADP-glucose. Glycogen/starch synthase activities are witnessed after iodine staining in the wild type's crude extract as two dark activity bands (black arrows). The fast migrating form disappears in the 187G11 mutant (black arrow). The total decrease in iodine staining was estimated through dilution to be between two to three orders of magnitude. This decrease is in line with that measured by quantitative radioactive assays in recombinant *E. coli* extracts.

Characterization of a glycogen/starch synthase mutation in the 187G11 genome

We proceeded to sequence all of the genes previously found in the *Cyanobacterium*sp. CLg1 genome related to either glycogen and (or) starch metabolism. These included all possible glycosyl hydrolases and glycosyl transferases found in the genome and known or suspected to be involved in glucan metabolism as listed in Table 1. In addition, we sequenced the unique gene encoding ADP-glucose pyrophosphorylase from the 187G11 strain. We found only one significant modification in the whole starch/glycogen metabolism network and no silent mutations. We thus found a one base pair deletion yielding a frameshift and a nonsense mutation toward the C-terminus of the GlgA2 glycogen synthase gene (Fig. 4). This mutation deletes a highly conserved region of the bacterial GT5 glycogen synthases and is therefore expected to impact the enzyme activity. We further investigated this impact on other GT5 glycogen synthases such as the *E. coli* enzyme by introducing mutations in one or both of the highly conserved tyrosine residues (Supplemental Fig. S4) of this region which resulted in a large decrease of the enzyme activity. Slow growth of the marine *Cyanobacterium* sp. CLg1 strain requires 4 years for a full cycle of mutant screening and purification which prevented us from selecting additional defective alleles. In addition chroococcales cyanobacteria are notorious for their resistance to genetic transformation precluding complementation of the effect by genetic transformation. Hence not only did we sequence all genes of the starch/glycogen metabolism network but in addition we assayed all possible enzymes of the network to check that undetected mutations in regulatory genes would not modify the balance of starch/glycogen metabolism enzymes. We found no evidence for any qualitative modification in all assayable enzyme activities through crude extract assays and zymogram procedures. However we did record a significant increase in starch (glycogen) phosphorylase activity which was also noted in other mutants of *Cyanobacterium* sp. CLg1. Similar increases have been noted in other cyanobacterial mutants by others (Fu and Xu, 2006; Cenci et al., 2013).

Activity	Gene	Cazy Classification	Accession N°	sequencing
ADP-glucose pyrophosphorylase	*glgC*	-	KR020055	+
glycogen/starch synthase	*glgA1*	GT5	AHB52787	+
glycogen/starch synthase	*glgA2*	GT5	AHB52788	K480N
glycogen/starch synthase	*gbss*	GT5	AHB52786	+
Branching enzyme	*glgB1*	GH13	AFP43334	+
Branching enzyme	*glgB2*	GH13	AFP43335	+
Branching enzyme	*glgB3*	GH13	AFP43336	+
Putative branching enzyme	*glgB4*	GH57	AHB52790	+
Debranching enzyme	*glgX1*	GH13	AGI19288	+
Debranching enzyme	*glgX2*	GH13	AGI19289	+
Debranching enzyme (amylopullulanase-GH13)	*apu13*	GH13	AHB52783	+
Putative debranching enzyme (amylopullulanase-GH57)	*apu57*	GH57	AHB52784	+
Debranching enzyme (amylo-1,6 glucosidase)	*amg*	GH133	AHB52785	+
Phosphorylase	*glgP*	GT35	AHB52789	+
α-1,4 glucanotransferase	*malQ*	GH77	AHB52791	+

Table 1: Summary of starch metabolism genes sequenced in the 187G11 mutant. Each gene was amplified using primers designed in the untranslated region. PCR products were cloned and sequenced on both strands using additional primers when required. GH and GT stand for Glycosyl Hydrolase and Glycosyl Transferase, respectively.

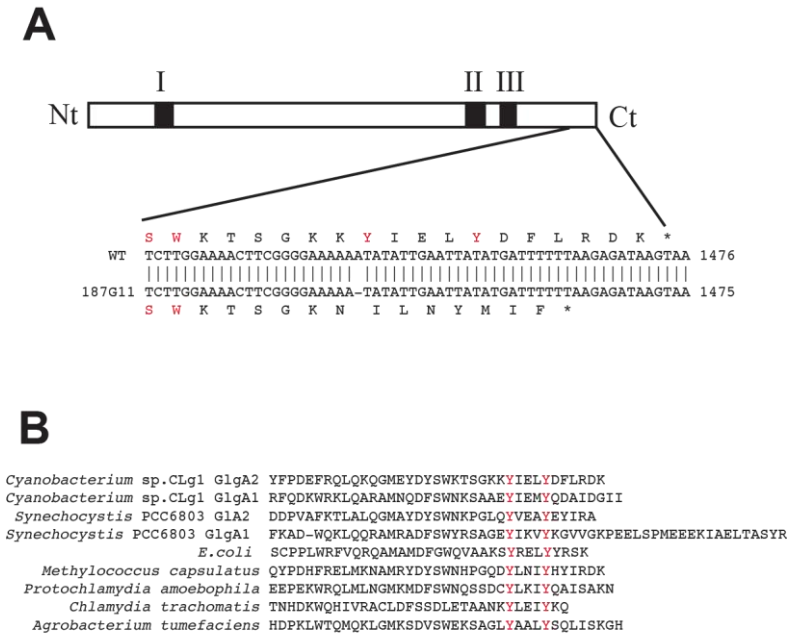


Figure 4. Molecular characterization of the starchless mutant 187G11. (A) One base pair deletion was identified in the glgA2 gene of 187G11. This point mutation results in a frameshift followed by appearance of a non-sense codon and the synthesis of truncated protein (GlgA2*) at the Carboxy terminus. Regions I, II and III (black boxes) previously characterized to be involved in the binding of ADP-glucose and in catalysis are conserved in GlgA2*. **(B)** Nevertheless, the YxxxY motif conserved throughout all bacterial GT5 glycosyl transferases has disappeared in GlgA2*.

Mutants of *Cyanobacterium* sp. CLg1 defective for GlgA2 display a phenocopy of *Synechocystis* PCC6803 strains disrupted for the GlgA2 gene

We compared the glycogen structures accumulated by the wild-type and mutant *Cyanobacterium* sp. CLg1 strains (Fig.2 panels C and D) to those of wild-type and mutant *Synechocystis* PCC6803 generated through targeted gene disruption of the Synechocystis GlgA2 structural gene (Fig.5A and B). Subtractive analyses of chain length distributions of *Cyanobacterium* sp .CLg1 (Fig. 5C) and Synechocystis (Fig. 5D) reveal that in both species modification or disappearance of GlgA2 yields the synthesis of normal amounts of glycogen with an increase of short chains of DP 4 to 9 and DP 7 to 15 and fewer long chains of DP 10 to 40 and DP 16 to 40, respectively. We believe this to strengthen our suggestion that the phenotype recorded in *Cyanobacterium* sp CLg1 resulted directly from the absence of normal GlgA2 activity. Indeed the Synechocystis mutant was generated by reverse genetics through selective disruption of the GlgA2 gene.

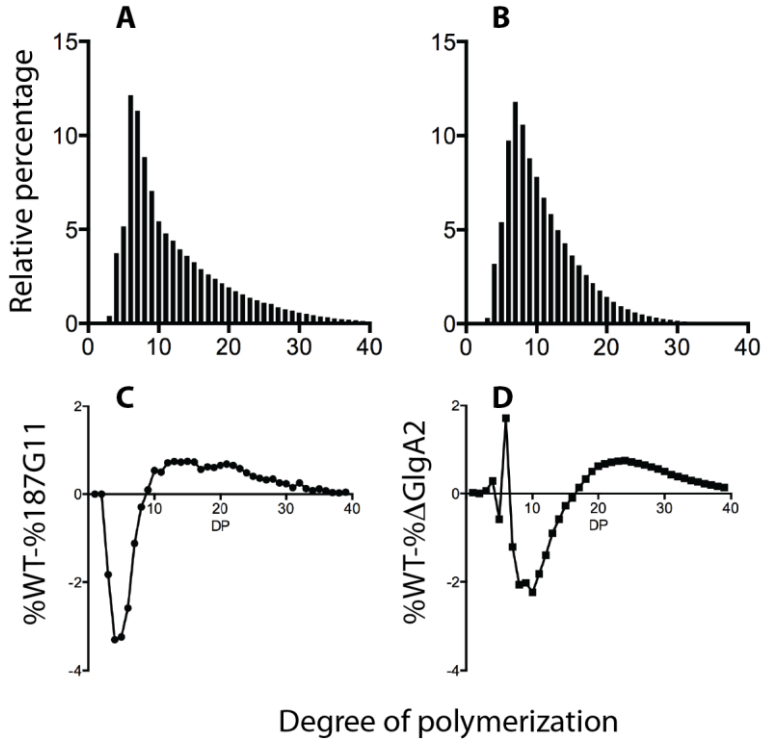
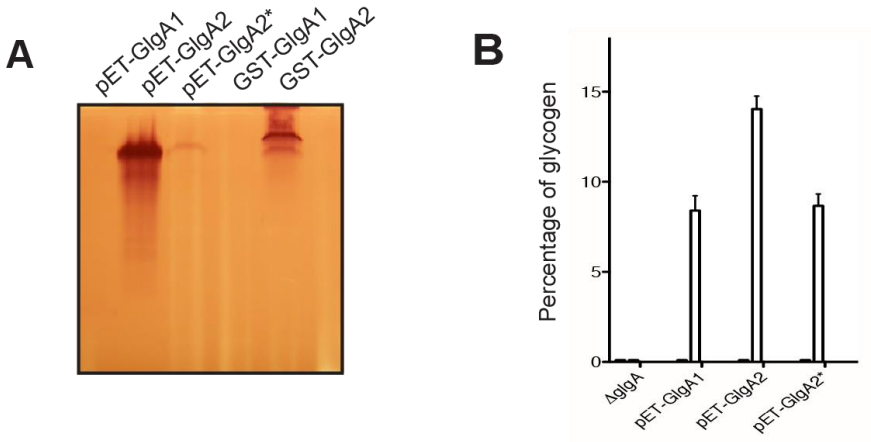


Figure 5. Structural analyses of water-soluble polysaccharide accumulated by the wild type and *GlgA2* mutant strains of *Synechocystis* PCC6803. Water-soluble polysaccharides purified from wildtype and the ΔglgA2 mutant strains were subjected to size exclusion chromatography analysis (TSKHW55 Toyopearl). After complete digestion with commercial isoamylase, glucan chains were separated according to their degree of polymerization (DP) by HPAEC-PAD. The relative abundance for each DP (black bars) was determined for the wild type **(A)** and the ΔglgA2 mutant **(B).** Subtractive analyses were performed between chain length distributions of wild-type and 187G11 strain of *Cyanobacterium* sp. CLg1 **(C)** and between wild-type and ΔglgA2 mutant strains of *Synechocystis* **(D).**

Biochemical characterization of glycogen/starch synthases in wild-type and mutant cyanobacteria

In order to verify that the nonsense mutation detected in *glgA2* explains both the disappearance of the major glycogen/starch synthase and the phenotype recorded in 187G11 we expressed wild-type GlgA1 and GlgA2 proteins as well as the mutant GlgA2* enzyme in *E. coli* (Fig. 6A). In addition we checked for complementation of the *E. coliglgA* mutation by our constructs. Interestingly the wild-type GlgA2 enzyme complemented the *E. coli* defect only when *E. coli* was supplemented with maltose and not with mannitol, a property which was shared by both GlgA1 and the mutant GlgA2* (Fig. 6B). All recombinant proteins cross-reacted in purified extracts as expected with antibodies directed against the phylogenetically related Synechocystis PCC 6803 GlgA1 and GlgA2 (named respectively GSII and GSI by Yoo et al. 2014) (Supplemental Fig. S5). However we were unable to distinguish the activities in crude extracts because of abundant cross reactions against other bacterial proteins. A

strongly decreased activity was scored for GlgA2* (0.303 µmol.min^{-1}.mg^{-1}) by comparison to the wild-type enzyme (327µmol.min^{-1}.mg^{-1}) while significant GlgA1 activities could be reproducibly measured only by quantitative radioactive assays (4.1 µmol.min^{-1}.mg^{-1}) (see methods). The absence of iodine stain in the strains expressing GlgA1 is suggestive of the synthesis of very short glucan chains (iodine staining of glucans starts developing at 20°C for chains longer than 12 Glc residues) (Fig. 6C). We also expressed GlgA1 and GlgA2 from Synechocystis PCC 6803. In a similar fashion, we found recombinant activity through iodine staining with GlgA2 but not with GlgA1, which thus behaved like the *Cyanobacterium* sp. CLg1 GlgA1. Finally we subjected the glycogen accumulated by *glgA* defective *E. coli* mutants complemented with either recombinant GlgA1 or GlgA2 to characterization of their CL distribution. Results displayed in supplemental Figure S6 did not yield any convincing difference between the two types of complemented strains



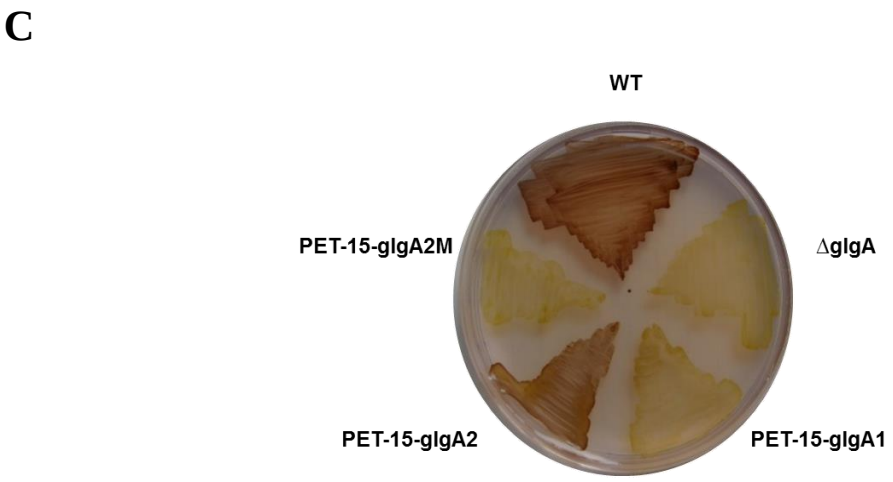


Figure 6. Complementation experiments and recombinant protein expressions of GlgA1 and GlgA2 of CLg1. (A) Recombinant protein expressions of GST tagged glycogen/starch synthase GlgA1and GlgA2 (pGEX-glgA1, pGEX-glgA2) and untagged proteins (pET-glgA1, pET-glgA2 and pET-glgA2*) were expressed in the ΔglgA mutant stain (JW3392-1) of *Escherichia coli*. Crude extracts were loaded on native PAGE containing glycogen. After migration, the gel was incubated overnight in the incubation buffer containing 3 mM ADP-glucose. Provided their glucan products are sufficiently long to form helices that stably trap iodine, glycogen/starch synthase activities may be revealed as black bands staining in an orange background after soaking the native gel in iodine solution. **(B)** Restoration of glycogen

synthesis in the presence of mannitol or maltose as carbon source in the ΔglgA mutant expressing untagged protein GlgA1, GlgA2 and GlgA2* (pET-glgA1, pET-glgA2 and pET-glgA2*). The amount of glycogen for each strain was determined by the amyloglucosidase assay. The glycogen measured in the wild-type strain was used as reference (70 ± 20 µg of glycogen/mg of protein). The results were expressed as percentages of glycogen amounts accumulated by our wild-type reference. **(C)** Iodine staining of the wild type strain (WT) and the ΔglgA strains of *Escherichia.coli* expressing the recombinant proteins GlgA1, GlgA2 and GlgA2* cultured in solid synthetic medium supplemented with 50mM of maltose.

GlgA2 is selectively bound to cyanobacterial starch and can prime polysaccharide synthesis in *Cyanobacterium* but not *E. coli.*

In a previous study, we found peptides from a 52 kDa starch bound protein that matched those from the *Cyanobacterium* sp. CLg1 GlgA2 (Fig. 7A) (Deschamps et al., 2008). We also found GlgA2 by similar means on the purified Cyanothece starch-like granules. We used the anti-GlgA1 and anti-GlgA2 antibodies raised against the Synechocystis enzymes that cross-reacted similarly with the corresponding GlgA1 and GlgA2 Cyanobacterium recombinant proteins to check for the presence of these proteins on the cyanobacterial starch granules. In both systems we found GlgA2 as a major starch bound protein with no GlgA1 detected (Fig. 7B).

The fact that the wild-type GlgA2 enzyme could complement a *glgA* defective *E. coli* mutant only in the presence of maltose suggested to us that this activity was dependent on the supply of MOS primers by the MalQ amylomaltase in *E coli*. This was confirmed by the absence of recombinant GlgA2 enzyme activity recorded on zymogram gels in the absence of glycogen primer (Fig. 8A). This property was shared also by the GlgA2* mutant activity. However when GlgA2 was purified partially from Cyanobacterium extracts the wild-type protein was always able to prime glucan synthesis in the absence of added polysaccharide primer (Fig. 8A). We conclude that in Cyanobacterium the GlgA2 glycogen synthase is either modified or interacts with a Cyanobacterium specific factor or substrate absent from both glucose or maltose grown *E. coli* cells. We do not know if GlgA2* would behave similarly in Cyanobacterium extracts since we never obtained enough residual activity in the mutant to assay its primer dependence. We therefore believe that the mutant phenotype could be explained either by the spectacular decrease of enzyme activity on its own or by a combination of the latter and a possible inability to prime polysaccharide synthesis in vivo in Cyanobacterium.

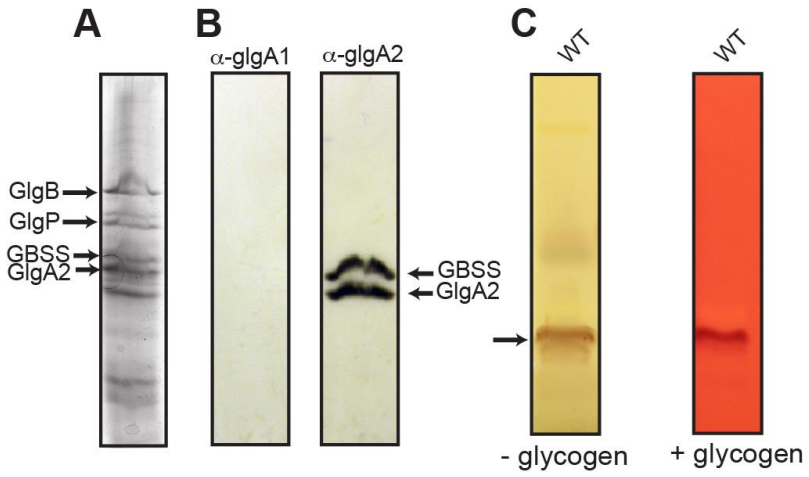


Figure 7. Western blot analysis of granule bound proteins and primer dependence of GlgA2 activity. (A) Proteins specifically bound to starch granules were analyzed onto SDS-PAGE. Major polypeptides were previously identified by mass spectrometry analysis (Deschamps et al., 2008): GlgB (85 kD branching enzyme); GlgP (72 kD glycogen phosphorylase); GBSS (57 kD granule bound starch synthase); GlgA2 (52 kD glycogen/starch synthase). **(B)** Western blot analysis was carried out on proteins attached to starch granules. Polypeptides were transferred onto PVDF membranes. Glycogen/starch synthases were immunodetected using antibodies raised against GlgA1 (α-GlgA1) and GlgA2 (α-GlgA2) of Synechocystis PCC6803. **(C)** Total proteins of the wild-type CLg1 (WT) strain were loaded and separated onto native-PAGE. After migration, the native PAGE gel was cut in two parts: one half of the gel was directly incubated in starch synthase buffer containing 3 mM of ADP-glucose (- glycogen) while the second half, was separately electro-transferred against another native PAGE containing glycogen (+ glycogen) during 2 hours. Starch synthase activity was revealed as black activity band with iodine solution after overnight of incubation (black arrow). Activities were too low to enable detection of GlgA2* in comparable experiments.

A Cyanobacterium protein is responsible for the modification of the ability to prime polysaccharide synthesis displayed by GlgA2

We further investigated the ability to prime polysaccharide synthesis and interact with glycogen of the wild-type recombinant and native Cyanobacterium GlgA2. First we compared the migration of recombinant and native enzymes on non denaturing PAGE gels (Fig. 8A and 8B). Both proteins co-migrated thereby suggesting that the differences between recombinant and wild-type enzymes were not due to extensive alterations or gross modification of the subunit structure (Fig. 8A). Nevertheless on glycogen containing zymogram gels both types of extracts yielded different migration patterns (Fig. 8B). The Cyanobacterium wild-type extracts generated a major slow high affinity zymogram band and a minor fast low affinity band. The recombinant enzyme displayed chiefly the fast migrating isoform (Fig. 8B). Hence the native enzyme differed by two criteria : affinity for glycogen and ability to prime polysaccharide synthesis. We attempted to understand the basis of these differences. We thus mixed Cyanobacterium extracts from the mutant strain 187G11 with defective GlgA2 activity with recombinant GlgA2 extracts. We then compared the mixed extracts with wild-type GlgA2 cyanobacterium extracts. We found that the mixing had transferred to the recombinant

enzyme the ability to prime polysaccharide synthesis but did not change significantly the ratio of fast to slow isoforms activities on glycogen containing gels (Fig. 9). The active element was sensitive to heat and to pre-treatment with proteases suggesting the presence of an active protein responsible for giving to GlgA2 it's primer independence within both wild-type and mutant extracts (Fig. 9A). In order to ascertain that the priming-inducing activity was not due to the presence of free malto-oligosaccharide primers, we mixed the recombinant enzyme with maltoheptaose (Fig. S7) and recovered no polysaccharide priming property after migration.

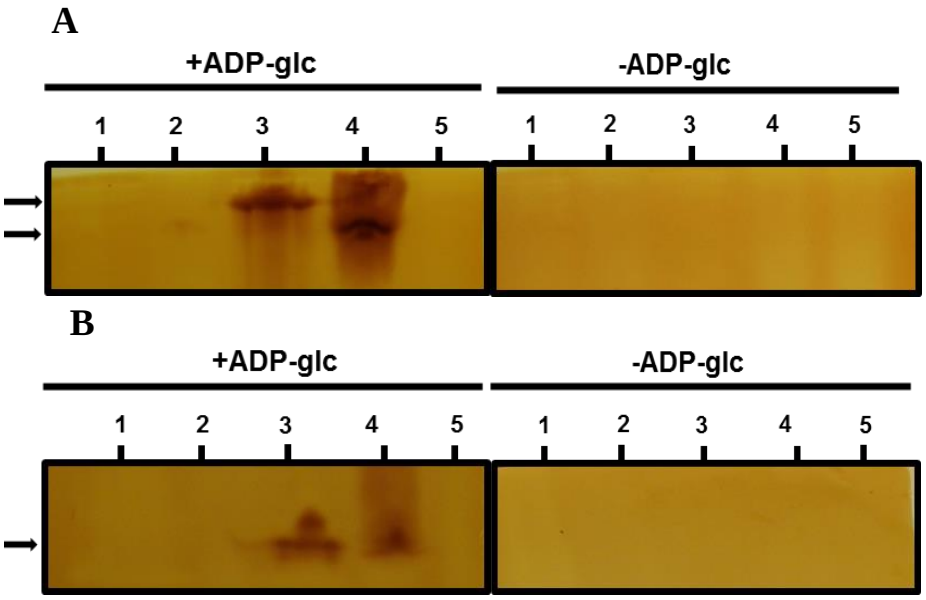


Figure 8. Zymogram analysis of glycogen/starch synthase activities. Total proteins of crude exctracts of the wild type strain (WT), the 187G11 mutant strain and the recombinant proteins GlgA1, GlgA2, GlgA2* were separated by native PAGE containing glycogen **(B)** or without glycogen then electro-transferred onto native PAGE containing 0.6% (p/v) of glycogen **(A).** The natives gels were then incubated with or without 3 mM ADP-glucose. Glycogen/starch synthase activities are witnessed after iodine staining as a dark activity bands. The wild type glycogen/starch synthase and the recombinant protein GlgA2 co-migrate when glycogen is absent in the zymogram gels **(A)**, while in presence of glycogen in the zymogram gels the wild type enzyme migrate slower then the recombinant protein GlgA2 **(B). 1:** 187G11, **2:** GlgA2*, **3:** WT, **4:** GlgA2, **5:** GlgA1.

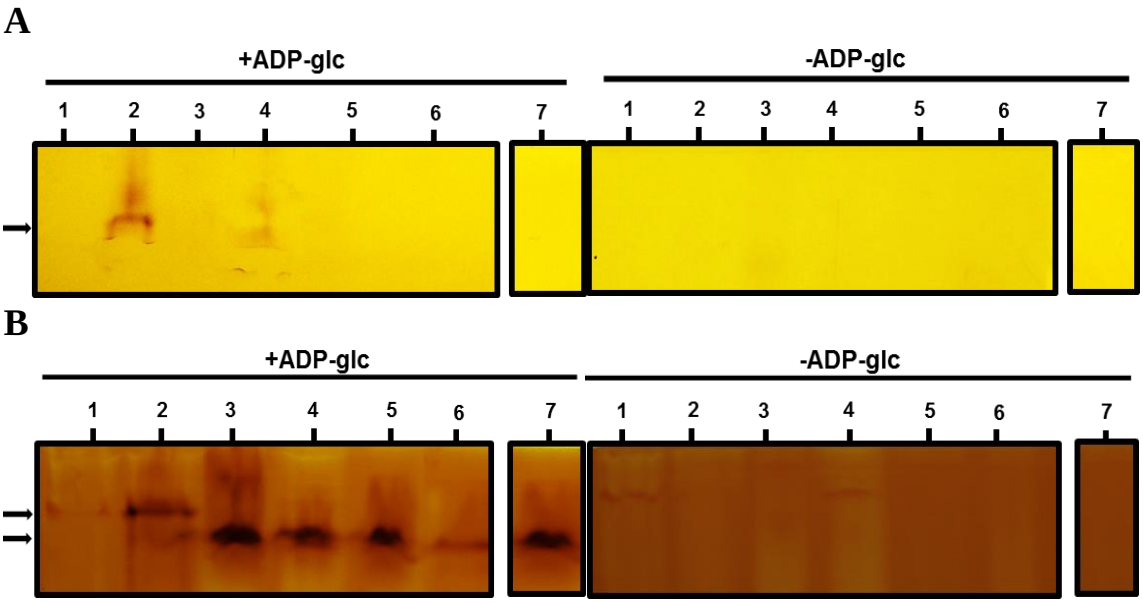


Figure 9. Zymogram analysis of glycogen/starch synthase activities. Total proteins of crude exctracts of the wild type strain (WT), the 187G11 mutant strain and the recombinant proteins GlgA2, a mixed of the crude extracts of the recombinant protein GlgA2 and the 187G11 mutant boiled (B) or not and the mixture (GlgA2 +187G11) incubated with the Amyloglucosidase (AGS) or with protease, were separated by native PAGE without glycogen **(A)** orcontaining 0.6% (p/v) of glycogen **(B).** The natives gels were then incubated with or without 3 mM ADP-glucose. Glycogen/starch synthase activities are witnessed after iodine staining as a dark activity bands. The wild-type extracts display two bands, a major band with high affinity to glycogen (arrow 1), and a minor band with a low affinity to glycogen (arrow 2) that co-migrate with the recombinant protein GlgA2. **1:** 187G11, **2:** WT, **3:** GlgA2, **4:** GlgA2+187G11, **5 :** GlgA2+187G11 boiled extract, **6 :** GlgA2+187G11 treated with AGS and boiled, **7 :** GlgA2+187G11 treated with protease.

Phylogenetic analysis of the SSIII-IV-GlgA2 glycogen/starch synthases

Several detailed and recent phylogenies of GT5 ADP-Glc requiring starch synthases have appeared (Ball et al., 2013). However the databases have considerably increased in size since our last analysis and now include a much larger diversity of cyanobacteria. For our phylogenetic analysis we have selected only the SSIII-SSIV-GlgA2 monophyletic subgroup of enzymes that was previously defined in these phylogenies with very high support and have restricted our alignment to these sequences. The GT5 ADP-Glc dependent glycosyl-transferases represent a distinctively prokaryotic group of enzymes with no representatives within eukaryotes with the noticeable exception of Archaeplastida. It is thus reasonable to assume that the green algae and plant sequences summarized in figure 10 were gained by LGT from a prokaryotic source. Because the tree is unrooted and because the phylogeny of GT5 glycogen/starch synthases shows many signs of signal erosion we cannot exclude the unlikely possibility that the sequence was donated to Chlamydiales by the Archaeplastida rather than the reverse. Nevertheless the phylogeny represented in figure 10 and detailed in the supplemental Figure S8 demonstrates that despite the growing databases, not only that the

Chlamydiales remain the most plausible donors for the ancestral plant SSIII-IV starch synthase gene but also that Cyanobacteria can be very confidently rejected as possible donors for the plant enzymes. The phylogeny now suggests in addition that GlgA2 may define a very ancient cyanobacterial enzyme as a substantial portion of the cyanobacterial diversity appears uninterrupted by foreign clades in a large size monophyletic group. Among the available cyanobacterial genome sequences, we estimate that 47 % of reported genomes lack GlgA2 while only 16 % lack GlgA1. Hence both sequences are largely distributed within cyanobacteria. Both GlgA1 and GlgA2 (2 cases) can function as sole glycogen/starch synthase. In addition both GlgA1 and GlgA2 absence is not confined to specific subgroups of cyanobacteria but is distributed throughout the cyanobacterial tree pointing to multiple gene loss events. A systematic search for GlgA2 in cyanobacteria has indeed yielded the finding of this enzyme in several of the most basal cyanobacterial clades (Colleoni and Suzuki, 2012) We further assayed the congruence of the GlgA2 phylogeny with the diversification of cyanobacteria (Supplemental Figures S9 and S10) as estimated through the 16S rRNA phylogeny and found a good level of congruence despite the intensive gene sharing and exchanges usually found in bacteria. A GlgA2-related sequence is also congruent with the diversification of Chlamydiales which contain no other glycogen/starch synthases (Ball *et al.*, 2013). On the other hand the rather restricted distribution within mostly marine gamma proteobacteria (purple sulfur bacteria (chromatiaceae) and methanotrophs) may argue for a more recent LGTs within these groups.

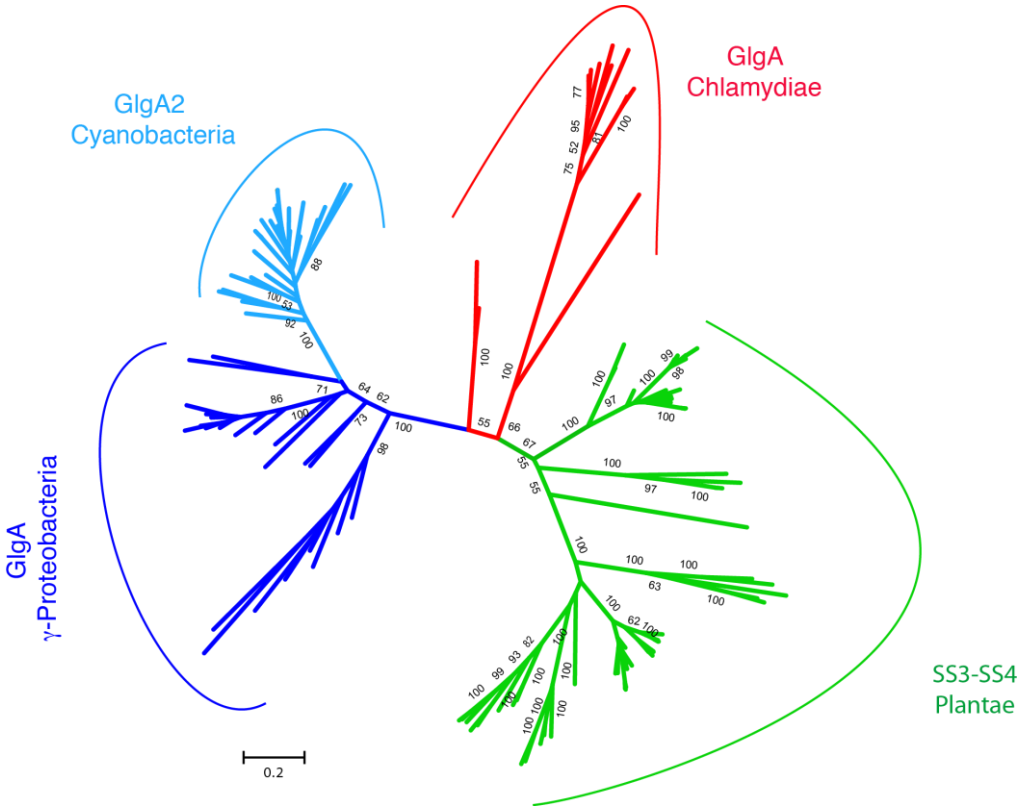


Figure 10. Unrooted phylogenetic tree of SSIII-IV-GlgA2 glycogen/starch synthases. Phylogenetic groups are color-coded according to their taxonomy: green corresponds to green linage, red corresponds to Chlamydiales, blue corresponds to Proteobacteria and turquoise corresponds to Cyanobacteria. Bootstrap values support, higher than 50 are indicated at nodes (1000 bootstrap replicates).

Discussion

A novel determinant of starch versus glycogen structure; glucan product length and particle seeding

In this work, we bring suggestive evidence that cyanobacterial starch depends on a function provided by GlgA2 that cannot be supplied by GlgA1. We believe that this function is distinct from polysaccharide synthesis priming since normal glycogen synthesis priming occurs in the 187G11 mutant. In plants, mutants defective for both SSIII and SSIV which are phylogenetically-related to GlgA2, are also starchless but do not produce any glycogen (Szydlowski et al., 2009). In that case, however, the missing function is thought to consist of polysaccharide synthesis priming. Indeed, transgenic expression of the self-priming Agrobacterium GlgA glycogen synthase in the Arabidopsis SSIII-IV double mutants restores starch synthesis (Crumpton-Taylor et al., 2013). Expression of both GlgA1 and GlgA2 was successfully achieved in *E. coli*. However successful complementation of the *E. coliglgA* mutation could only be achieved in the presence of maltose and no complementation was observed on mannitol or glucose-grown *E. coli* cultures. Maltose is known to induce the maltose operon *per se* the MalQ amylomaltase which elongates malto-oligosaccharides by a series of transfer reactions at the expense of glucose formation. The synthesis of long glucans leading to glycogen production by action of glycogen branching enzyme is prevented by the presence of the MalP and MalZ gene products yielding glucose and glucose-1-P thereby feeding bacterial metabolism and recessing the long chains to maltotetraose yielding a MOS pool consisting of small-size glucans. Hence *E. coli* cells growing on maltose contain a significant pool of small-size MOS. Selective complementation of *glgA* in maltose grown *E.coli* cells suggest that *in vivo,* these bacteria are using the MOS pool to elongate glucans for glycogen synthesis. In the presence of mannitol or glucose the absence of a sizeable MOS pool prevents glycogen synthesis priming. This suggests that neither recombinant GlgA2 nor recombinant GlgA1 or GlgA2* are able to prime polysaccharide synthesis in the absence of MOS. This was confirmed for GlgA2 through zymogram analysis. However when GlgA2 was purified from *Cyanobacterium* sp Clg1, GlgA2 was systematically able to prime polysaccharide in the same zymogram analysis. We therefore conclude that, unlike the *Agrobacterium tumefasciens* enzyme, the ability to prime does not define an intrinsic property of the cyanobacterial glycogen synthases and that this ability is dependent on either enzyme modification or supply of specific primers by other cyanobacterial factors. The GlgA2* mutant activity may or may not lack this essential property but nevertheless the very substantial decrease of its specific activity (at minima two orders of magnitude (Figure 3)) precludes its normal function for cyanobacterial starch synthesis. Interestingly our work suggests that the ability to prime polysaccharide synthesis appears to be selectively transferred by a cyanobacterial protein to the recombinant enzyme. This factor, hereby named factor X, deserves further attention and may be purified and analyzed from Cyanobacterium extracts in future work.

Our work further suggests that *Cyanobacterium* sp. CLg1 displays two separate pathways for polysaccharide synthesis affording the possibility to regulate both of these pathways at least partly independently. The first pathway leads to the production of short-chain glycogen and the second yields starch. The starch specific pathway consists at minima of GlgA2 (this work) and of GlgX2 (Cenci et al., 2013). As we previously proposed, glycogen, because of the instant accessibility of the glucose stores of its outer chains, defines an optimal structure to

ensure a fast adaptation of the carbon sink strength to optimize photosynthetic activity. On the other hand starch offers the opportunity to trap five-fold more (Cenci et al., 2013) carbon into a slow turnover storage polysaccharide form to ensure high respiration rates during the dark phase. These high respiration rates have been proposed by others to be needed in diazotrophic single cell cyanobacteria not only to supply nitrogenase with the required high levels of ATP and reducing power but also to further lower locally the oxygen levels through its respiratory consumption (Schneegurt et al., 1994). By reaching anoxia this would induce Nitrogenase synthesis and activity. We therefore predict that the absence of starch would abolish diazotrophy. Unfortunately the loss of diazotrophy of our axenic Cyanobacterium strain does not allow us to test this in a straightforward fashion. However in this respect, we wish to stress that the most abundant class of mutants of the green alga *Chlamydomonas reinhardtii* that are defective for hydrogen production under anoxic conditions by the oxygen-sensitive hydrogenase are those that we reported to substitute starch by glycogen synthesis (Posewitz et al., 2004).

All 6 cyanobacteria that have been proven to accumulate starch contain GlgA2. However many glycogen accumulating cyanobacteria also contain both GlgA1 and GlgA2 (Colleoni & Suzuki, 2012). We would like to propose that GlgA2 has evolved mainly to allow nitrogen fixation through the synthesis and mobilization of starch. Frequent loss of both diazotrophy and starch in cyanobacteria may not necessarily have been accompanied by that of GlgA2. In some cases GlgA2 may indeed have been lost as in *Prochlorococcus* and many related *Synechococcus* strains (Colleoni & Suzuki, 2012). In other cases, the single loss of the GlgX2 debranching enzyme would have converted the synthesis of high levels of starch into smaller levels of "phytoglycogen" (a polymer resembling glycogen but with slightly longer chains that results from impaired amylopectin crystallization) as evidenced in the GlgX2 mutants (Cenci *et al.*, 2013). The pool of phytoglycogen induced by the loss of GlgX2 function would lead to the production of increased glycogen amounts made of slightly longer chains that escapes the hypothesized tight regulation of GlgA1 by photosynthesis. Hence maintenance of both long-chain and short-chain glycogen may have been desirable in some glycogen accumulating cyanobacteria such as *Synechocystis* sp. PCC6803.

Our work emphasizes that the intrinsic properties of the glycogen/starch synthase possibly defines a novel determinant of starch versus glycogen synthesis. We believe that the specific properties concerned consist in the synthesis of chains with a length (DP > 12) compatible for their selective debranching by GlgX2. Indeed GlgX2 was demonstrated to display little or no activity toward glycogen chains and to require the longer chains present in amylopectin-like molecules (Cenci et al., 2013). Our results concerning the biochemical properties of GlgA2 are in agreement with these speculations. The absence of iodine stained polysaccharide product in zymogram analysis of recombinant GlgA1 despite the presence of significant activity measured in our radioactive ADP-glucose incorporation assays strongly suggests the presence of a distributive mode of action for GlgA1. Indeed an hypothetical distributive mode of action had been solely deduced previously by others from the detailed glycogen structures produced in the single GlgA1 and GlgA2 mutants of *Synechocystis* (Yoo et al., 2014). That GlgA1 is responsible for short chain glycogen synthesis is indeed suggested by the measure of significant (20%) residual crude extract glycogen synthase activity found in the 187G11 mutant which is also in agreement with its function in the synthesis of the remaining short-chain glycogen pool. The small increase of short chains in the glycogen structure of the GlgA2 mutant of Cyanobacterium when compared to the wild-type can be explained either by

the mutation of GlgA2 contributes in a minor fashion to glycogen synthesis or by the observed induction of phosphorylase activity in the mutant. We presently cannot distinguish between these two possibilities.

In addition to long chain synthesis, we believe that an additional function for starch versus glycogen synthesis carried by GlgA2 may consist of starch granule seeding. Indeed, a need exists at the core of the granule for the organization of the 3D crystalline growth of the individual granules. There is also a need to control starch granule size and hence starch granule seeding independently from glycogen to avoid physically blocking bacterial division.

The cyanobacterial origin of the SSIII-IV-GlgA2 subfamily of GT5 glycogen/starch synthases.

The phylogeny published in this and our previous work shows that two prokaryotic groups show a significant level of congruence between their diversification and the GlgA2 phylogeny. These are the Chlamydiales and the cyanobacteria. The few bacteria mainly basal gamma proteobacteria members presenting a GlgA2-SSIII-IV group can be easily explained through lateral gene transfers (LGTs) from cyanobacteria in a common marine environment. This work extends the distribution of cyanobacteria to the point where we can show that the phylogeny of GlgA2 shows an appreciable level of congruence with this group of bacteria especially when considering the high level of gene exchanges considered to occur in bacteria. However some cyanobacteria lack either GlgA1, GlgA2 or both through selective gene losses, but nevertheless, both enzymes are largely distributed within this group. Since diversification of cyanobacteria was initiated between 2 to 3 billion years (Sanchez-Baracaldo et al., 2014), the node uniting all cyanobacteria in Figure 7 is vastly more ancient than that uniting the green algae and land-plants. This conclusion invalidates that the root of the GlgA2-SSIII-IV group could lie within the Archaeplastida since the latter diversified after plastid endosymbiosis (dated between 0.9 to 1.6 billion years). A GlgA2-like gene is on the other hand universally distributed in all Chlamydiales where it defines the sole starch/glycogen synthase present. Chlamydiales are considered to be members of the bacterial PVC clade (consting of Planctomycetes, Verrucomicrobia, Chlamydiales). However no other glycogen accumulating PVC members are reported to contain enzymes of similar phylogenetic origin suggesting that the last common Chlamydiales ancestor may have received the gene by LGT from other bacteria. Hence the GlgA2 type of enzyme displays a very ancient origin in both Chlamydiales and cyanobacteria. GT5 (Glycosyl Transferase Cazy family 5) glycogen/starch synthases that use ADP-Glc as substrate are very largely distributed in Bacteria and Archea. This type of enzyme is only very distantly related to the UDP-Glc requiring GT5 or GT3 glycogen synthases distributed in glycogen storing eukaryotes. It is thus reasonable to conclude that the green algae land plants and glaucophyte SSIII-SSIV glycogen/starch synthases must have received the ancestor of SSIII-IV from a bacterial source, the most likely being an ancient Chlamydiale. The unrooted tree presented in this work does not clarify the origin of the SSIII-SSIV-GlgA subgroup of glycogen/starch synthases and a chlamydial proteobacterial or cyanobacterial origin remain possible. Nevertheless if we now exclude the Archaeplastida as the source for this enzyme, all these hypotheses agree with the presence of a chlamydial LGT to the Archaeplastida. We believe that among the 3 possible origins a cyanobacterial source defines the most probable scenario. Indeed while both Chlamydiales

and cyanobacteria display some level of congruence between their diversification and the phylogeny of their SSIII-IV-like glycogen synthase, the specialized function of GlgA2 in cyanobacteria evidenced in this work points to a possible link between the latter and diazotrophy in single cell cyanobacteria. This suggests a cyanobacterial origin since the conflict between oxygenic photosynthesis and diazotrophy probably predates the evolution of Chlamydiales from a PVC ancestor. This is further suggested by a suspected more ancient diversification of cyanobacteria and the presence of GlgA2 in some of the most basal clades. Also supporting such an origin is the unusual abundance of glycosyl hydrolases and glycosyltransferases related to storage polysaccharide metabolism in several cyanobacterial lineages including *Cyanobacterium* sp CLg1 which often display over twofold more enzymes than those found in most other bacteria and archaea (Colleoni & Suzuki, 2012). This higher complexity may reflect a very ancient subfunctionalization of storage polysaccharide metabolism into two types of distinct pathways: one producing glycogen controlled by photosynthesis and the other producing starch controlled independently possibly by Nitrogen metabolism. Gene losses, acquisition of multicellularity and adaptation and diversification of cyanobacteria may have yielded more simple networks from a subset of this very ancient and complex pathway.

Convergent evolution of starch aggregation in cyanobacteria and plants

We have previously noted that aggregation of storage polysaccharide into starch evolved several times independently in cyanobacteria, Archaeplastida, alveolates and cryptophytes and also possibly several times within the cyanobacteria and Archaeplastida. It is remarkable to note that on different occasions similar tools have been selected to achieve analogous functions. Natural selection led at least twice to the recruitment of GH13 GlgX-like proteins to achieve selective debranching of amylopectin precursors. This work shows that although the SSIII-SSIV-GlgA2 subfamily of GT5 glycogen synthases is of rather restricted distribution, such enzymes have also been recruited at least twice independently in cyanobacteria and Archaeplastida to fulfill a required function of starch synthesis. Does this mean that these two types of CAZymes will always be recruited to achieve analogous essential functions of starch metabolism through convergent evolution? The answer to this question seems at first glance to be no! Indeed some cyanobacteria, cryptophytes and alveolates apparently lack the GH13 GlgX-like enzymes while SSIII-IV-GlgA2 glycogen(starch) synthases are lacking in red algae, cryptophytes and alveolates (Ball *et al*., 2015). We believe that other more distantly related CAZymes will be recruited to do the very same job but this still needs to be demonstrated.

Materials and Methods

Strains and culture conditions

187G11 mutant of *Cyanobacterium* sp. CLg1 was obtained through UV mutagenesis campaign and grown in Artificial Sea Water medium (Rippka et al., 1979) in the absence (AS0 medium) or in the presence of nitrogen source provided by 0.88 mM of sodium nitrate (ASNIII medium) as described in Cenci et al., 2013. Wild-type and GlgA2 mutant of Synechocystis PCC6803 were kindly provided by Dr Zilliges. Both strains were grown in the presence (BG11) or in the absence of nitrogen (BG0) as described in Grundel et al., 2012.

Transmission electron microscopy observation

Cyanobacteria were cultivated in 50 mL nitrogen-deprived medium (ASO) medium and harvested by centrifugation (5 min at 4000 g at 4 °C) after 2 weeks. The cells were fixed with glutaraldehyde, post-fixed with osmium tetroxide and embedded in Epon resin. 70 nm-thin sections were cut with a diamond knife in a Leica UC6 microtome and post-stained with periodic acid thiosemicarbazide silver proteinate (PATAg) (Gallant & Guilbot, 1969). Drops of dilute suspensions of water-soluble polysaccharide fractions were deposited on glow-discharged carbon-coated copper grids and allowed to dry after negative staining with 2% uranyl acetate. All specimens were observed with a Philips CM200 transmission electron microscope (TEM) operating at 80 kV. Images were recorded on Kodak SO163 films.

Purification and structural analysis of water-soluble polysaccharide (WSP)

WSP of *Cyanobacterium* sp. CLg1 and Synechocystis were purified from wild type and mutant strains cultivated in 300 mL of ASO during 12 days and BG0 during 4 days respectively and harvested at the middle of the day by centrifugation at 3600 g during 15 min at 4 °C. The cell suspension (10 mL) was disrupted through a French Press. Starch-pellets were separated from WSP by spinning the lysate at 16000 g for 15 min at 4 °C. WSP in the supernatant and starch-pellet were quantified by amyloglucosidase assay following the instructions of R-Biopharm. Results are expressed in mg of polysaccharide/milligram of total protein. The total protein concentration was determined in the supernatant using the Bradford method (Bio-Rad). WSP were sized on exclusion chromatography (Toyopearl TSK HW 55) pre-equilibrated at 1 $mL.min^{-1}$ in 10% DMSO (D = 1.8 cm L = 60 cm). Polysaccharides were quantified in each fraction (1 mL) by the phenol-sulfuric acid method (Fox & Robyt, 1991). Polysaccharides contained in fractions 35 to 45 were pooled and further incubated *Pseudomonas* sp. isoamylase (1U) and pullulanase (1U) (Megazyme), in Sodium Acetate 55 mM pH 3.5. The linear glucan chains were separated according to their degree of polymerization by HPAEC-PAD as described previously (Colleoni et al., 1999).

Zymogram analysis

Cells were grown for 10 days in 3 liters liquid ASNIII medium and harvested at the middle of the day by centrifugation (3000 g at 4°C during 15 min.). The cell pellets were washed three times with 20 mL of cold Tris-acetate buffer (25 mM Tris-acetate, pH 7.5, 10 mM DTT) before disrupting by French press at 1250 psi. The lysate was centrifuged 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 (150 mM NaCl, 25 mM Tris-acetate, pH 7.5, 5 mM dithiothreitol, 10% glycerol). The proteins were eluted at 4 $ml.min^{-1}$ using buffer B (150 mM NaCl, 25 mM Tris-acetate, pH 7.5, 5 mM dithiothreitol, 10% glycerol, 1 M NaCl) in 25 mL. Eluted proteins were desalted and concentrated to 1 milliliter using ultrafiltration system (Millipore). The semi-purified crude extracts were separated by non-denaturing polyacrylamide gel electrophoresis (PAGE) containing 0.6% of rabbit glycogen (Sigma-Aldrich). After electrophoresis, gels were incubated overnight at room temperature in starch synthase buffer (70 mM Gly-gly pH 7.5, 135 mM $(NH_4)_2SO_4$, 280 mM NaF, 330 mM trisodium citrate, 290

mM sodium acetate, 3 mM ADP-glucose, 67 mM β-mercapto-ethanol). Starch synthase activities were then visualized as dark activity bands after soaking native PAGE in iodine solution (0.5 g I_2, 10 g KI).

Gene cloning and sequencing

Starch metabolism genes: glgC (KR020055), glgA1 (AHB52787), glgA2 (AHB52788), gbss (AHB52786), glgB1 (AFP43334), glgB2 (AFP43335), glgB (AFP43336), glgB4 (AHB52790), glgX1 (AGI19288), glgX2 (AGI19289), apu13 (AHB52783), apu57 (AHB52784), amg (AHB52785), glgP (AHB52789), malQ (AHB52791) were amplified from genomic DNA of mutant strain using primers designed in the untranslated region as described in Cenci et al., 2013. Starch/glycogen synthase genes (*glgA1*, *glgA2* and *glgA2**) were amplified from genomic DNA of wild type (*glgA1* and *glgA2*) and 187G11 mutant strains (*glgA2**). Primers include restriction sites in order to clone the *glgA* genes either in pGex (GE Healthcare) or pET15 (Novagen) expression vectors (underlined letters) :

BamHI-*glgA*1-pGexF GGATCCATTCCCTCTGAGTCTGTGTGGCAG GCAA,

NcoI-*glgA*1-pET15F CCATGGGCAAAATATTATTTGTGGCGGCAGAAGCATC,

XhoI-*glgA*1R CTCGAGTTAAATAATTCCATCGATCGCATCTTGATAC,

EcoRI-*glgA*2-pGexFGAATTCTATATAGTTCAAATTGCCTCCGAATGTCCT,

NcoI-*glgA*2/*glgA*2*-pET15F CTCGAGTATATAGTTCAAATTGCCTCCGAATGTCCT,

XhoI-*glgA*2/*glgA2**R CTCGAGTTACTTATCTCTTAAAAAATCATATAATTCA,

The PCR experiments were conducted at 95 °C for 5 min; 30 cycles of denaturation at 98 °C for 30 s, annealing 30 s at 59.6 °C for *glgA*1, *glgA*2 and *glgA**, and extension 1 min 30 s 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™-TR, 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 mutation was identified by alignment with the wild type gene using the BLASTn program. The insertion of the cloning product in pGEX or pET15 was done by the T4 ligase (Thermo Scientific) and transferred into the chemical competent *E. coli* TOP10 Mach1TM-TR, and plated on LB agar with Ampicillin. Purified plasmids were also sequenced by GATC Biotech Company.

Protein expressions in *Escherichia coli* JW3392-1 Δ*glgA*

Escherichia coli wild type strain (BW25113) and the derivative single knockout *ΔglgA* mutant(JW3392-1) of Keio collection (Baba et al., 2006) were provided by *E.coli* Stock center (http://cgsc.biology.yale.edu). Δ*glgA* mutant was lysogenized with ΔDE3 phage in order to insert the inducible T7 RNA polymerase gene (□DE3 lysogenization kit, Novagen). Δ*glgA* DE3mutantwas thentransformed with pET-15-*glgA*1, pET-15-*glgA*2, pET-15-*glgA*2* pGEX-*glgA*1, pGEX-*glgA*2. pET and pGEX expression vectors allow the synthesis of

recombinant protein without and with N-terminus GST tag, respectively. Transformed *E.coli* strains were grown in 200 mL autoinductible medium (Formedium^{TM}) in the presence of ampicillin (100 µg/mL) at 30°C during 36 hours. The cells were harvested by centrifugation 16000g for 10 min at 4 °C and the pellets were resuspended in 5 mL of cold buffer (Tris/acetate 25 mM pH 7.5 10 mM DTT) before lysing the cells by sonication. Crude extacts were fractionnated and stored at -80°C for further analyses. Starch synthase activities in the *E.coli* crude extract were determined either by ^{14}C-glucose incorporation assay (described below) or by zymogram analysis. The complementation experiment was carried out in 250 mL M9 liquid medium (38 mM Na_2HPO_4, 22mM KH_2PO_4, 8.5 mM NaCl, 18 mM NH_4Cl chloride, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 0.4% casaminoacids.) supplemented with either 2% glucose, mannitol or maltose. After 12 h incubation at 37 °C, the cells are harvested by centrifugation (15 min at 4000 g) and the pellets are washed and resuspended in cold buffer (Tris/acetate 25 mM pH 7.5 10 mM DTT). After sonication (three times for 30 seconds) and centrifugation, the amount of glycogen and protein were measured in the supernatants using amyloglucosidase assay (R-biopharm) and bradford method (Biorad), respectively. The results are expressed as mg WSP.mg^{-1} total protein.

Ability of GlgA2 to prime polysaccharide synthesis

Crude extracts of wild type and mutant strain 187G11 of *Cyanobacterium* sp. CLg1 were produced as mentioned above (Zymogram analysis). Recombinant proteins GlgA1, GlgA2, GlgA2* were produced in *Escherichia coli* JW3392-1 Δ*glgA* as mentionned above (Protein expressions in *Escherichia coli* JW3392-1 ΔglgA). Recombinant protein GlgA2 (50 µg) was mixed either with a crude extract of the mutant strain 187G11 (50 µg) or with a crude extract of mutant strain 187G11 pre-heated at 99°C for 5 minutes, or with a crude extract of the mutant strain 187G11 treated with Amyloglucosidase (Megazyme) then pre-heated at 99°C for 5 minutes, or with a crude extract of mutant strain 187G11 incubated with proteinase K (ThermoFisher) at 60°C during 30 minutes, the proteinase K was inactivated by the addition of 2 mM of EGTA (Ethylene glycol tetraacetic acid). Samples were then separated onto non-denaturing polyacrylamide gels electrophoresis (PAGE) containing or not 0.6% of rabbit glycogen (Sigma-Aldrich). Crude extracts of wild type and mutant strains of *Cyanobacterium* sp. CLg1, were used as references. After electrophoresis, gels were incubated overnight at room temperature in starch synthase buffer (70 mM Gly-gly pH 7.5, 135 mM $(NH_4)_2SO_4$, 280 mM NaF, 330 mM trisodium citrate, 290 mM sodium acetate, 3 mM ADP-glucose, 67 mM β-mercapto-ethanol). Starch synthase activities were then visualized as dark activity bands after soaking native PAGE in iodine solution (0.5 g I_2, 10 g KI).

Western blot analysis

Proteins bound to starch granules of CLg1 were extracted by denaturing 1 mg of purified starch granules in 50 µL of SDS/β-mercaptoethanol buffer for 10 min at 95 °C. After centrifugation at 10000 g during 10 min, proteins specifically attached to starch granules are found in the supernatant. Granule-bound proteins were loaded onto 9 % SDS-PAGE. Western blot was then carried out as described presiously (Ral et al., 2006). Polyclonal primary antibodies raised against GlgA1 and GlgA2 of synechocystis PCC6803 and secondary antibody were diluted at 1:1000 and 1:20000 in blocking buffer, respectively. The immunocomplexes were detected by chimioluminescence following the instructions of ECL^{TM} prime western blotting reagent kit (GE healthcare).

Starch/glycogen synthase assay

The starch synthase activities were measured by following the incorporationof ^{14}C-glucose onto glycogen particles. The reaction was carried out at the initial velocity by incubating 40 µL of enzyme preparation and 60 µL of incubation buffer (50 mM HEPES-NaOH pH 7, 10 mg.mL^{-1} glycogen, 100 mM $(NH_4)_2SO_4$, 10 mM Dithiothreitol , 0.5 mg.mL^{-1} bovine serum albumin, 3 mM of ADP-Glucose and 2 µM ADP-^{14}C-[U]-Glc) for 15 minutes at 35°C. The reaction is stopped by precipating labeled glycogen with 1 mL methanol-KCl (75%v/v; 1%wt/v). The samples are stored at -20 °C during 10 min and then centrifuged 5 min at 3000 g at 4 °C. After centrifugation, the glycogen-pellets are resuspended with 200 µL of distilled water. This step was repeated twice before mixing the sample with 2.5 mL of scintillation liquid. The radioactivity incorporated onto glycogen was determined by liquid scintillation counting.

Phylogenetic tree

Homologs of GlgA 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 (final alignment of 595 amino acids available from SGB). 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 et al., 2008) model with heterogeneous gamma rate distribution across sites (+ G) 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.

Supplemental data

Supplemental Figure S1: Comparison of growth rates of 187G11 mutant and wild-type strains in continuous and day/night cycles.

Supplemental Figure S2: Analysis of starch metabolizing enzymes in the crude extract of 187G11.

Supplemental Figure S3: Semi-quantitative assay of GlgX2 activity in the 187G11 extract.

Supplemental Figure S4: Highly conserved tyrosine residues in starch/glycogen synthase

Supplemental Figure S5: GST-GlgA1 and GST-A2 recombinant proteins.

Supplemental Figure S6: Structural analysis of the glycogen accumulated by the *Escherichia.coli* strain complemented with GlgA1 or GlgA2.

Supplemental Figure S7: Zymogram analysis of glycogen/starch synthase activities in presence of maltoheptaose.

Supplemental Figure S8: Detailed phylogenetic tree of glycogen/starch synthases belonging to SSIII/SSIV/GlgA2 family

Supplemental Figure S9: Maximum likelihood phylogenies of GlgA2 and 16S RNA of cyanobabacteria.

Supplemental figure S10: Congruence of maximum likelihood phylogenies of GlgA2 and 16S RNA of cyanobacteria.

Acknowledgments

SB and CC were supported by CNRS, Université des Sciences et Technologies de Lille, Région Nord-Pas-de-Calais, and Agence Nationale de la Recherche (grant # ANR-BLAN07-3_186613).

Authors contributions: C.C. and S. G. B. designed research and wrote the paper; K.D. and M.D. have performed the molecular characterization of 187G11 mutant strain; U.C and C.T. carried out the mutagenesis campaign and the screening process; E.S and Y.N. sequenced the CLg1 genome; TEM observations were performed by J-L.P and A.D.T.; A.S. and S.G. characterized the water soluble polysaccharide; S.D-T and F.J.F. produced the antibodies raised against glycogen/starch synthases. M-C.A. performed phylogenetic analysis; MD, KD, AS and MP have expressed GlgA1 and GlgA2 recombinant proteins.

References

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

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:** 2006 0008

Ball SG, Colleoni C, Arias MC (2015) The transition from glycogen to starch metabolism in cyanobacteria and eukaryotes. In: Nakamura Y, ed. Starch: metabolism and structure. Springer, 93-158.

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber AP, Gehre L, Colleoni C, Arias MC, Cenci U, Dauvillee D (2013) Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? Plant Cell **25:** 7-21

Bertoft E, Laohaphatanalert K, Piyachomkwan K, Sriroth K (2010) The fine structure of cassava starch amylopectin. Part 2: building block structure of clusters. Int J Biol Macromol **47:** 325-335

Cenci U, Chabi M, Ducatez M, Tirtiaux C, Nirmal-Raj J, Utsumi Y, Kobayashi D, Sasaki S, Suzuki E, Nakamura Y, Putaux JL, Roussel X, Durand-Terrasson A,

Bhattacharya D, Vercoutter-Edouart AS, Maes E, Arias MC, Palcic M, Sim L, Ball SG, Colleoni C (2013) Convergent evolution of polysaccharide debranching defines a common mechanism for starch accumulation in cyanobacteria and plants. Plant Cell **25:** 3961-3975

Cenci U, Nitschke F, Steup M, Minassian BA, Colleoni C, Ball SG (2014) Transition from glycogen to starch metabolism in Archaeplastida. Trends Plant Sci **19:** 18-28

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 Physiol **120:** 1005-1014

Colleoni C, Suzuki E (2012) Storage polysaccharide metabolism in Cyanobacteria.In: Tetlow I, ed. Essential reviews in experimental biology: Starch: Origins, Structure and metabolism. Society Experimental Biology, 217-254.

Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE (2013) Atmospheric oxygenation three billion years ago. Nature **501:** 535-538

Crumpton-Taylor M, Pike M, Lu KJ, Hylton CM, Feil R, Eicke S, Lunn JE, Zeeman SC, Smith AM (2013) Starch synthase 4 is essential for coordination of starch granule formation with chloroplast division during Arabidopsis leaf expansion. New Phytol **200:** 1064-1075

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

D'Hulst C, Wattebled F, Szydlowski N (2015) Starch biosynthesis in leaves and its regulation. In: Nakamura Y, ed. Starch: metabolism and structure. Springer, 161-210.

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

Fox JD, Robyt JF (1991) Miniaturization of three carbohydrate analyses using a microsample plate reader. Anal Biochem **195:** 93-96

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

Gallant D, Guilbot A (1969) Etude de l'ultrastructure du grain d'amidon à l'aide de nouvelles méthodes de préparation en microscopie électronique. Die Stärke **6:** 156-163

Grundel M, Scheunemann R, Lockau W, Zilliges Y (2012) Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the *Cyanobacterium* Synechocystis sp. PCC 6803. Microbiology **158:** 3032-3043

Gupta RS (2009) Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades. Int J Syst Evol Microbiol **59:** 2510-2526

Hizukuri S (1986) Polymodal distribution of the chain lengths of amylopectin and the crystalline structure of starch granules. Carbohydr Res **147:** 342-347

Laohaphatanaleart K, Piyachomkwan K, Sriroth K, Bertoft E (2010) The fine structure of cassava starch amylopectin. Part 1: Organization of clusters. Int J Biol Macromol **47:** 317-324

Le SQ, Lartillot N, Gascuel O (2008) Phylogenetic mixture models for proteins. Philos Trans R Soc Lond B Biol Sci **363:** 3965-3976

Melendez-Hevia E, Waddell TG, Shelton ED (1993) Optimization of molecular design in the evolution of metabolism: the glycogen molecule. Biochem J **295 (Pt 2):** 477-483

Nakamura Y (2015) Biosynthesis of reserve starch. In: Nakamura Y, ed. Starch: metabolism and structure. Springer, 161-210.

Posewitz MC, Smolinski SL, Kanakagiri S, Melis A, Seibert M, Ghirardi ML (2004) Hydrogen photoproduction is attenuated by disruption of an isoamylase gene in Chlamydomonas reinhardtii. Plant Cell **16:** 2151-2163

Ral JP, Colleoni C, Wattebled F, Dauvillee D, Nempont C, Deschamps P, Li Z, Morell MK, Chibbar R, Purton S, d'Hulst C, Ball SG (2006) Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in Chlamydomonas reinhardtii. Plant Physiol **142:** 305-317

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

Sanchez-Baracaldo P, Ridgwell A, Raven JA (2014) A neoproterozoic transition in the marine nitrogen cycle. Curr Biol **24:** 652-657

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 **176:** 1586-1597

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:** 2688-2690

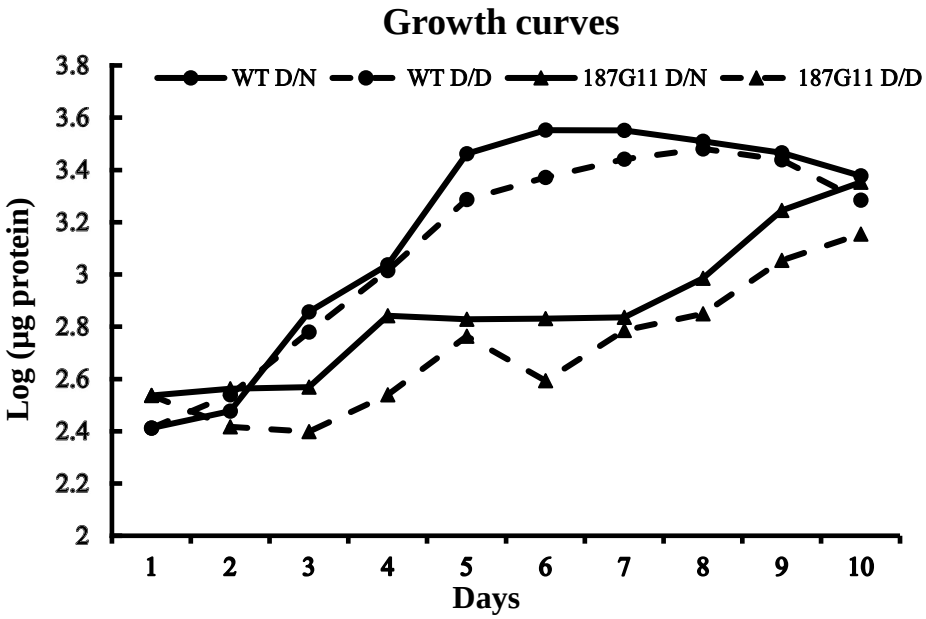
Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature **400:** 554-557

Szydlowski N, Ragel P, Raynaud S, Lucas MM, Roldan I, Montero M, Munoz FJ, Ovecka M, Bahaji A, Planchot V, Pozueta-Romero J, D'Hulst C, Merida A (2009) Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases. Plant Cell **21:** 2443-2457

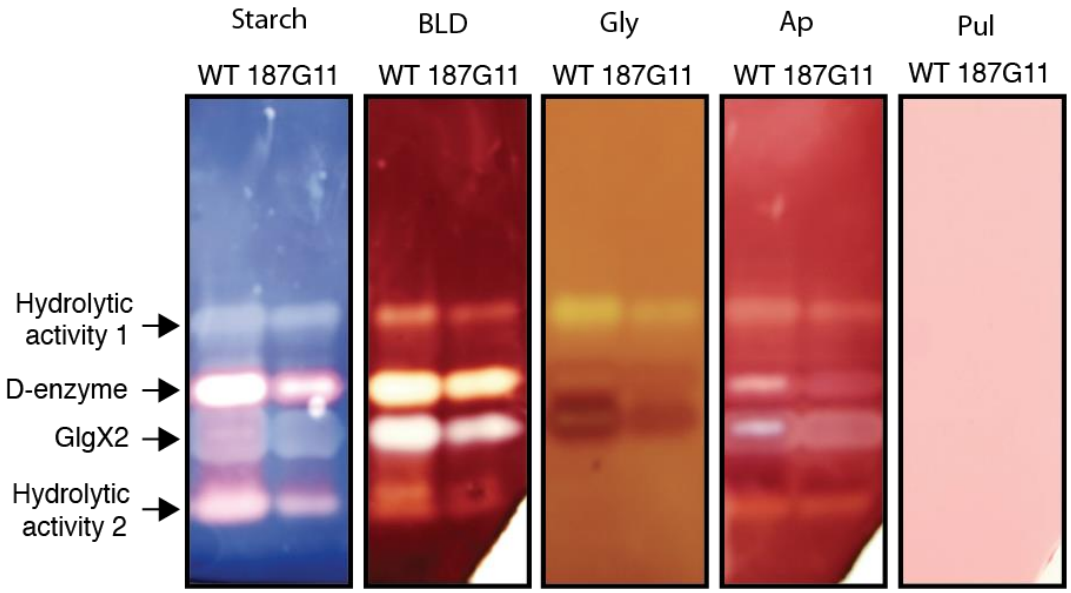
Welkie DG, Lee BH, Sherman LA (2016) Altering the Structure of Carbohydrate Storage Granules in the Cyanobacterium Synechocystis sp. Strain PCC 6803 through Branching-Enzyme Truncations. J Bacteriol **198:** 701-710

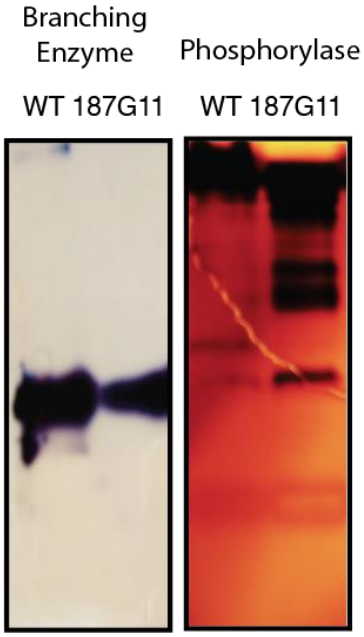
Yoo SH, Lee BH, Moon Y, Spalding MH, Jane JL (2014) Glycogen Synthase Isoforms in Synechocystis sp. PCC6803: Identification of Different Roles to Produce Glycogen by Targeted Mutagenesis. PLoS One **9:** e91524

Supplementary figures:

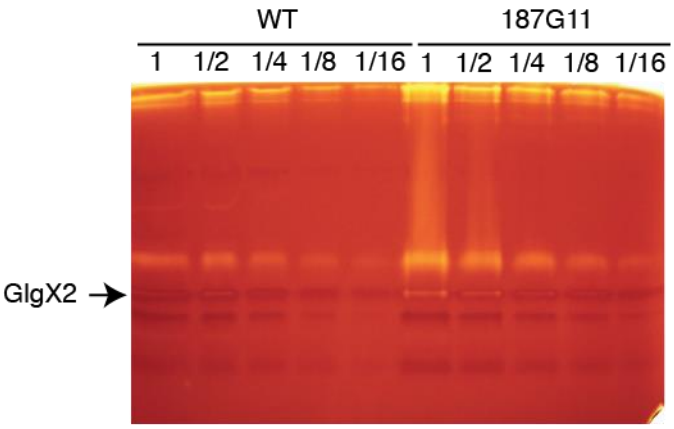


Supplemental Figure 1. Growth curves of wild type (WT) and mutant (187G11) strains. Both strains were grown for 10 days in Artificial Sea Water ASNIII at 22°C and subjected to light/dark cycle (12h/12h) (WT D/N, 187G11 D/N) or continued light (24h) (WT D/D, 187G11 D/D). The cells were harvested every day for 10 days. The protein concentration was determined in the supernatant using the Bradford method (Bio-Rad). The results are expressed as Log (µg protein).

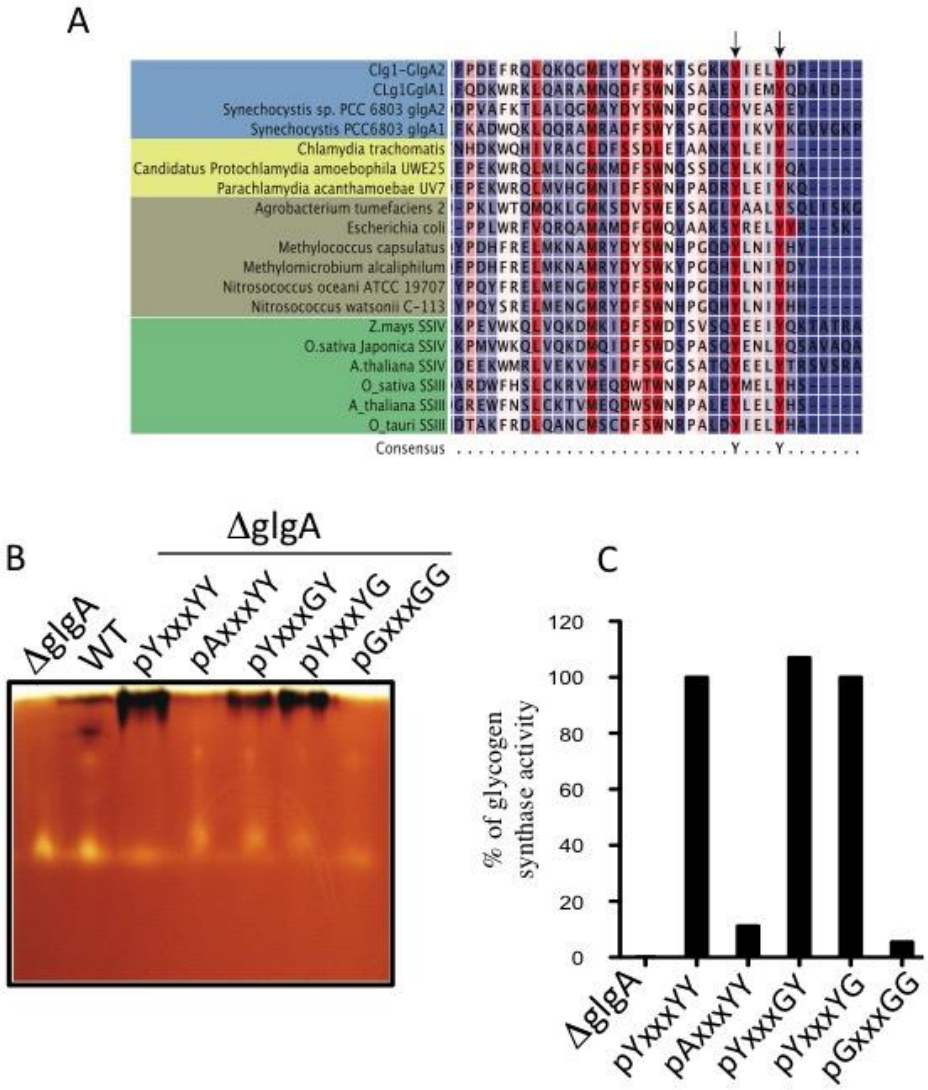




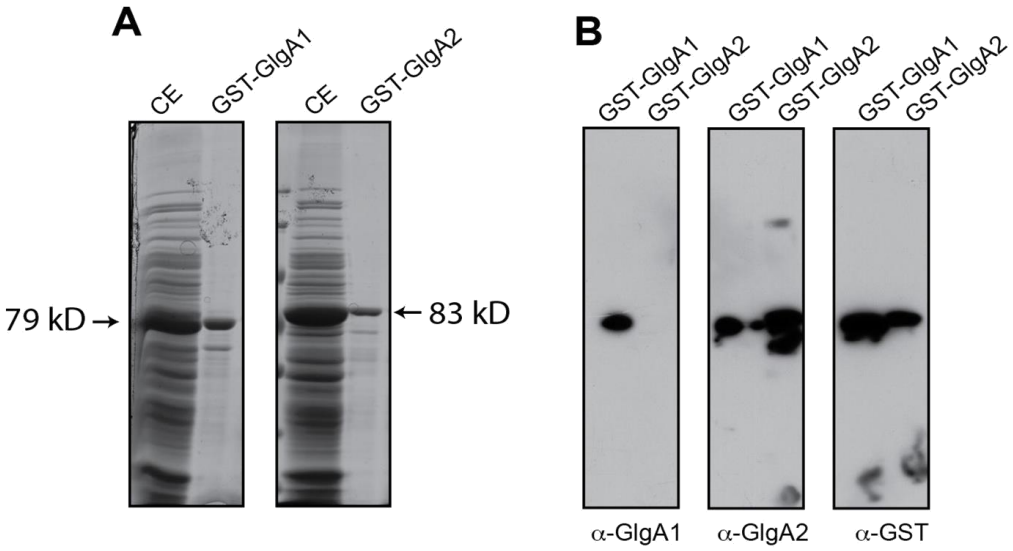
Supplemental Figure 2. Zymogram analysis of the different enzymatic activities of glycogen/starch metabolism (synthesis and degradation) of the WT and mutant 187G11 strains. Protein extracts are partially purified from both strains (see methods section). Partially purified proteins are separated on a native polyacrylamide gel, then transferred on a gel containing either 0.3% (weight/vol)of starch, beta limit dextrin (BLD), glycogen (Gly), amylopectin (Ap) or red pullulan (Pul). The gels are incubated on a buffer containing 25 mM Tris/acetate pH 7,5, in order to determine the hydrolytic activities, or in a buffer containing 10 mM of glucose-1-phosphate and rabbit phosphorylase a to reveal the branching activity, or in a buffer containing 10 mM of glucose-1-phosphate to assay phosphorylase activity. After incubation, the gels are stained with iodine solution. Colorless bands characterize hydrolytic activities while phosphorylase and branching activities revealed by glucan synthesis are characterized by a black activity bands.



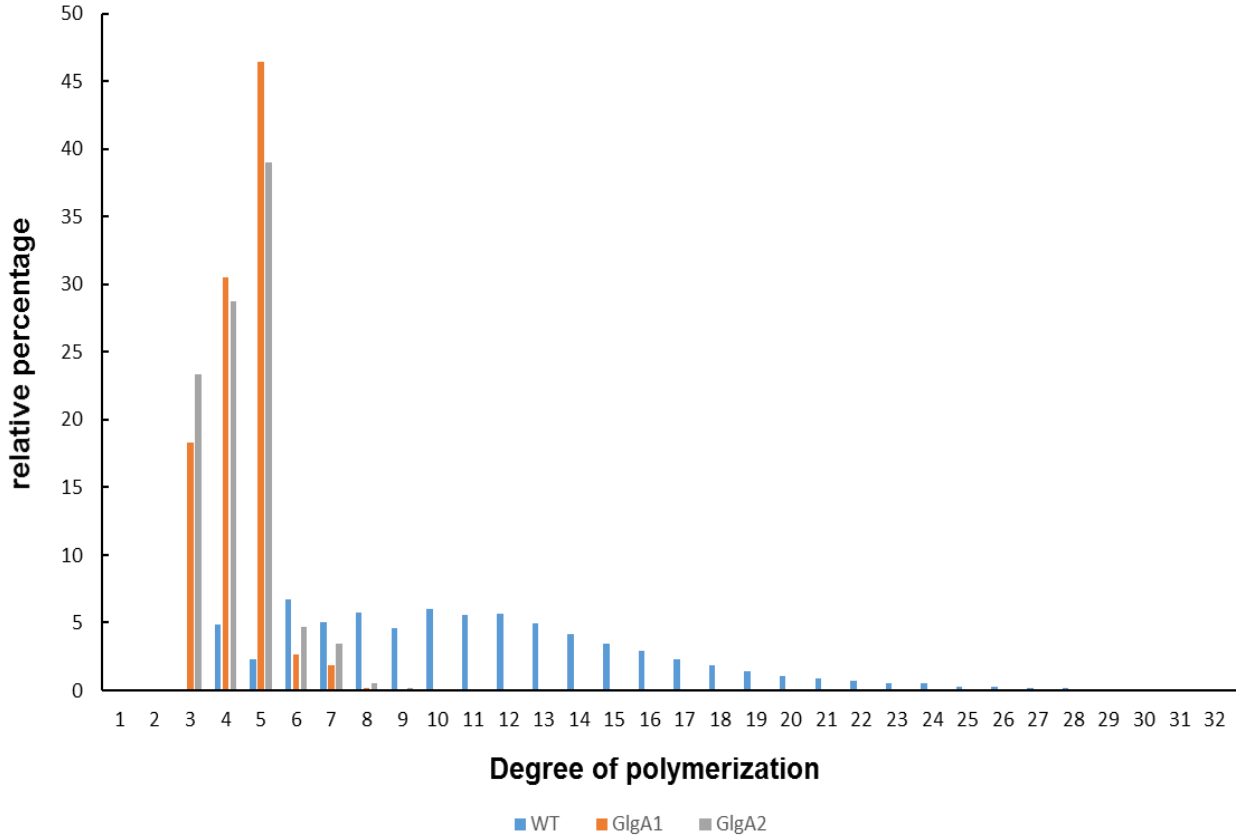
Supplemental Figure 3. **Zymogram analysis of the GlgX2 hydrolytic activity in the WT and mutant 187G11 strains.** Total proteins of crude extracts from both strains (52µg of total proteins) were diluted (1/2, 1/4, 1/8, 1/16) and separated on a native gel containing 0.6% (weight/vol) of glycogen. After migration, the gel is incubated in a buffer containing 25 mM Tris/acetate pH 7.5, and then the gel was stained with iodine solution.



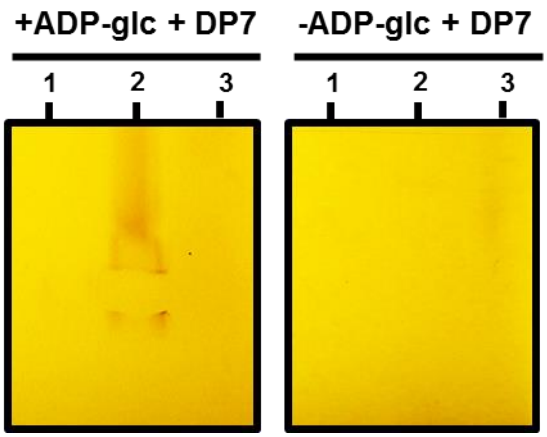
Supplemental Figure 4. (**A**) a YxxxY motif is highly conserved at the C-terminus of glycogen/starch synthase of plants and bacteria. Several allelic mutants were produced to investigate the role of tyrosine residues in the glycogen synthase activity of *E.coli*. Reverse primers were designed to replace either the first tyrosine residue (AxxxYY) or second tyrosine residue (YxxGY; YxxxYG) and all tyrosine residues (GxxxGG) by alanine or glycine residues. As control, the *GLGA* gene was amplified using reverse primer (YxxxYY). PCR products were cloned into the pET15 expression vector and used to transform a ΔglgA mutant strain of *E.coli*. (JW3392-1 from *E.coli* stock center). (**B**) Crude extracts of ΔglgA, WT and recombinant proteins were loaded onto glycogen synthase activity gel (i.e zymogram) and (**C**) glycogen synthase activities were measured by 14C-ADP-glucose incorporation assay as described in the methods section.



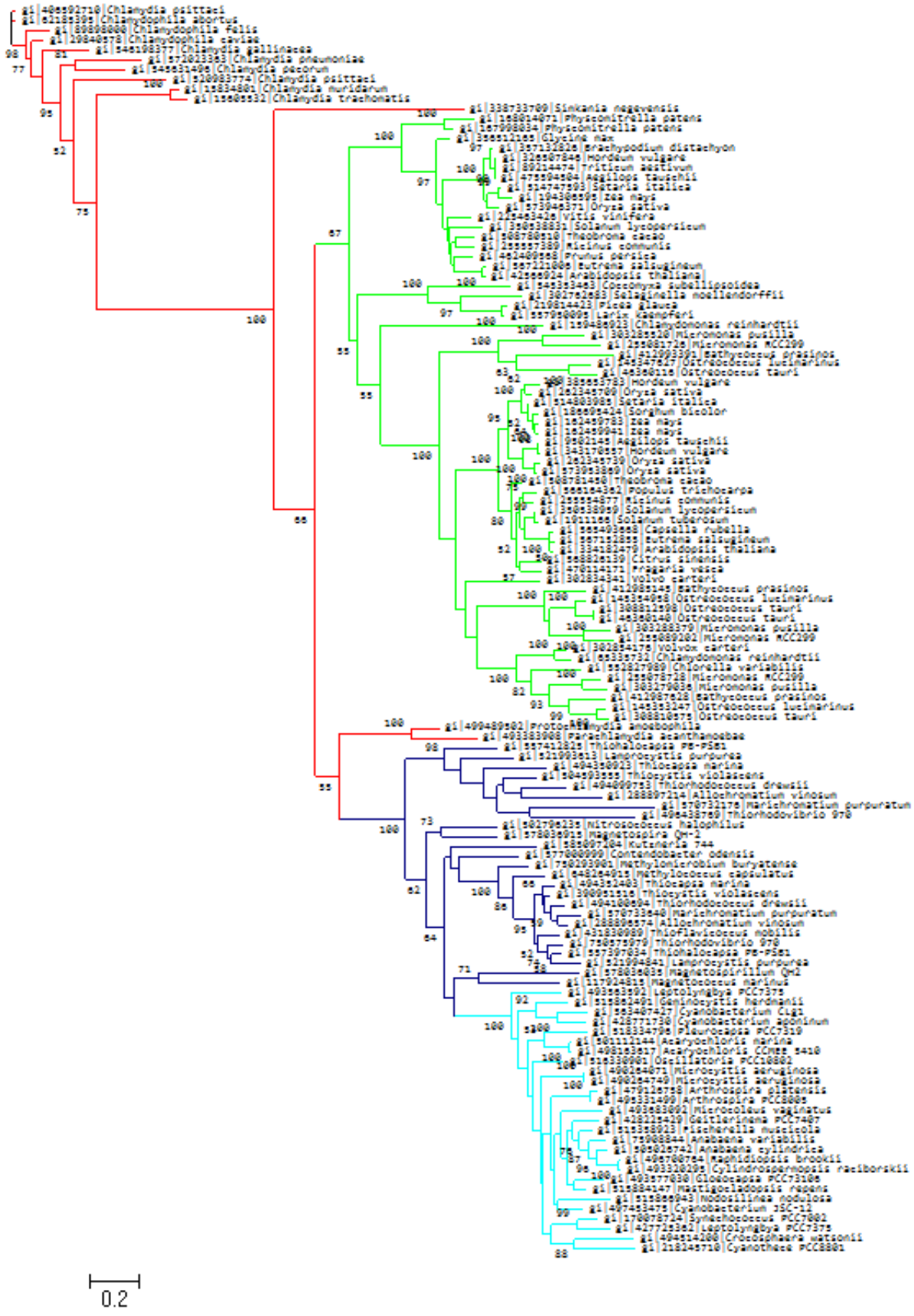
Supplemental Figure 5: (A) Recombinant proteins GlgA1 and GlgA2. The recombinant proteins GSTGlgA1 and GST-GlgA2 were extracted and observed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (9% SDS-PAGE). (B) Western blot analysis of purified recombinant protein GSTGlgA1 and GST-GlgA2 of *Cyanobacterium* sp CLg1 using the antibodies α-GlgA1 and α-GlgA2 of Synechocystis PCC6803 and polyclonal antibodies raised against gluthatione S-transferase (α-GST).



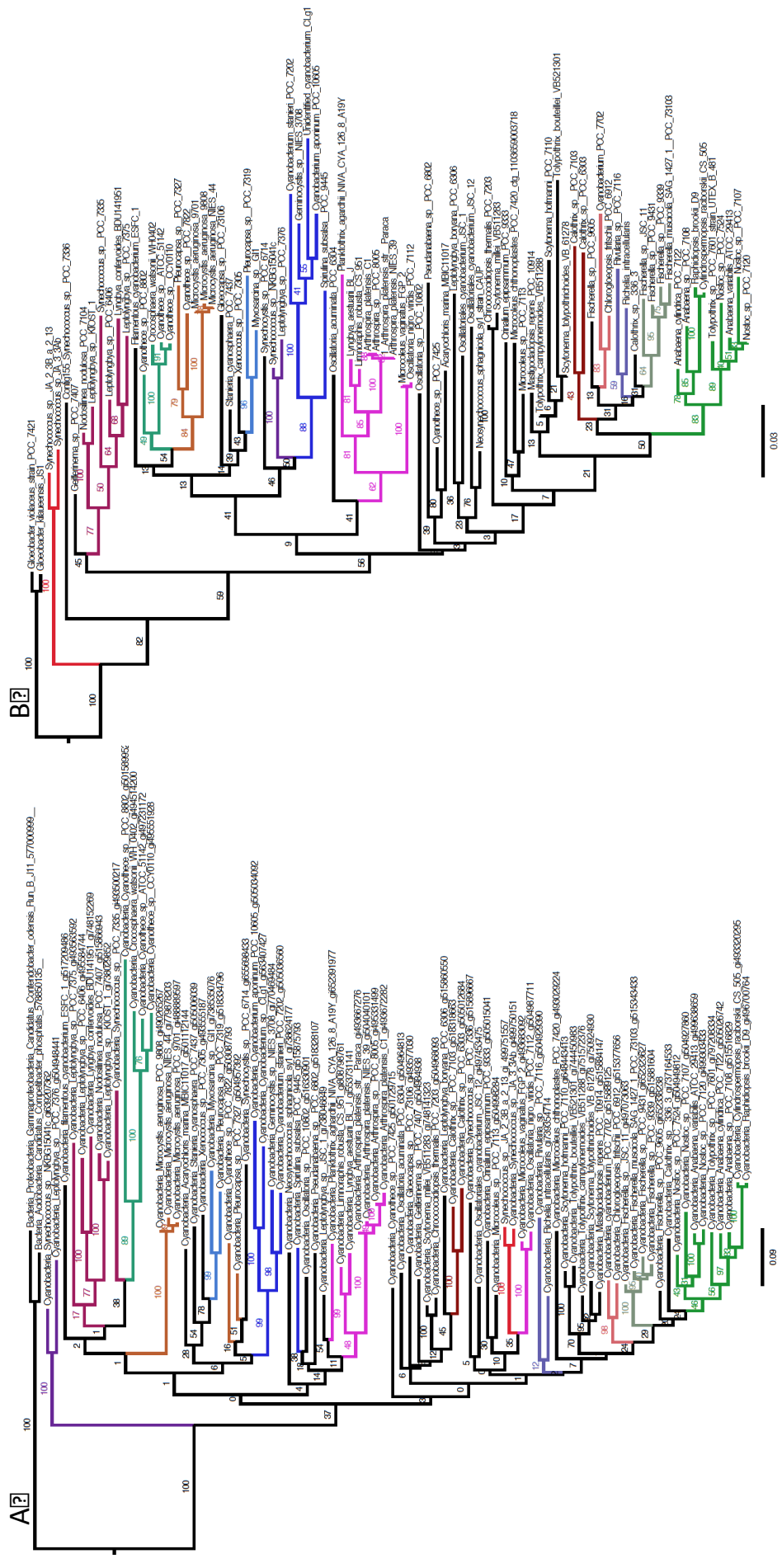
Supplemental Figure 6. Structural analysis of the glycogen accumulated by the *Escherichia.coli* wild type strain and the mutant strain ΔglgA deleted in glycogen synthase complemented with GlgA1 or GlgA2.The mutant strain ΔglgAwas complemented with pET-glgA1 or pET-glgA2, the cells were cultivated overnight in M9 medium at 37°C then harvested by centrifugation at 4000g during 10 minutes at 4°C. The cells were disrupted through a French Press then centrifuged at 16000 g for 10 min at 4 °C. Water-soluble polysaccharides present in the supernatant were purified from the wild type or the ΔglgA mutant complemented with pET-glgA1 or pET-glgA2. After complete digestion with commercial isoamylase, glucan chains were separated according to their degree of polymerization (DP) by FACE (Fluorescence anisotropy capillary electrophoresis). The wild type strain produces short and long chains, no difference was observed between the mutant strains complemented with GlgA1 or GlgA2, both produces a polysaccharide with short chains (DP3 to DP8).



Supplemental Figure 7. Zymogram analysis of glycogen/starch synthase activities. Total proteins of crude exctracts of the wild type strain (WT), the 187G11 mutant strain and the recombinant protein GlgA2, were separated by native PAGE without glycogen. The natives gels were then incubated with or without 3 mM ADP-glucose and 1.2mM of maltoheptaose (DP7). Glycogen/starch synthase activities are witnessed after iodine staining as a dark activity bands. The Wild type is able to initiate the glycogen synthesis while the 187G11 mutant and the recombinant protein GlgA2 are enable to initiate the glycogen synthesis. **1 :** 187G11, **2 :** WT, **3 :** GlgA2.



Supplemental Figure 8. Detailed representation of phylogenetic tree of glycogen/starch synthases belonging to the SIII/SSIV/GlgA2 family displayed in a summarized version in the main text (Fig. 10). See legend of Fig 10 for details.



Supplemental Figure 9. Maximum likelihood phylogeny of GlgA2 **(A)** and 16S RNA **(B)** of the same strains of Cyanobacteria. All the sequence for each tree was taken from the NCBI database, aligned with MUSCLE and then blocs were selected with BMGE (Criscuolo and Gribaldo 2010) and checked manually. The maximum likelihood phylogenies were built using IQTREE (Nguyen et al., 2015), and either LG4X matrix (Le et al. 2012) **(A)** and GTR+I+G matrix (Lanave et al. 1984, Yang 1994) **(B).** The bootstrap values are indicated on each branch, based on 1000 bootstrap replicas. **(A)** The outgroup was chosen among the closest sequences of cyanobacteria GlgA2 present in the NCBI database and that belongs to GlgA2 family. **(B)** The outgroup is composed of Gloeobacter 16S RNA sequences, that is known to be an early branching cyanobacterial group. The groups of organisms found in both trees are highlighted in different colors.

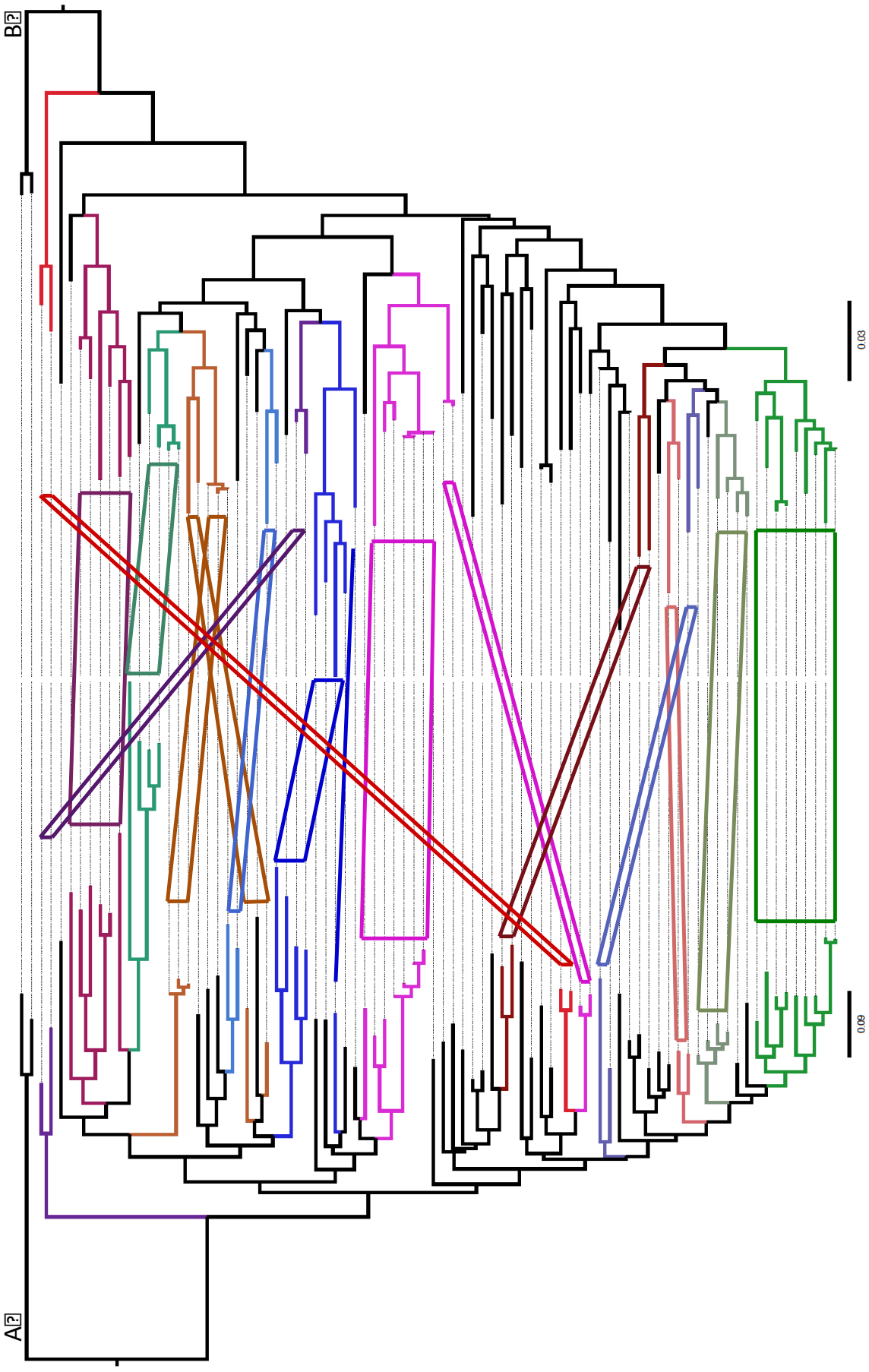
Relevant supplemental references

Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010 Jul 13;10:210.

Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. J Mol Evol. 1984;20:86-93.

Le SQ, Dang CC, Gascuel O. Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol Biol Evol. 2012 Oct;29(10):2921-36. **Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015 Jan;32(1):268-74.

Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 1994;39:306-314.



Supplemental Figure 10. Congruence of maximum likelihood phylogeny of GlgA2 (**A**) and 16S RNA (**B**). To better appreciate the (relative) congruence between GlgA2 tree **(A)** and 16SRNA tree (**B**), the same trees as figure S1 were positioned in front of each other. Parallel lines highlights congruence while disruption of such lines highlights congruence disruption.The groups colored with the same color show an appreciable level of topological congruence between their GlgA2 sequence (**A**) and their 16S RNA (**B**). Imperfect congruence between both trees can result either from LGT among cyanobacteria or can result from signal erosion problems witnessed selectively in the GT5 glucan synthase phylogenies. Such disruptions are more evident in the deepest branches. Low bootstrap values are due both to signal erosion and also to the patchy distribution of GlgA2.

General conclusions

General conclusions

In this thesis I have tried to bring additional insights into our understanding of storage polysaccharide synthesis in Cyanobacteria and Chlamydiales. I have shown that extensive (processive) glucan elongation is an additional property of glycogen/starch synthases that is required to make the synthesis of starch possible. Until very recently, starch debranching enzyme was thought to define the sole enzyme that was required to make the difference between synthesis of hydrosoluble glycogen and semi-crystalline starch. However, all attempts at producing starch-like material through heterologous expression of starch producing debranching enzymes starting from glycogen metabolism networks have presently failed (Welkie et al., 2016). We believe that in addition to polysaccharide debranching, a suitable balance of glycogen/starch synthases and of branching enzymes is required to achieve aggregation into large size starch granules in the presence of isoamylase. The finding of Chlamydiales with exactly the same enzyme properties as the maize starch generating Isa1 protein is a particularly striking result of my thesis, directly stemming from my own experimental work. It is quite obvious that the Archaeplastida have received the ancestor of the Isa1 gene from a Chlamydiales ancestor. We do not believe that the function of this enzyme was to trigger synthesis of starch-like material in Chlamydiales or their hosts. Indeed, these pathogens do not encode any enzyme able to degrade semi-crystalline starch material. As mentioned above, because Chlamydiales do not infect starch accumulating algae, but rather glycogen accumulating protists, we do not think that the presence of this enzyme effector immediately triggered starch synthesis. It nevertheless pre-adapted the host of plastid endosymbiosis to switch from the ancestral glycogen accumulating stage to that of starch. It onlyrequired a slight modification of the elongation to branching ratio, the pathogen's isoamylase would, then, have made the difference. This switch was possibly also allowed because eukaryotes had previously evolved mechanisms that enable them to clear deleterious abnormal glycogen structures known as lafora bodies. Hence the only additional thing that was needed was to evolve a protein that rendered starch more similar to these highly phosphorylated lafora polysaccharide aggregates. This immediately fed starch into a novel catabolic pathway. The rather straightforward fusion of a dikinase domain with a CBM (carbohydrate binding module) domain would have achieved this, thereby generating glucan-water dikinase. We now postulate that the initial function of the β-amylase dpe2 hydrolytic pathway evidenced in amoebozoa and excavates and lost from opisthokonts was degradation of such lafora bodies. The reason why natural selection favored starch over glycogen accumulation was possibly initially to bypass Chlamydiales access to glycogen degradation and restore host control of host cytosolic carbohydrate stores degradation as initially proposed by Cenci et al, 2014. Hence host-pathogen biotic interactions clearly governed evolution of novel important biochemical pathways.

Welkie DG, Lee BH, Sherman LA. Altering the Structure of Carbohydrate Storage Granules in the Cyanobacterium Synechocystis sp. Strain PCC 6803 through Branching-Enzyme Truncations.J Bacteriol. 2015 Dec 14;198(4):701-10. doi: 10.1128/JB.00830-15.

Cenci, U., Nitschke, F., Steup, M., Minassian, B. A., Colleoni, C., and Ball, S. G. (2014). Transition from glycogen to starch metabolism in Archaeplastida. Trends Plant Sci. 19, 18–28

References

References:

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

Almagro G, Viale AM, Montero M, Rahimpour M, Munoz FJ, Baroja-Fernandez E, Bahaji A, Zuniga M, Gonzalez-Candelas F, Pozueta-Romero J (2015) Comparative genomic and phylogenetic analyses of Gammaproteobacterial glg genes traced the origin of the *Escherichia coli* glycogen glgBXCAP operon to the last common ancestor of the sister orders Enterobacteriales and Pasteurellales. PLoS One **10:** e0115516

Arias MC, Danchin EGJ, Coutinho P, Henrissat B, Ball S. (2012) Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants. Mobile genetic elements **2**:81-87

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:** 2006 0008

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

Ball SG, Bhattacharya D, Weber AP (2016) EVOLUTION. Pathogen to powerhouse. Science **351:** 659-660

Ball SG, Colleoni C, Arias MC (2015) The transition from glycogen to starch metabolism in cyanobacteria and eukaryotes. *In* Y Nakamura, ed, Starch: metabolism and structure. Springer, pp 93-158

Ball SG, Colleoni C, Kadouche D, Ducatez M, Arias MC, Tirtiaux C (2015) Toward an understanding of the function of Chlamydiales in plastid endosymbiosis. Biochim Biophys Acta

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber AP, Gehre L, Colleoni C, Arias MC, Cenci U, Dauvillee D (2013) Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? Plant Cell 25: 7-21

Bertelli C, Aeby S, Chassot B, Clulow J, Hilfiker O, Rappo S, Ritzmann S, Schumacher P, Terrettaz C, Benaglio P, Falquet L, Farinelli L, Gharib WH, Goesmann A, Harshman K, Linke B, Miyazaki R, Rivolta C, Robinson-Rechavi M, van der Meer JR, Greub G (2015) Sequencing and characterizing the genome of *Estrella lausannensis* as an undergraduate project: training students and biological insights. Front Microbiol **6:** 101

Bertelli C, Collyn F, Croxatto A, Ruckert C, Polkinghorne A, Kebbi-Beghdadi C, Goesmann A, Vaughan L, Greub G (2010) The *Waddlia* genome: a window into chlamydial biology. PLoS One **5:** e10890

Bertoft E, Laohaphatanalert K, Piyachomkwan K, Sriroth K (2010) The fine structure of cassava starch amylopectin. Part 2: building block structure of clusters. Int J Biol Macromol **47:** 325-335

Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell **60:** 115-122

Brown DH, Illingworth B, Cori CF (1961) The Mechanism of the De Novo Synthesis of Polysaccharide by Phosphorylase. Proc Natl Acad Sci U S A **47:** 479-485

Brushia RJ, Walsh DA (1999) Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure. Front Biosci **4:** D618-641

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

Cenci U, Chabi M, Ducatez M, Tirtiaux C, Nirmal-Raj J, Utsumi Y, Kobayashi D, Sasaki S, Suzuki E, Nakamura Y, Putaux JL, Roussel X, Durand-Terrasson A, Bhattacharya D, Vercoutter-Edouart AS, Maes E, Arias MC, Palcic M, Sim L, Ball SG, Colleoni C (2013) Convergent evolution of polysaccharide debranching defines a common mechanism for starch accumulation in cyanobacteria and plants. Plant Cell **25:** 3961-3975

Cenci U, Nitschke F, Steup M, Minassian BA, Colleoni C, Ball SG (2014) Transition from glycogen to starch metabolism in Archaeplastida. Trends Plant Sci **19:** 18-28

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 Physiol 120: 1005-1014

Colleoni C, Linka M, Deschamps P, Handford MG, Dupree P, Weber AP, Ball SG (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

Colleoni C, Suzuki E (2012) Storage polysaccharide metabolism in *Cyanobacteria*.In: Tetlow I, ed. Essential reviews in experimental biology: Starch: Origins, Structure and metabolism. Society Experimental Biology, 217-254.

Cori CF, Cori GT, Green AA (1943) Crystalline muscle phosphorylase III Kinetics. J. Biol. Chem.**:** 33-55

Cori GT, Cori CF (1943) Crystalline muscle phosphorylase IV. Formation of glycogen. J. Biol. Chem.**:** 57-63

Cori GT, Larner J (1951) Action of amylo-1,6-glucosidase and phosphorylase on glycogen and amylopectin. J Biol Chem **188:** 17-29

Criscuolo A, Gribaldo S.(2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010 Jul 13;10:210.

Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R, Canfield DE (2013) Atmospheric oxygenation three billion years ago. Nature **501:** 535-538

Crumpton-Taylor M, Pike M, Lu KJ, Hylton CM, Feil R, Eicke S, Lunn JE, Zeeman SC, Smith AM (2013) Starch synthase 4 is essential for coordination of starch granule formation with chloroplast division during Arabidopsis leaf expansion. New Phytol **200:** 1064-1075

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

D'Hulst C, Wattebled F, Szydlowski N (2015) Starch biosynthesis in leaves and its regulation. In: Nakamura Y, ed. Starch: metabolism and structure. Springer, 161-210.

Facchinelli F, Colleoni C, Ball SG, Weber AP (2013) Chlamydia, cyanobiont, or host: who was on top in the menage a trois? Trends Plant Sci **18:** 673-679

Facchinelli F, Pribil M, Oster U, Ebert NJ, Bhattacharya D, Leister D, Weber AP (2013) Proteomic analysis of the *Cyanophora paradoxa* muroplast provides clues on early events in plastid endosymbiosis. Planta **237:** 637-651

Facon M, Lin Q, Azzaz AM, Hennen-Bierwagen TA, Myers AM, Putaux JL, Roussel X, D'Hulst C, Wattebled F (2013) Distinct functional properties of isoamylase-type starch debranching enzymes in monocot and dicot leaves. Plant Physiol **163:** 1363-1375

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

Fischer EH (2013) Cellular regulation by protein phosphorylation. Biochem Biophys Res Commun **430:** 865-867

Fischer EH, Krebs EG (1955) Conversion of phosphorylase b to phosphorylase a in muscle extracts. J Biol Chem **216:** 121-132

Fox JD, Robyt JF (1991) Miniaturization of three carbohydrate analyses using a microsample plate reader. Anal Biochem **195:** 93-96

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

Gallant D, Guilbot A (1969) Etude de l'ultrastructure du grain d'amidon à l'aide de nouvelles méthodes de préparation en microscopie électronique. Die Stärke **6:** 156-163

Gehre L, Gorgette O, Perrinet S, Prevost MC, Ducatez M, Giebel AM, Nelson DE, Ball SG, Subtil A (2016) Sequestration of host metabolism by an intracellular pathogen. Elife **5**

Grundel M, Scheunemann R, Lockau W, Zilliges Y (2012) Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the *CyanobacteriumSynechocystis* sp. PCC 6803. Microbiology 158: 3032-3043

Gupta RS (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, Coutinho PM (2002) Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? Trends Genet **18:** 437-440

Hizukuri S (1986) Polymodal distribution of the chain lengths of amylopectin and the crystalline structure of starch granules. Carbohydr Res **147:** 342-347

Illingworth B, Brown DH, Cori CF (1961) The de novo synthesis of polysaccharide by phosphorylase. Proc Natl Acad Sci U S A **47:** 469-478

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

Jones TH, Wright BE (1970) Partial purification and characterization of glycogen phosphorylase from *Dictyostelium discoideum*. J Bacteriol **104:** 754-761

Krebs EG, Fischer EH (1955) Phosphorylase activity of skeletal muscle extracts. J Biol Chem **216:** 113-120

Kwak JY, Kim MG, Kim YW, Ban HS, Won MS, Park JT, Park KH (2016) Properties of a glycogen like polysaccharide produced by a mutant of *Escherichia coli* lacking glycogen synthase and maltodextrin phosphorylase. Carbohydr Polym **136:** 649-655

Lanave C, Preparata G, Saccone C, Serio G. (1984) A new method for calculating evolutionary substitution rates. J Mol Evol. 1984;20:86-93.

Laohaphatanaleart K, Piyachomkwan K, Sriroth K, Bertoft E (2010) The fine structure of cassava starch amylopectin. Part 1: Organization of clusters. Int J Biol Macromol **47:** 317-324

Le SQ, Dang CC, Gascuel O. (2012) Modeling protein evolution with several amino acid replacement matrices depending on site rates. Mol Biol Evol. 2012 Oct;29(10):2921-36.

Le SQ, Lartillot N, Gascuel O (2008) Phylogenetic mixture models for proteins. Philos Trans R Soc Lond B Biol Sci **363:** 3965-3976

Melendez-Hevia E, Waddell TG, Shelton ED (1993) Optimization of molecular design in the evolution of metabolism: the glycogen molecule. Biochem J **295 (Pt 2):** 477-483

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

Melendez R, Melendez-Hevia E, Cascante M (1997) How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. J Mol Evol **45:** 446-455

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. Plant Cell **8:** 1353-1366

Nakamura Y (2015) Biosynthesis of reserve starch. In: Nakamura Y, ed. Starch: metabolism and structure. Springer, 161-210.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. (2015)IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015 Jan;32(1):268-74.

Nielsen TJ, Pradhan P, Brittingham A, Wilson WA (2012) Glycogen accumulation and degradation by the trichomonads *Trichomonas vaginalis* and *Trichomonas tenax*. J Eukaryot Microbiol **59:** 359-366

Park JT, Shim JH, Tran PL, Hong IH, Yong HU, Oktavina EF, Nguyen HD, Kim JW, Lee TS, Park SH, Boos W, Park KH (2011) Role of maltose enzymes in glycogen synthesis by *Escherichia coli*. J Bacteriol **193:** 2517-2526

Posewitz MC, Smolinski SL, Kanakagiri S, Melis A, Seibert M, Ghirardi ML (2004) Hydrogen photoproduction is attenuated by disruption of an isoamylase gene in *Chlamydomonas reinhardtii*. Plant Cell 16: 2151-2163

Ral JP, Colleoni C, Wattebled F, Dauvillee D, Nempont C, Deschamps P, Li Z, Morell MK, Chibbar R, Purton S, d'Hulst C, Ball SG (2006) Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii*. Plant Physiol 142: 305-317

Rall TW, Sutherland EW, Wosilait WD (1956) The relationship of epinephrine and glucagon to liver phosphorylase. III. Reactivation of liver phosphorylase in slices and in extracts. J Biol Chem **218:** 483-495

Rathore RS, Garg N, Garg S, Kumar A (2009) Starch phosphorylase: role in starch metabolism and biotechnological applications. Crit Rev Biotechnol **29:** 214-224

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

Roach PJ (2015) Glycogen phosphorylation and Lafora disease. Mol Aspects Med **46:** 78-84

Schinzel R, Nidetzky B (1999) Bacterial alpha-glucan phosphorylases. FEMS microbiology letters **171:** 73-79

Sanchez-Baracaldo P, Ridgwell A, Raven JA (2014) A neoproterozoic transition in the marine nitrogen cycle. Curr Biol **24:** 652-657

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 **176:** 1586-1597

Shepherd D, Segel IH (1969) Glycogen phosphorylase of *Neurospora crassa*. Arch Biochem Biophys **131:** 609-620

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690

Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng Des Sel **19:** 555-562

Subtil A, Parsot C, Dautry-Varsat A (2001) Secretion of predicted Inc proteins of *Chlamydia pneumoniae* by a heterologous type III machinery. Mol Microbiol **39:** 792-800

Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. Nature 400: 554-557

Szydlowski N, Ragel P, Raynaud S, Lucas MM, Roldan I, Montero M, Munoz FJ, Ovecka M, Bahaji A, Planchot V, Pozueta-Romero J, D'Hulst C, Merida A (2009) Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases. Plant Cell 21: 2443-2457

Welkie DG, Lee BH, Sherman LA (2016) Altering the Structure of Carbohydrate Storage Granules in
the *Cyanobacterium Synechocystis* sp. Strain PCC 6803 through Branching-Enzyme
Truncations. J Bacteriol 198: 701-710

Wilson WA, Roach PJ, Montero M, Baroja-Fernandez E, Munoz FJ, Eydallin G, Viale AM, Pozueta-Romero J (2010) Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev **34:** 952-985

Wu G, Muller M (2003) Glycogen phosphorylase sequences from the amitochondriate protists, *Trichomonas vaginalis*, *Mastigamoeba balamuthi*, *Entamoeba histolytica* and *Giardia intestinalis.* J Eukaryot Microbiol **50:** 366-372

Yoo SH, Lee BH, Moon Y, Spalding MH, Jane JL (2014) Glycogen Synthase Isoforms in Synechocystis sp. PCC6803: Identification of Different Roles to Produce Glycogen by Targeted Mutagenesis. PLoS One **9:** e91524

Yang Z.(1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 1994;39:306-314.