Master Programme in Molecular Biology and Biomedicine

University of Crete
Departments of Biology and Medicine
&
Institute of Molecular Biology and Biotechnology
Foundation for Research and Technology - Hellas

Eirini Areti Karapidaki (441)

Master Thesis

Overexpression, purification and biochemical characterization of native Parhyale hawaiensis glycosyl hydrolases

Υπερέκφραση, καθαρισμός και βιοχημικός χαρακτηρισμός ενδογενών ενζύμων διάσπασης κυτταρίνης του οργανισμού *Parhyale hawaiensis*

Supervisor: Dr. Anastasios Pavlopoulos
Developmental Morphogenesis Lab
IMBB-FORTH

Committee Members: Dr. Giorgos Gouridis
IMBB-FORTH

Ass. Prof. Ioannis Pavlidis
Department of Chemistry, University Crete

Heraklion, November 2023

Acknowledgements

The present master dissertation, titled "*Overexpression, purification and biochemical characterization of native Parhyale glycosyl hydrolases*", was conducted within the framework of the postgraduate program, Molecular Biology and Biomedicine, during the academic year 2022-2023. This endeavor has been achieved through collaboration and engagement with several individuals, each of whom has made a significant contribution to its progress. Therefore, I would like to dedicate the subsequent section to honor and express my heartfelt gratitude to these people, who, in their own unique ways, have contributed to the successful completion of this work.

First of all, I would like to express my deepest gratitude to my supervisor and group leader of Developmental Morphogenesis Lab, Dr. Anastasios Pavlopoulos, who provided me with unwavering support, guidance, and valuable feedback throughout the entire research process. Additionally, I am sincerely thankful to Dr. Valia Stamataki, Postdoctoral Researcher in Anastasios Pavlopoulos' Lab, for her continuous support, feedback, and patience during the various stages of this research. Her constructive criticism and valuable suggestions have greatly contributed to the quality and rigor of this thesis. Subsequently, I genuinely appreciate the invaluable support and beneficial advice of Marina Ioannou, Technician and Lab manager, that has been instrumental in this endeavor. Their mentorship has not only shaped my research skills but has also had a profound impact on my overall academic journey.

I extend my heartfelt appreciation to all the members of Developmental Morphogenesis Lab, who have been an essential part of my educational adventure and the successful completion of this master's thesis. I truly desire to express my gratitude especially to John Rallis (PhD Student), Maria Kalogeridi (PhD Student) and Niovi Rafailidou (MSc Student) for their support and encouragement during the ups and downs of this academic endeavor. Their collaboration, and shared enthusiasm have enriched this journey in so many ways.

Furthermore, I would like to sincerely thank the members of my examination committee, Dr. Giorgos Gouridis and Ass. Prof. Ioannis Pavlidis.

Last but not least, I am thankful from the bottom of my heart for my friends and family, who have been my pillars of strength. Their support, understanding, and unwavering belief in my abilities have been a constant source of motivation.

Table of Contents

Abstract

Cellulose is the main component of lignocellulosic plant biomass presenting an abundant and renewable resource for biofuel production and various biotechnological applications. However, the compact, crystalline structure and chemical composition of lignocellulose make it resistant to chemical breakdown. To address this limitation, the biofuel industry utilizes enzymatic cocktails produced by living organisms to promote the degradation of cellulose into fermentable sugars. Recent research in certain marine crustaceans feeding on wood has revealed that they encode in their genome all the necessary enzymes for extracting sugars from cellulose. Remarkably, unlike most other animals, these organisms do not depend on symbiotic microorganisms. Among these marine crustaceans, the genetically tractable *Parhyale hawaiensis* provides an excellent model to investigate autonomous cellulose biodegradation in a simple digestive gland. My Master thesis research concentrated on the analysis of *Parhyale* glycosyl hydrolase genes from the GH7 family, known as cellobiohydrolases or exoglucanases, that are uniquely present in crustaceans among metazoans. First, a gain-of-function strategy was devised to explore the activity of an in vivo overexpressed, tagged GH7 protein in cellulose digestion. This approach involved the generation of stable *Parhyale* transgenic lines using the Minos transposable element, in combination with a heat-inducible system for conditional overexpression of *Parhyale* GH7. Upon heat-shock, a significant increase in cellulolytic activity was observed with the dinitrosalicylic acid assay, suggesting that overexpressed GH7 enhances the cellulolytic capacity of the enzymatic cocktail in *Parhyale* digestive glands. Second, the overexpressed, tagged GH7 enzyme was affinity purified and activity assays with the cellobiohydrolase-specific p-nitrophenyl-β-D-cellobioside substrate confirmed its functionality. Comparisons between the *Parhyale* GH7 and a commercial, recombinant, fungal cellobiohydrolase indicated that they exhibit comparable activities. Finally, a CRISPR non homologous end joining knock-in method was employed for midgut-specific expression of the tagged *Parhyale* GH7 cellobiohydrolase, further expanding the prospects for understanding and utilizing these enzymes. This research represents the first integrated genetic and biochemical study of the autonomous cellulolytic capacity of *Parhyale* that opens up various avenues for future investigations and potential biotechnological applications.

Keywords: Cellulose digestion; Cellobiohydrolase/exoglucanase GH7; Heat-inducible gene overexpression; CRISPR NHEJ knock-in; Affinity purification; Biochemical assays.

Περίληψη

Η κυτταρίνη είναι το κύριο συστατικό της λιγνοκυτταρινικής φυτικής βιομάζας, η οποία αποτελεί άφθονο και ανανεώσιμο πόρο για διάφορες βιοτεχνολογικές χρήσεις και συγκεκριμένα στην παραγωγή βιοκαυσίμων. Ωστόσο, η περίπλοκη της δομή και η χημική της σύσταση την καθιστούν ανθεκτική στη χημική διάσπαση. Για να αντιμετωπιστεί αυτό, η διάσπαση της κυτταρίνης μέσω ενζυματικών κοκτέιλ, που παράγονται από ζωντανούς οργανισμούς, έχει αναδειχθεί ως μια πολλά υποσχόμενη λύση για τη βιομηχανία βιοκαυσίμων. Πρόσφατη έρευνα σε θαλάσσια καρκινοειδή, που τρέφονται με ξύλο, αποκάλυψε ότι κωδικοποιούν στο γονιδίωμά τους όλα τα απαραίτητα ένζυμα για την εξαγωγή σακχάρων από την κυτταρίνη. Είναι αξιοσημείωτο ότι αυτοί οι οργανισμοί δεν εξαρτώνται από συμβιωτικούς μικροοργανισμούς, σε αντίθεση με άλλα ζώα. Μεταξύ αυτών των καρκινοειδών, ο οργανισμός *Parhyale hawaiensis* είναι ένα εξαιρετικο μοντέλο για τη διερεύνηση της πέψης της κυτταρίνης χωρίς μικροβιακή συνεισφορά. Η ανάλυση επικεντρώθηκε στο πιο άφθονο γονίδιο υδρολάσης γλυκοζυλίου, μια εξωγλουκανάση, που στο εξής αναφέρεται ως GH7. Επινοήθηκε μια στρατηγική υπερέκφρασης της GH7 για να διερευνηθεί ο ρόλος της στην πέψη της κυτταρίνης. Αυτή η προσέγγιση περιλάμβανε τη δημιουργία σταθερών διαγονιδιακών σειρών *Parhyale* χρησιμοποιώντας το μεταθετό στοιχείο Minos, σε συνδυασμό με θερμο-επαγόμενη ρυθμιστική αλληλουχία για την υπό όρους υπερέκφραση της *Parhyale* GH7. Μετά από έκθεση σε υψηλότερη θερμοκρασία, παρατηρήθηκε σημαντική αύξηση στη διάσπαση πολυσακχαριτών υποδηλώνοντας ότι η υπερεκφρασμένη GH7 ενισχύει την ικανότητα διάσπασής τους από το ενζυματικό κοκτέιλ που παράγουν οι πεπτικοί αδένες. Επιπλέον, ο καθαρισμός της GH7 πρωτεΐνης από τα υπόλοιπα ενδογενή πεπτικά ένζυμα του εντέρου και οι δοκιμασίες δραστικότητας με ειδικά υποστρώματα επιβεβαίωσαν τη λειτουργικότητά της. Οι συγκρίσεις μεταξύ της *Parhyale* GH7 και μιας αντίστοιχης μυκητιακής GH7 εξωγλουκανάσης έδειξαν ότι η πρώτη παρουσίασε πιο αποτελεσματική δράση, υπό συγκεκριμένες συνθήκες. Παράλληλα με τη μέθοδο υπερέκφρασης, χρησιμοποιήθηκε CRISPR NHEJ knock-in προσέγγιση, με στόχο τη σήμανση των ενδογενών GH7, διευρύνοντας περαιτέρω τις προοπτικές κατανόησης και χρήσης αυτών των ενζύμων. Αυτή η έρευνα υπογραμμίζει τη σημασία των γλυκοζυλο υδρολασών στην πέψη της κυτταρίνης και ανοίγει διάφορους δρόμους για μελλοντικές έρευνες και βιοτεχνολογικές εφαρμογές.

Λέξεις κλειδιά: Διάσπαση κυτταρίνης, Εξωγλουκανάση GH7, Θερμοεπαγώμενη υπερέκφραση γονιδίων, CRISPR NHEJ knock-in, Πρωτεϊνικός καθαρισμός, Βιοχημική ανάλυση.

Introduction

An Emerging Versatile Model Organism for Development, Evolution and Genetic Biology Research: The case of crustacean *Parhyale hawaiensis*

Research in the field of biology has greatly benefited from model organisms like *Drosophila melanogaster* and *Mus musculus*, which have facilitated the study of basic genetic principles and developmental processes (Sun & Patel, 2019). Over time, researchers have chosen additional model organisms to study specific topics, such as *Caenorhabditis elegans* for nervous system development and *Danio rerio* (zebrafish) for vertebrate genetics and development. However, these classical model systems represent only a small fraction of biological diversity, and researchers are increasingly turning to novel model organisms to bridge phylogenetic gaps and investigate unique biological processes (Sun & Patel, 2019). Arthropods, the most diverse group of animals on Earth, have been a subject of scientific curiosity for centuries (Sun & Patel, 2019). While *Drosophila* has been the primary research system for arthropod development, it has limitations, and other arthropod species offer remarkable insights into the evolution of developmental procedures (Sun & Patel, 2019).

To enable broader comparative studies of arthropod development, *Parhyale hawaiensis*, an amphipod crustacean, was established as a novel research system. Belonging to the Malacostraca class, which includes renowned crustaceans like crabs, lobsters, shrimps, and crayfish, as well as mantis shrimps, woodlice, and krill, *Parhyale* stands out as a representative model within this diverse group of marine animals (Rehm et al., 2009). Its adaptability and the expanding toolkit of experimental techniques and resources available for research have contributed to its status as a valuable research model (Rehm et al., 2009).

Figure 1. Parhyale hawaiensis male adults.
(reproduced from Averof, 2022)

Parhyale hawaiensis (Fig. 1) was originally documented by James D. Dana in 1853 on Maui, Hawaii (Myers, 1985). In 1997, it was introduced to laboratory research from a population collected in Chicago by Professor Nipam Patel (Rehm et al., 2009; Rallis et al., 2021). *P. hawaiensis*, a detritivorous species, is widely distributed across tropical coastlines and can adapt to various environmental conditions, including changes in salinity, temperature, and nutrient availability (Rallis et al., 2021). *Parhyale*'s suitability for research is further enhanced by its relatively short generation time, ease of cultivation in laboratory settings, and the availability of individuals at different developmental stages year-round (Rallis et al., 2021). Similar to the rest of amphipods, *Parhyale* undergoes direct development, bypassing intermediate larval stages. After approximately ten days of embryogenesis at 26°C, newly hatched juveniles closely resemble miniature adult forms (Rallis et al., 2021). These juveniles grow through successive molts and attain sexual maturity within six to seven weeks of hatching (Rallis et al., 2021).

Since the early 2000s, *P. hawaiensis* has gained popularity as a versatile model organism for a wide range of developmental, evolutionary, behavioral, and environmental studies (Wolff et al., 2018; Bruce & Patel, 2020; Kwiatkowski et al., 2023; Artal et al., 2017). More specifically,

research in *P. hawaiensis* has addressed questions about embryonic axis formation, segmentation and germ layer identity during development (Kao et al., 2016). It has also shed light on the evolution of arthropod body plan diversity by studying the expression and function of Hox genes (Pavlopoulos et al., 2009). Furthermore, *Parhyale*'s ability to regenerate limbs has provided insights into regenerative processes (Konstantinides & Averof, 2014).

In evolutionary developmental biology, having tools to manipulate gene activity and create mutant or transgenic animals is crucial for studying gene function (Kontarakis & Pavlopoulos, 2014). *P. hawaiensis* is amenable to a variety of experimental manipulations, including microinjection, transgenesis, lineage tracing, ablation, and more (Kontarakis & Pavlopoulos, 2014). Specifically, transgenesis in *Parhyale* can be achieved using the Minos transposable element. Additionally, an endogenous heat-inducible promoter from *Parhyale* has been isolated and is used routinely for conditional misexpression studies (Kontarakis & Pavlopoulos, 2014). Moreover, the field of *Parhyale* research continues to benefit from recent advancements in gene manipulation technologies, like CRISPR-Cas9, offering a diverse toolkit for targeted genetic investigations in this emerging research system (Serano et al. 2016; Kao et al. 2016; Martin et al., 2016). Moreover, tools and resources have been developed for studying early *Parhyale* embryo fate map and embryonic development stages (Rehm et al., 2009).

Overall, *Parhyale* has become a valuable model organism for understanding the genetic and developmental processes that drive organismal diversification (Sun & Patel, 2019).

The importance of Lignocellulose Digestion

Land plants fix the majority of carbon they capture through photosynthesis into lignocellulose, which consists of cellulose, hemicellulose, pectin, and lignin. Throughout a plant's life, this intricate structure provides it with architectural strength and protection against herbivores and pathogens (Cragg et al., 2015). The main component of lignocellulose is cellulose, corresponding to 40–60% of the dry weight, which is a-1,4-linked chain of glucose monosaccharides. Additionally, cellulose, with the aid of lignin, is organized in structures called microfibrils, which in turn form macrofibrils (Kao et al., 2016). Plant cellulose, being the most abundant organic polymer on Earth and one of the richest sources of carbon fixed through photosynthesis on the planet, is the prime target for bioconversion into fermentable sugars (saccharification) and offers a sustainable reservoir of polysaccharides that can be used for the fermentation of biofuels, catering to both industrial and domestic demands (Chang & Lai, 2018).

Nowadays, the strong global need for energy, the instability of petroleum supplies, and apprehensions about climate change on a global scale have prompted renewed efforts in the advancement of alternative energy sources that can replace fossil-based transportation fuels (Himmel et al., 2007). Consequently, numerous nations have launched comprehensive research and development initiatives in the field of biofuels, which represent a sustainable and renewable energy solution, in order to replace other environmentally damaging methods (Himmel et al., 2007).

It is known that production of first-generation biofuels competes with the human food and animal feed industries for the use of biomass and agricultural land. On the contrary, second-generation biofuels aim to exploit lignocellulosic material (plant biomass) that is plentiful from agricultural, forest and industrial residues and inedible crops. Therefore, it is essential to shift

from first-generation biofuels, criticized for their land use impact, to second-generation biofuels that use inedible parts of plants, specifically lignocellulose, as main component of biomass, for that kind of energy production (Bugg et al., 2011).

Biomass has the potential to serve as a raw material for the production of biofuels. However, the process of extracting simpler metabolites from lignocellulose is challenging, because enzymes cannot readily access the crystalline cellulose structure (Chang & Lai, 2018). The crystalline nature of cellulose, the complex hemicellulose layer covering cellulose microfibrils, and the coating of polysaccharide components by lignin hinder enzymatic catalysis (Cragg et al., 2015). In industrial applications, this challenge is resolved through harsh chemical and physical pre-treatment methods. However, in nature, organisms achieve the degradation of lignocellulose under conditions that are physiologically manageable (Cragg et al., 2015).

In summary, overcoming biomass recalcitrance and improving conversion processes are essential for cost-effective and sustainable biofuel production, addressing pressing global challenges (Himmel et al., 2007).

Lignocellulose Digestion across the Tree of Life

Lignocellulose digestion strategies are diverse across the Tree of Life. Various organisms use different mechanisms for cellulose degradation, including oxidative attack, mechanical disruption and enzymatic catalysis. The breakdown of woody plant tissues necessitates at least a partial disruption of lignin before the polysaccharides become accessible. Many fungal species accomplish this through the use of oxidative enzymes (King et al., 2010). In nature, cellulose and hemicellulose are broken down by a combination of enzymes, mainly glycoside hydrolases (GHs), with the help of polysaccharide esterases and, in some cases, polysaccharide lyases (Cragg et al., 2015). The specific enzymes and their composition vary across organisms and environments. Recently, a new type of enzyme called lytic polysaccharide monooxygenases (LPMOs) was discovered, which can digest cellulose and hemicellulose through oxidative attacks on polymer chains (Cragg et al., 2015).

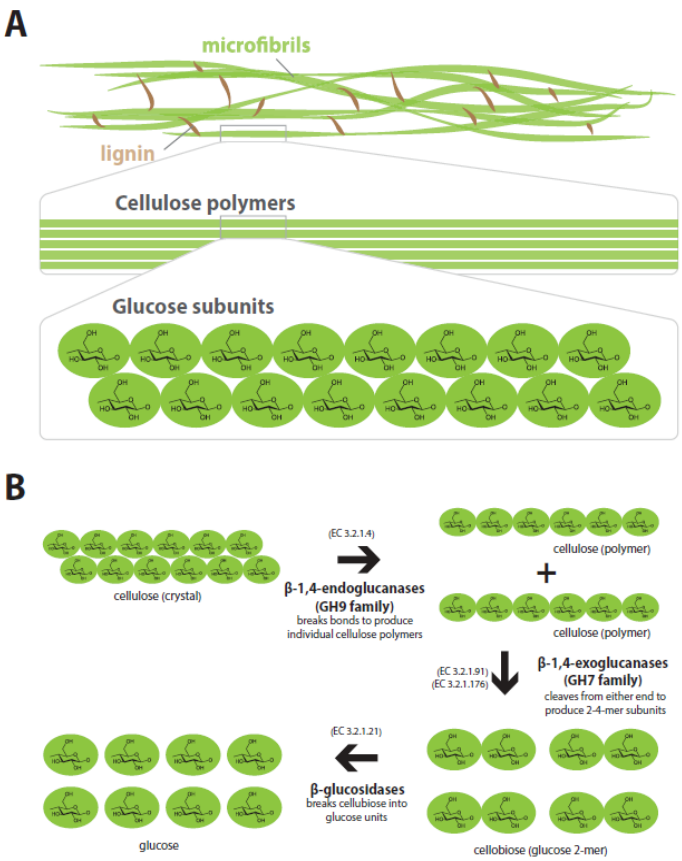

Figure 2. Overview of lignocellulose biodegradation. *Schematics of (A) lignocellulose structure and (B) reactions catalyzed by cellulolytic enzymes (reproduced from Kao et al., 2016).*

For the efficient enzymatic digestion of cellulose, three main types of enzymes collaborate and are typically required. **β-1,4-endoglucanases** hydrolyze glycosidic bonds within the cellulose structure, resulting in oligomers of various lengths and producing free chain ends. These chain ends can then be broken down by **β-1,4-exoglucanases** or **cellobiohydrolases**, which release soluble cellobiose. Subsequently, cellobiose is further cleaved to yield glucose by **β-glucosidases** (Fig. 2) (Kao et al., 2016; Bui & Lee, 2015).

To achieve complete enzymatic digestion of lignocellulose, hemicellulases are also necessary. These enzymes are responsible for breaking down matrix polysaccharides like xylans, mannans, arabinans, and galactans (King et al., 2010).

i. Bacteria

Recent microbiology research and meta-omics analyses of bacteria have opened up new possibilities for exploring the genetic diversity of prokaryotic communities in lignocellulose-rich environments, such as compost. These techniques allow for the discovery of robust lignocellulose-degrading enzymes that could be valuable for industrial applications. Researchers have used comparative meta-transcriptomic analysis to identify highly expressed genes in compost-derived microbial communities capable of breaking down rice straw under challenging conditions (Simmons et al., 2014). These enzymes, especially those from *Micromonospora*, containing carbohydrate binding module (CBM) domains, show promise for improving lignocellulose deconstruction processes in high-temperature and high-solids conditions for biofuel production (Simmons et al., 2014). Moreover, certain soil bacteria, including Actinobacteria, α-Proteobacteria, and γ-Proteobacteria, are known lignin catalyzers, utilizing their peroxidases and laccases -lignin-modifying enzymes-, which are essential for lignin degradation (Cragg et al., 2015). In marine environments, wood-degrading prokaryotes can be divided into tunnelling and erosion bacteria, each with distinct ways of breaking down plant cell walls (Cragg et al., 2015). These bacteria primarily target cellulose and hemicellulose, leaving lignin relatively intact. In deepwater settings, pressure-tolerant bacteria form unique communities on exposed wood, distinct from those found in fecal material produced by wood-boring organisms (Cragg et al., 2015).

i. Archaea

Another Tree of Life's enormous branch, Archaea, found in composts, also have lignocellulose-degrading capabilities, particularly at high temperatures, and some Archaea have genes encoding laccase enzymes (Graham et al., 2011). Additional studies have to be conducted in order to elucidate more aspects of lignocellulose degradation by Archaea.

ii. Single celled eukaryotes and protists

Some single-celled eukaryotes and protists, such as *Dictyostelium* and *Chlamydomonas*, possess cellulases that help break down cellulose (Cragg et al., 2015). Similarly, the pathogenic oomycete *Phytophthora* produces various enzymes that target hemicellulose and cellulose, including enzymes from different GH (glycoside hydrolases) families (1, 5, 6, 7, 10) and AA9 and 10 (Auxiliary Activities — redox enzymes that act with GHs, often in a synergistic manner) (Cragg et al., 2015). In addition, up to 19 species of flagellate parabasalian and oxymonadid protists have been found in lower termites' digestive system, which phagocytize wood particles (Brune, 2014). These protists' enzymatic machinery consists of endoglucanases, GH7 cellobiohydrolases, β-glucosidases, xylanases, mannosidases, and arabinosidases, aiding in wood digestion (Brune, 2014).

iii. Fungi

Given that the resistance of lignin to breakdown is the main bottleneck in lignocellulose digestion, a plethora of white-rot fungi, like bacteria, have been reported to degrade lignin by employing different enzymes and catabolic pathways (Pollegioni et al., 2015). In typical white rot degradation, cellulases and enzymes that oxidize lignin components are produced, such as ligninase, manganese peroxidase, versatile peroxidase, or laccase (Pollegioni et al., 2015). In contrast, brown rot fungi, which have evolved from ancestral white rot fungi multiple times, have lost lignolytic enzyme systems and certain cellulases (Arantes & Goodell, 2014). Instead, a chelator-mediated Fenton (CMF) system that allows a non-enzymatic deconstruction of biomass provides brown rot fungi with advantages in ecological niches, enabling them to outcompete white rot fungi in degrading conifer wood (Arantes & Goodell, 2014).

iv. Animals

Various categories of animals capable of lignocellulose digestion caught the attention of many researchers. Invertebrates, including plant-parasitic nematodes, cockroaches, termites, insects, Gastropoda, Crustacea, and Annelida, possess distinct digestive mechanisms, which are essential for herbivorous and detritivorous animals (Cragg et al., 2015). Different organisms have acquired diverse lignocellulose deconstruction strategies, involving mutualistic host-symbiont interactions or endogenous processes. While some animals possess large digestive gut chambers for microbial gut symbionts, others exclusively rely on endogenous cellulases due to the absence of such chambers (Cragg et al., 2015). In most cases, mechanical breakdown of substrate by mouthparts or shells aids in overcoming lignin recalcitrance (Cragg et al., 2015).

Shipworms. Firstly, shipworms are wood-boring bivalves that consume wood particles using specialized cells in their gills housing endosymbiotic γ-proteobacteria (O'Connor et al., 2014). These bacteria produce lignocellulose-degrading enzymes, like glycoside hydrolases, GHs, and carbohydrate esterase families, as well as lytic polysaccharide monooxygenases, LPMOs (Cragg

et al., 2015). The unique setup of separation of bacterial residence from digestive process may allow shipworms to collect liberated sugars from wood without competition from gut microbes (Fig. 3). These symbiotic bacteria also fix nitrogen, aiding in nitrogen-poor wood environments (O'Connor et al., 2014).

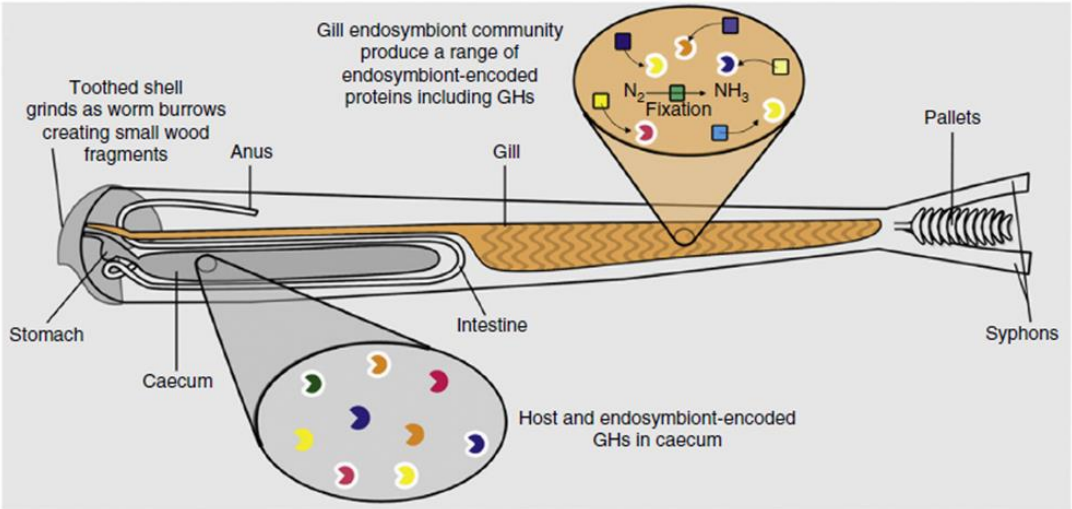

Figure 3. Lignocellulose degradation by shipworms. *These bivalves create small wood particles with their ridged shells as they burrow inside wood. Endosymbiotic bacteria in their gills produce enzymes transported from the gills to the gut, where they mix with host enzymes to digest the wood particles. (reproduced from Cragg et al., 2015)*

Termites. Meta-omics has provided insights into the complex interactions within gut-resident microbial consortia, with termite digestive systems serving as prime examples (Cragg et al., 2015). Termites are well-known for their symbiotic relationships with gut microbes, although these microbes are not solely responsible for digestion. Lower termites have three-part symbioses with host, bacteria, and protozoa, while higher termites have simpler two-way associations with host and bacteria and sometimes fungal organisms (Scharf, 2015). The host contributes with a toolkit of enzymes and maintains a conducive gut environment for digestion. Viewing termites as "holobionts," where host and symbionts are tightly connected, is crucial (Scharf, 2015). Termites use a combination of their own cellulases (from salivary glands and midgut) and microbial enzymes (from flagellates and bacteria in the hindgut) to break down cellulose, aided by oxidative breakdown of lignin (Fig. 4) (Cragg et al., 2015). Hemicellulases found in termites are mainly of bacterial origin, while some termite species cultivate *Termitomyces* fungi, which produce glycoside hydrolases (GHs) for breaking down polysaccharides (Cragg et al., 2015). Termite workers host bacteria that digest the oligosaccharides released by the fungus, facilitating efficient digestion of plant materials (Cragg et al., 2015). Recent advancements in omics research provides a comprehensive understanding of gut digestomes, essential for studying the termite holobiome for practical applications.

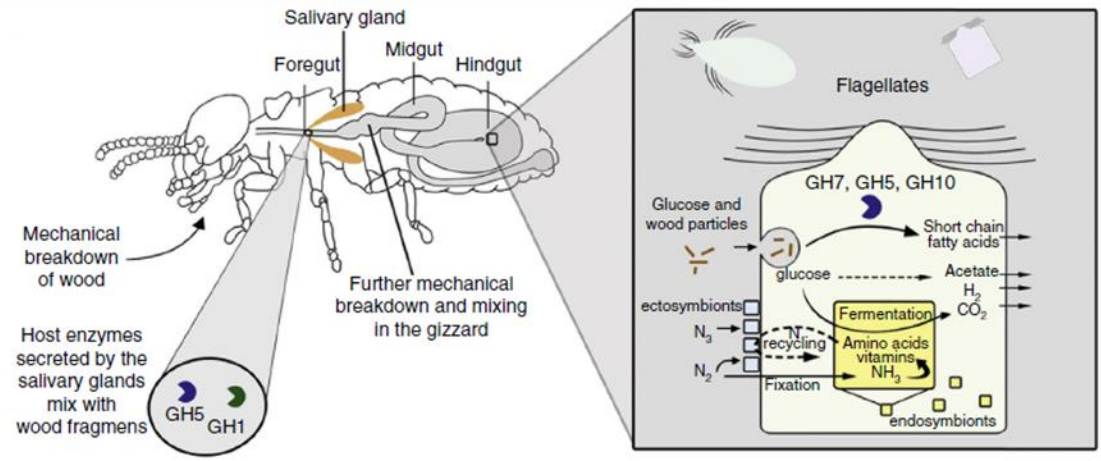

Figure 4. Lignocellulose degradation by termites. *In the termite's digestive system, wood particles mix with enzymes from salivary glands and are further broken down in the gizzard. Glucose is absorbed in the midgut, while partially digested wood particles proceed to the hindgut. There, cellulolytic flagellates engulf and break down the remaining polysaccharides using cellulases and hemicellulases in their digestive compartments. The resulting microbial fermentation products, mainly short-chain fatty acids, are absorbed by the termite host, and the lignin-rich residues are expelled as feces. (reproduced from Cragg et al., 2015)*

Vertebrates. Lignocellulose digestion is uncommon among vertebrates, but some, like beavers, pandas, and porcupines, consume significant amounts of lignocellulose in their diets. Nevertheless, the diverse microbiomes that enable lignocellulose digestion in vertebrates are currently under investigation (Cragg et al., 2015).

Lignocellulose Digestion in Crustaceans

Given that immense amounts of lignocellulose enter marine ecosystems, originating from sources like river estuaries and mangrove forests, it serves as a vital food source for various marine organisms, including a plethora of crustaceans (King et al., 2010). A recurring pattern in the digestion of lignocellulose by animals is the necessity for symbiotic relationships with microorganisms, which play a crucial role in providing the digestive capabilities (direct or indirect microbiome-dependent) to thrive on a lignocellulose-rich diet. On the contrary, wood-consuming crustaceans like *Chelura* and *Limnoria* lack resident microorganisms in their digestive tracts, making them an interesting model for studying enzyme function without microbial interactions (Fig. 5, Cragg et al., 2015). The sterile gut environment, which is unique in terms of gut physiology, offers opportunities for biotechnology, particularly in identifying enzymes and reaction conditions for lignocellulose degradation (King et al., 2010).

Based on research studies, the hepatopancreas in crustaceans plays a significant role in a plethora of processes such as synthesizing digestive enzymes, absorbing nutrients, storing carbon reserves, facilitating ion transport, maintaining osmotic balance, sequestering heavy metals, and generating the oxygen-binding pigment hemocyanin (King et al., 2010; Brunet et al., 1994).

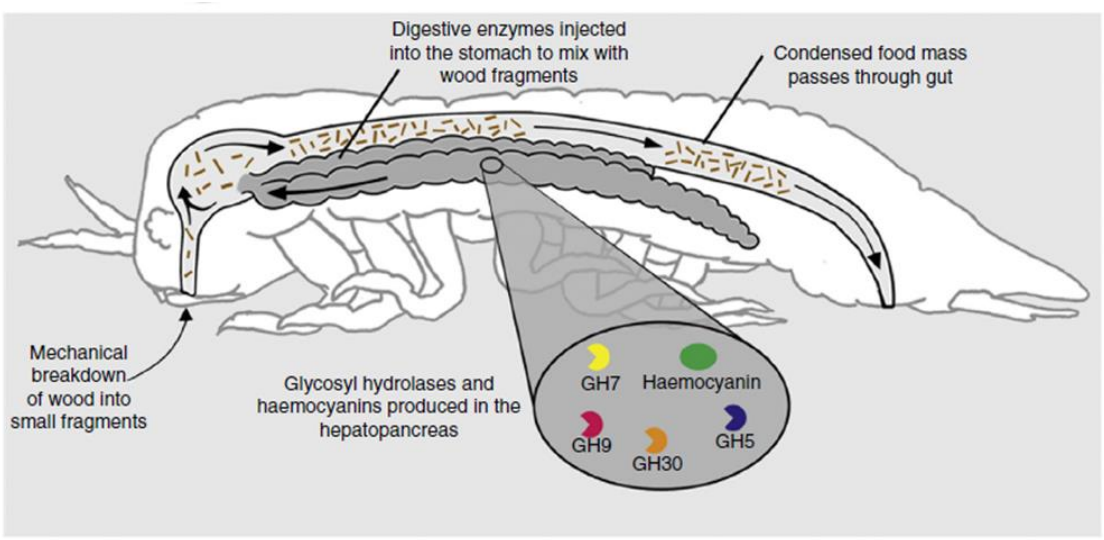

Figure 5. Lignocellulose degradation by Limnoriids. *They possess a straightforward, linear digestive tract featuring two paired hepatopancreas lobes or caeca, which converge with the stomach near the head region. As these crustaceans feed, they mechanically break down the wood into small pieces. In the stomach, these small wood fragments combine with the digestive enzymes released by the hepatopancreas. The wood fragments are compacted, and any parts that cannot be digested are expelled as fecal pellets. (reproduced from Cragg et al., 2015)*

The *Limnoria* hepatopancreas transcriptome is primarily composed of a few dominant gene sequences, including glycosyl hydrolases and hemocyanins (Fig. 7A, King et al., 2010). More specifically about GHs, the transcriptome contains sequences from 12 recognizable glycosyl hydrolase families, with the majority belonging to three families: GH7, GH9, and GH5 (Fig. 7B, King et al., 2010). These families are involved in lignocellulose digestion. GH9 genes encode β-1,4-endoglucanases and are found in arthropods like termites and crayfish. GH5 family genes likely encode β-1,4-mannanases involved in hemicellulose digestion and are found in aquatic animals like molluscs and arthropods. The most abundant GH7 family of glycosyl hydrolases, which encode cellobiohydrolases, are unexpectedly present in *Limnoria*, as they were previously only reported in fungi and protozoan mutualists in termite digestive tracts (King et al., 2010).

Through investigation of multiple aspects of lignocellulose breakdown by a GH7 cellulase from *Limnoria quadripunctata* (LqCel7B), it has been reported that this enzyme exhibits several noteworthy features that set it apart from typical fungal GH7 enzymes and reflect its adaptation to the marine environment (Kern et al., 2013). Some of LqCel7B's remarkable traits are the distinct protein sequence, lack of a carbohydrate-binding module (CBM) domain, high abundance in the gut lumen, efficient kinetic properties and salt tolerance (Kern et al., 2013). Overall, LqCel7B is an impressive candidate for potential utility in industrial processes, especially those involving challenging conditions like high salt concentrations (Kern et al., 2013).

Lignocellulose Digestion in Crustaceans: The case of *Parhyale hawaiensis*

Parhyale hawaiensis, one of the limited marine animal models available, stands out as the most suitable crustacean species for genetic research. An impressive arsenal of embryological techniques and functional genetic methods has been developed for *Parhyale*. Gene activity

can be effectively observed and altered in vivo using high-efficiency methods such as knock-down, knock-out, and knock-in approaches involving transposons, integrases and CRISPR/Cas (Kao et al., 2016). Additionally, the research community has been actively developing a comprehensive set of standardized genome-wide resources for this species.

Unpublished SEM (Scanning electron microscopy) data of *Parhyale*'s digestive system (kindly provided by V. Stamataki) confirm the absence of microorganisms inside hepatopancreatic glands (Fig. 6C-E), in agreement with previous reports from other crustaceans (King et al., 2010). Even after feeding on bacteria for a transient period of time, these can be found only in the stomach and hindgut (HG) lumen (Fig. 6F). Hepatopancreatic (midgut, MG) glands remain bacteria-free (Fig. 6F), as only irregular food particles can be detected like in standard conditions (Fig. 6E). There is an elaborate mesh of filaments between the hepatopancreas and the stomach, preventing bacterial entrance. Therefore, microorganisms have the ability to pass through stomach and hindgut lumen, but they are filtered away from midgut glands.

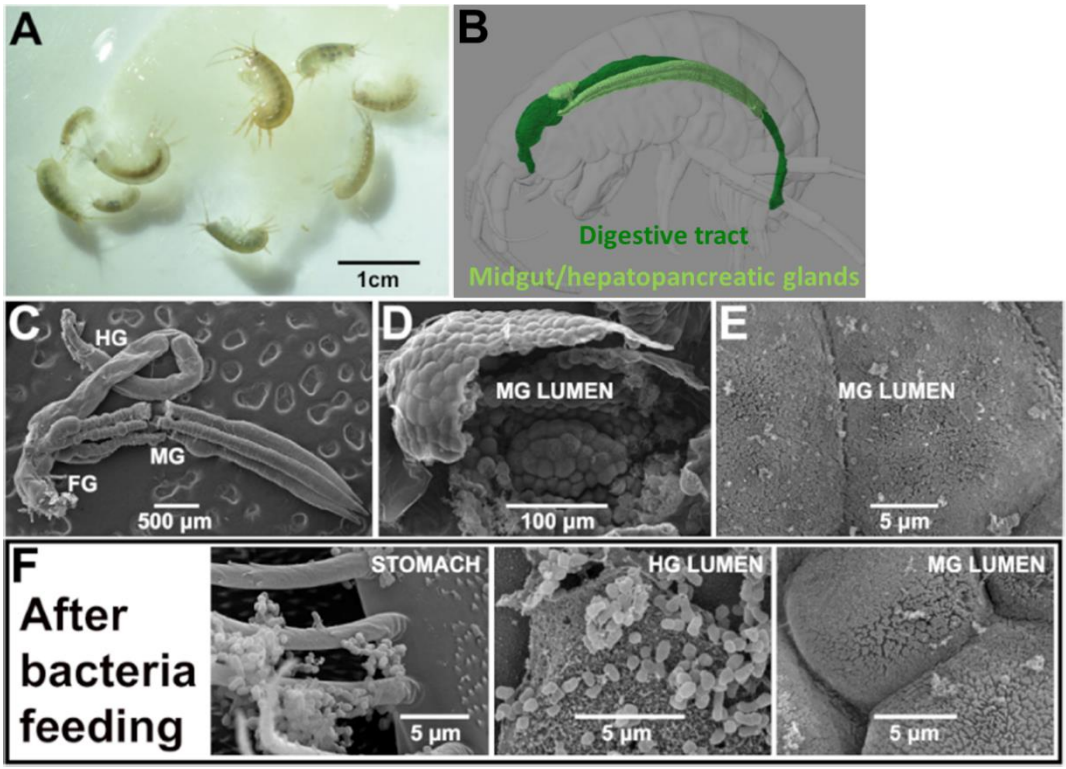

Figure 6. Parhyale microbiome-free digestive system*. (A) Parhyale feeding on cotton cellulose. (B) Micro-CT scan indicating the Parhyale digestive system in green. (C) Scanning electron micrograph (SEM) of dissected Parhyale gastrointestinal system. Abbreviations: Foregut (FG), Midgut glands (MG), Hindgut (HG). (D) SEM of the MG lumen and (E) higher magnification SEM showing the microbe-free apical microvilli of MG cells. (F) SEM images indicating the presence of bacteria only in the FG and HG, but not in the MG lumen, after bacteria feeding to Parhyale. Panels C-F kindly provided by Evangelia Stamataki and panel B by Carsten Wolff.*

Additionally, a comprehensive analysis of the *Parhyale* digestive midgut glands involved high-depth Illumina RNA sequencing and proteomics with Liquid Chromatography-Tandem Mass Spectrometry on the fluid, tissue and total components of dissected digestive glands. *Parhyale's* digestome predominantly features lignocellulolytic enzymes (Valia Stamataki, unpublished data). In Figure 7 C and D, the most abundant transcripts and proteins in the *Parhyale* digestive system are demonstrated. Based on these analyses, the most prevalent

enzymes include cellulases (comprising endoglucanases and exoglucanases / cellobiohydrolases), hemicellulases (specifically mannosidases), and ligninases (like hemocyanins) (personal communication V. Stamataki, J. Rallis). Of particular interest are the *Parhyale* exoglucanases / cellobiohydrolases belonging to the glycosyl hydrolase family 7 (GH7), because they are uniquely present in crustaceans, previously identified exclusively in fungi and protozoans. According to a phylogenetic analysis of *Parhyale* GH7s, there is high sequence similarity to the known GH7 genes in *L. quadripunctata* and the amphipod *C. terebrans* (Kao et al., 2016).

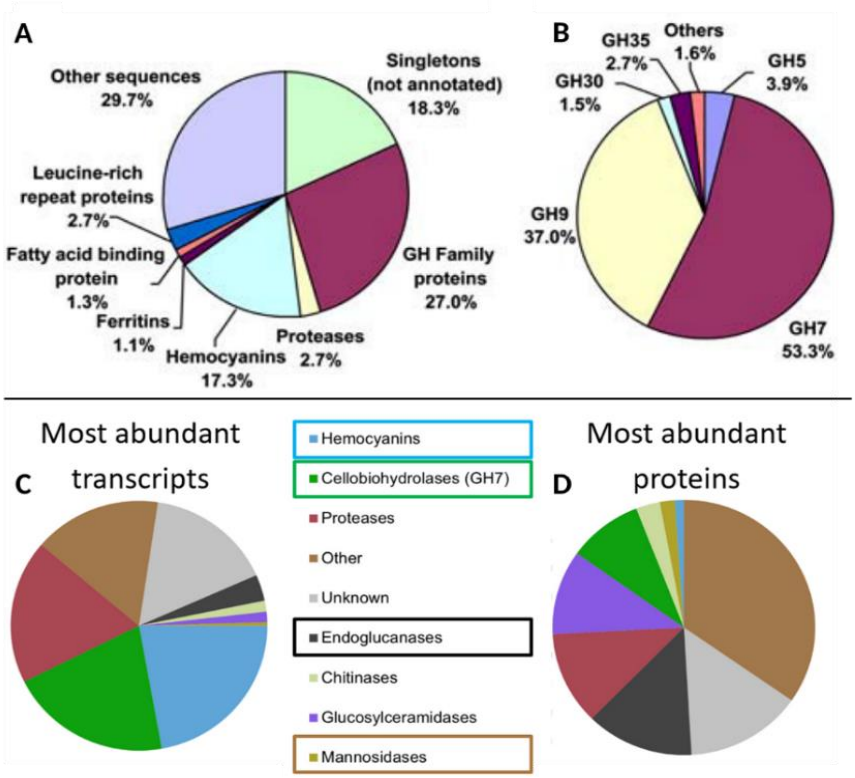

Figure 7. Multi-omics analyses of crustacean hepatopancreatic glands*. Pie charts showing (A) the relative abundance of expressed gene families in the Limnoria hepatopancreatic transcriptome, (B) the relative abundance of expressed glycosyl hydrolase families in the Limnoria hepatopancreatic transcriptome, (C) the top expressed gene families in the Parhyale hepatopancreatic transcriptome, and (D) the relative abundance of protein families in the Parhyale hepatopancreatic secreted proteome. Panels A and B reproduced from King et al., 2010 and panels C and D kindly provided by Evangelia Stamataki and John Rallis.*

Multiple clustered *Parhyale* GH7 genes

In the most recent Phaw5.0 assembly of the *Parhyale* genome, three cellobiohydrolases / β-1,4-exoglucanases GH7 genes, were identified, sharing similarity from 44% to 67%. In transcriptomic and proteomic analysis, striking differences were observed in the expression levels of these genes, with the first *Parhyale* GH7 gene exhibiting significantly higher expression, roughly two orders of magnitude more than the second GH7 gene and nearly three orders of magnitude higher than the third GH7 gene (Valia Stamataki, personal communication). Given these outcomes, the first target of a CRISPR/Cas9 knock-out genetic analysis was the most prevalent GH7 gene. Surprisingly, when the mutant lines were genotyped, more sequence variants were discovered than anticipated. This unexpected result motivated a detailed exploration of the assembled *Parhyale* genome, where multiple

fragmented and unannotated GH7-like genes were identified within the same scaffold as the highly expressed one. To analyze the genomic organization of the locus more thoroughly, high-molecular-weight genomic DNA was prepared from a single adult *Parhyale* male and sequenced, using Oxford Nanopore Technologies. Alignment of long overlapping reads revealed the existence of nine clustered genes, labeled GH7a to GH7i, with amino acid similarities ranging from 95% to 99% (Fig. 8). Additionally, it is noteworthy that the selected sgRNA perfectly matches seven out of these nine genes, which explains the sequence variability observed in the CRISPR/Cas9 knock-out experiments.

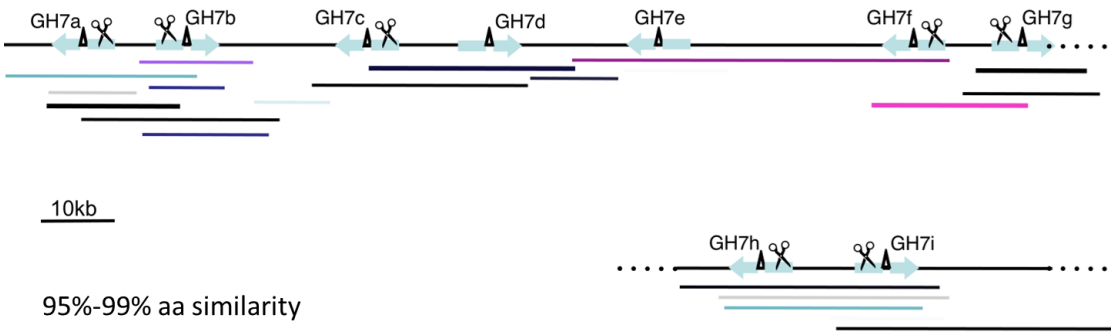

Figure 8. Genomic organization of Parhyale GH7 genes. *Revised annotation of nine paralogous GH7a-GH7i genes identified by long-read next-generation sequencing organized in two (so far unlinked) clusters and exhibiting 95%-99% similarity at the amino acid level. Cyan block arrows depict the two exons and orientation of each gene, black triangles mark the locations of introns, and lines of varying colors represent the assembled long reads. Scissors indicate perfect matches to the sgRNA used in the CRISPR knock-in experiments. Figure kindly provided by Evangelia Stamataki.*

Aim of the study

Based on recent reports about the capacity of certain marine crustaceans for the microbe-free autonomous digestion of lignocellulosic biomass, my Master thesis focused on the functional genetic and biochemical analyses of GH7 cellobiohydrolases produced by the genetically tractable crustacean model *Parhyale hawaiensis*.

Following a multi-omics analysis of the *Parhyale* digestome, previously carried out in the lab, my primary goals were the in vivo and in vitro analyses of glycosyl hydrolase activity in wildtype and transgenic *Parhyale* overexpressing a GH7 cellobiohydrolase. In order to achieve these objectives, I created transgenic lines encoding heat-inducible, tagged versions of a *Parhyale* GH7 protein for their *in vivo* fluorescent labelling, isolation via affinity purification, biochemical characterization and comparison to a recombinant fungal cellobiohydrolase. Furthermore, a CRISPR knock-in approach has been employed for the tagging of an endogenous *Parhyale GH7* gene for midgut-specific expression and future characterization. This work paves the way to exploit cellulolytic enzymatic cocktails from *Parhyale* for significant biotechnological applications.

Materials and Methods

Design and generation of plasmid constructs

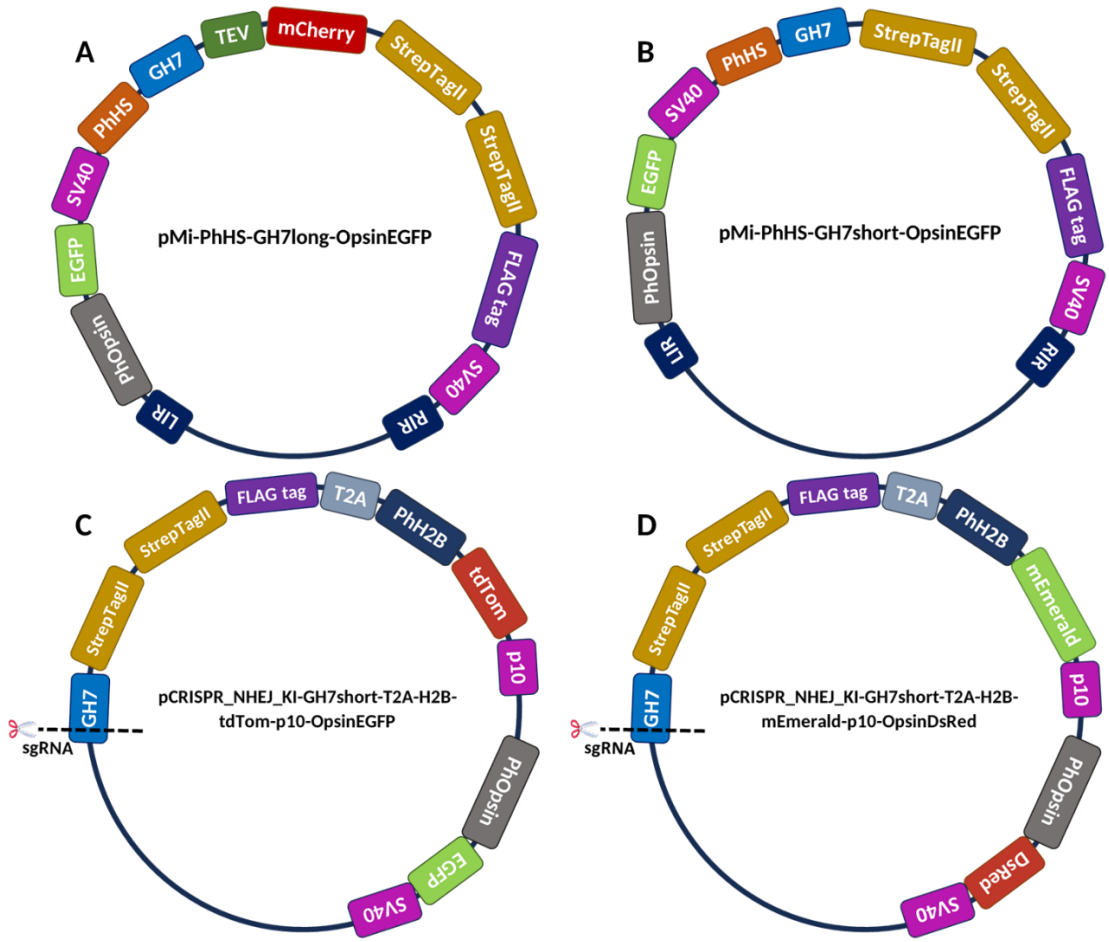

Figure 9. Schematic representation of plasmid constructs created in this project. *(A) pMi-PhHS-GH7long-OpsinEGFP, (B) pMi-PhHS-GH7short-OpsinEGFP, (C) pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP, (D) pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed. LIR and RIR: left and right Minos-transposase Inverted Repeats, respectively. PhOpsin: PhOpsin1 regulatory element. EGFP: enhanced green fluorescent protein. SV40: transcriptional terminator. PhHS: heat-inducible cis-regulatory sequence. GH7: PhGH7a open reading frame without stop codon. TEV: cleavage site for Tobacco Etch Virus (TEV) protease. mCherry: fast maturing, monomeric red fluorescent protein. StrepTagII-StrepTagII: eight amino acid double motif with affinity to streptavidin connected by a linker. FLAG tag: artificial tag DYKDDDDK. T2A: 2A viral self-cleaving peptide. PhH2B: Parhyale H2B histone. tdTom: exceptionally bright red fluorescent protein. p10: transcriptional terminator. mEmerald: fast maturing, bright green fluorescent protein. DsRed: DsRedT1 red fluorescent protein.*

A. pMi-PhHS-GH7long-OpsinEGFP

This plasmid is used for overexpression of GH7 with a marker gene and appropriate tags for protein purification and is constructed as follows:

Initially, with a three-fragment Gibson assembly, the cassette containing the heat-inducible *cis*-regulatory elements, the coding sequence for GH7, the sequence for TEV protease cleavage, mCherry, Twin-Strep-tag and FLAG tag is assembled in the intermediate vector pSL-PhHS-DsRed.

The following fragments were generated:

The GH7 sequence was PCR-amplified from a plasmid containing the 1374bp GH7a cDNA using the primers GH7_GBS_PhHS_F and GH7-C_R. (The first primer introduced a Homologous Arm (HA) to PhHS at the 5'end of GH7. The second primer lacked the stop codon at the end of GH7 and was homologous to gBlock-long C-terminal tag.

GH7_GBS_PhHS_F (40nt) 5'-CACAGAACAATCCACAACcATGAAGTCCCTGTTCTTCGTC-3'
Homology Arm

GH7-C _R (20nt) 5'-CTGGATGGGTCCGATCTTGA-3'

Long C-terminal tag (943 bp, hereafter referred to as long tag or, briefly, long), consists of a 7-amino acid linker, cleavage site for Tobacco Etch Virus (TEV) protease, mCherry monomeric red fluorescent protein, a 5-amino acid linker, StrepTagII- StrepTagII or Twin-Strep-tag including the eight amino acid motif (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) two times in series connected by a 12-amino acid linker and is accordingly composed of 28 amino acids, a 5-amino acid linker and FLAG tag, an artificial tag, also called DYKDDDDK-peptide tag. It was synthetically generated as a gBlock by IDT - Integrated DNA Technologies. gBlocks Gene Fragments are double-stranded DNA fragments designed for affordable and easy gene construction or modification. The gBlock-long fragment had homologous arm to GH7 (C-end without stop) and homologous arm to SV40 (in plasmid vector pSL).

pSL-PhHS-DsRed vector, in which DsRed is removed with NcoI-NotI.

After the 3-fragment Gibson assembly, pSL-PhHS-GH7long is created. Next, the GH7long cassette was cloned in pMi-OpsinEGFP. Both pSL-PhHS-GH7long and pMi-OpsinEGFP were digested with AscI. Then, the ligation between PhHS-GH7long fragment (5040 bp) and linear pMi-OpsinEGFP vector created the final construct, **pMi-PhHS-GH7long-OpsinEGFP** (Fig. 9 A), ready to use in microinjections.

B. pMi-PhHS-GH7short-OpsinEGFP

This plasmid is used for heat-inducible overexpression of GH7 protein with Twin-Strep-tag and FLAG tag for protein purification and is constructed the same way as pMi-PhHS-GH7long-OpsinEGFP.

Short C-terminal tag (hereafter referred to as short tag or, briefly, short), with 193bp length, consists of a 5-amino acid linker, a StrepTagII- StrepTagII or Twin-Strep-tag which includes the eight amino acid motif (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) two times in series connected by a 12-amino acid linker and is accordingly composed of 28 amino acids, a 5-amino acid linker and FLAG tag, an artificial tag, also called DYKDDDDK-peptide tag. Short tag was synthetically generated as a gBlock by IDT- Integrated DNA Technologies. The gBlock-short fragment had homologous arm to GH7 (C-end without stop) and homologous arm to SV40 (in plasmid vector pSL). The following fragments: GH7 sequence with HA arms, short tag with HA arms and pSL-PhHSfragment, (from pSLPhHS-DsRed after removal of DsRed with NcoI-NotI) were used in a 3-fragment Gibson assembly to create the intermediate pSL-PhHS-GH7short. Ultimately, the cassette containing HS regulatory element and GH7short was isolated as an AscI fragment and cloned in AscI site of pMi-OpsinEGFP vector thus creating the final construct, **pMi-PhHS-GH7short-OpsinEGFP** (Fig. 9 B), ready to use in microinjections.

C. pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP

i. This plasmid is used for expression and monitoring of *Parhyale* GH7 in native conditions with a marker gene and appropriate tags.

ii. The initial step for the generation of pCRISPR_NHEJ_KI-DS3-T2A-H2B-Ruby2-p10, via PCR (5 cycles annealing at 66 °C and 30 cycles at 72 °C) was the amplification of the p10 terminator (685bp) from plasmid pTGV9 (kindly provided by M. Grillo) with the following primers:

NotI-p10-F	aattGCGGCCGCCTAGAATG
p10-SalI-R	atttgtcgaCGTTAACTCGAATCGCTATCC

Then, PCR product was digested with NotI/SalI and cloned into NotI/SalI-cut pCRISPR_NHEJ_KI-DS3-T2A-H2B-Ruby2 (4123bp) to replace PhDlle-3UTR with p10. Afterwards, pCRISPR_NHEJ_KI-DS3-T2A-H2B-tdTom-p10 was created, through replacement of Ruby2 with tdTom by digesting pBACretro-PhHS-tdTom-attB with XhoI/NotI, purifying with gel extraction of the 1437bp TdTom band and cloning into XhoI/NotI-digested pCRISPR_NHEJ_KI-DS3-T2A-H2B-Ruby2-p10 (4077bp). The next step was the production of pCRISPR_NHEJ_KI-DS3-T2A-H2B-tdTom-p10-OpsEGFP. This was achieved by digestion of pCRISPR_NHEJ_KI-DS3-T2A-H2B-tdTom-p10 with SalI-MluI (5491bp) and insertion of SalI-MluI Opsin-GFP fragment (2617bp) from pMi-PhOps-EGFP. Finally, the goal was the generation of pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsEGFP. In order to accomplish that, it was necessary to digest pCRISPR_NHEJ_KI-DS3-T2A-H2B-tdTom-p10-OpsEGFP with SphI-BspEI (8083bp) to remove DS3+PAM-Dll coding and replace it with PCR (5 cycles annealing at 65 °C and 30 cycles at 72 °C) fragment GH7short (1517bp) with primers containing SphI in the F primer and and AgeI in R primer (there was a BspEI site in GH7 coding sequence and AgeI is compatible with BspEI). The inclusion of 2nt in GH7short-AgeI-R (noted in red) was required in order to restore the correct open reading frame after GH7short insertion.

GH7short-SphI-F	5'-ATTAgcatgCATGAAGTCCCTGTTCTTC-3'
GH7short-AgeI-R	5'ATTAaccggtgaCTTGTCGTCGTCATCCTTG-3'

After all these steps, **pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP** (Fig. 9 C), was ready to use in microinjections.

D. pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed

As an alternative to the pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP, pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed was generated by replacement of tdTom with mEmerald and OpsinEGFP with OpsinDsRed. In order to achieve this, first tdTom was replaced with mEmerald. Therefore, pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP was digested with XhoI to linearize CRISPR_KI vector. Then, partial digestion with NotI (normally, 2 cut sites in pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP) was conducted by addition of 0.5 enzyme unit/ul and incubation for 15 min at 37 °C. Afterwards, vector without tdTom sequence (8166bp) was extracted. Ligation of this cut (XhoI/NotI) vector (lacking tdTom) with mEmerald sequence (720 bp), already digested with XhoI/NotI, led to the creation of pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinEGFP. Subsequently, OpsinEGFP was substituted with OpsinDsRed, though, initially,

digestion with EcoRV to linearize CRISPR_KI vector and, then, partial digestion with NotI (normally, 2 cut sites in pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinEGFP) was conducted by addition of 0.6 enzyme unit/ul and incubation for 15 min at 37 °C and vector without OpsinGFP (6550 bp) was extracted. Ligation of this cut (EcoRV/NotI) vector (lacking OpsinGFP) with OpsinDsRed sequence (2301 bp), already digested with EcoRV/NotI from vector pMi-Ops-DsRed, led to the creation of **pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed** (Fig. 9 D), ready to use in microinjections.

Collection and Microinjections of Early-Stage *Parhyale* Embryos

The collection and microinjection of *Parhyale* embryos, were performed following procedure steps as previously described by Kontarakis & Pavlopoulos, 2014.

The day before microinjections, it was required to collect over 100-150 WT *Parhyale* pairs using a plastic Pasteur pipette with a large opening from the main cultures into 150 mm Petri dishes with artificial sea water (ASW) and few corals. A lot of the collected Parhyale pairs were separated the day of injections and gravid females with eggs in their ventral brood pouch were easily identifiable by eye. Gravid females were transferred into a different Petri dish with filtered artificial sea water (FASW) and anaesthetized by bubbling CO2 gas for 30–60 s. All pregnant females were dissected using tongs into a Petri dish filled with filtered artificial seawater with antibiotics and antimycotics (FASWA), then 1-cell embryos were separated and collected into a different Petri dish with FASWA. Dissected females were transfer into another Petri dish with ASW, until they were fully awake and mobile, in order to return them into the main *Parhyale* cultures. These steps were repeated every 3-4 h with the aim of collection adequate number of 1-cell WT embryos for microinjections.

A. Heat inducible system using Minos transposase

The mRNA encoding Minos transposase was prepared according to Kontarakis & Pavlopoulos, 2014, and stored at -20 °C in 100% isopropanol or ethanol. To precipitate and use Minos transposase mRNA, it was needed to centrifuge it for 30 min at 4°C and at 13.000rpm, remove 100% isopropanol or ethanol, rinse pellet with ice-cold 70% ethanol, centrifuge for 10min at 4°C and at 13.000rpm, remove 70% ethanol, resuspend pellet with RNase and DNase free ddH2O (~8ul), measure RNA concentration via Nanodrop and, finally, store at -80°C until use.

The injection mix for obtaining overexpression transgenic lines included 300 ng/uL Minos mRNA, 500 ng/uL pMi-PhHS-GH7long-OpsinEGFP or pMi-PhHS-GH7short-OpsinEGFP plasmid, 1-2/10 dilution of stock inert dye phenol red, which allowed better visualization of the injected material and RNase and DNase free ddH2O until the final volume was achieved (usually 5 uL). The mix was prepared in a tube and centrifuged for at least 30 min at 13000 rpm for any insoluble material to precipitate and not clog the microneedle. The supernatant was transferred carefully in another clean microcentrifuge tube and during the day of injections, mix was constantly on ice. The remaining mix was stored at -80˚C and reused, but could be frozen and thawed max 2 times.

B. CRISPR NHEJ knock-in system

The sgRNA used in these experiments was prepared, using Ambion MAXIscript SP6 Transcription Kit (catalog number AM1310, Thermo Fisher Scientific) and stored at -20 ℃ in 100% isopropanol. To precipitate and use sgRNA, it was needed to centrifuge it for 30min at 4^{o}C and at 13.000rpm, remove 100% isopropanol, rinse pellet with ice-cold 70% ethanol, centrifuge for 10min at 4^{o}C and at 13.000rpm, remove 70% ethanol, resuspend pellet with RNase and DNase free ddH2O (~6ul) measure RNA concentration via Nanodrop and, finally, store at -80^{o}C until use.

The injection mix for the CRISPR NHEJ knock-in experiment contained 400 ng/ul Cas9 protein, 40 ng/ul sgRNA, 10 ng/ul pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP or pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed plasmid, 1-2/10 dilution of stock inert dye phenol red and RNase and DNase free ddH2O until the final volume was achieved (usually 5 uL). Initially, for the preparation of CRISPR NHEJ knock-in mix, it was necessary to combine 400 ng/µl Cas9 protein and 40 ng/µl sgRNA in a microcentrifuge tube and incubate it for 5 min at 37C to form the sgRNA/Cas9 complexes. Afterwards, it was transferred to ice for 1 min, spinned down for 30 sec, transferred back to ice and then the plasmid template (10 ng/ul) for NHEJ-KI, phenol red and RNase and DNase free ddH2O were added. It was mixed gently by flicking the tube and centrifuged for at least 30 min at 13000 rpm for any insoluble material to precipitate and not clog the microneedle. The supernatant was transferred carefully in another clean microcentrifuge tube and during the day of injections, mix was constantly on ice. The remaining mix was stored at -80˚C and reused, but could be frozen and thawed 2 times, at the most, to avoid RNA degradation. To increase the efficiency of the knock-ins, at the end of each injection's session (i.e., 1-2 hours from the beginning of injections), screening and collection only of embryos, that were still at the 1-cell stage was suggested.

Henceforward, the same experimental procedure was followed for injecting both heat inducible and CRISPR NHEJ knock-in microinjection mixes. Microneedles from borosilicate glass capillaries were prepared using a Sutter P-87 puller. Furthermore, 2% agarose in FASW steps were already prepared in order to provide a convenient solution for immobilizing *Parhyale* embryos, which are extremely sensitive to desiccation, while keeping them wet. Microinjection needle was loaded with approximately 1µl of the injection mix using a microloading pipette tip (Eppendorf) and the filled needle was mounted onto the needle holder of a Narishige micromanipulator at a small downward angle. For injections, the FemtoJet 5247 Microinjector (Eppendorf) and an upright stereoscope setup were used. The embryos, with the aid of a cut pipette tip coated with bovine serum albumin (BSA) to prevent eggs from sticking to the plastic, were placed on an 2% agarose step that was covered with FASWA throughout the process. After each injection round the embryos were transferred with a cut pipette tip coated with BSA in a 35 mm cell culture petri dish with FASWA and kept at 26˚C. Renewal once per day of FASWA of each petri dish with injected embryos and removal of dead ones were of vital importance in order to enhance embryos' viability.

Screening of transgenic *Parhyale* embryos

The screening of *Parhyale* embryos was performed under Leica M205 FA Fluorescence stereomicroscope using DSR and GFP filters. Transgenesis marker, PhOpsin regulatory element combined with a fluorescent protein, GFP or DsRed, could be detected 10 days post injection, with the aid of GFP and DSR filters, respectively. In addition to that, tdTom and mEmerald marker genes for visualizing native GH7 expression pattern, could be detected 10 days post injection, utilizing the same filters. The 10^{th} day of development was selected as the ideal day of screening both transgenesis marker and expression reporter of GH7, because, at this point, embryo's eyes were entirely developed and visible by stereoscope and the expression of GH7 in digestive glands was increasing. Injected embryos (G0), that were screened on the 10^{th} day and found positive for EGFP fluorescence in their eyes, driven by the PhOpsin1 regulatory element, were labeled appropriately and kept in individual petri dishes containing artificial sea water which was renewed once or twice a week. After reaching sexual maturity, about 2 months after hatching, they were crossed with WT *Parhyale* of the opposite sex and their progeny were screened following the same strategy. On the other hand, injected embryos of CRISPR NHEJ knock-in approach, that were screened the 10th day and found positive for either EGFP (or DsRed) fluorescence in their eyes or tdTom fluorescence in their digestive glands (or mEmerald) or both of them, were labeled appropriately and kept, individually, in petri dishes containing artificial sea water which was renewed once or twice a week. After reaching sexual maturity, they were crossed with WT *Parhyale* of the opposite sex and their progeny were screened following the same strategy.

Bradford Protein assay

Before the protein analysis experiments, all protein samples were measured using Bradford assay and equal total protein concentration was loaded for each sample. First of all, 0.2 ug/ul BSA stock solution and Bradford reagent 1:4 diluted with ddH2O, enough for the number of samples, plus one blank, plus five for BSA Standards, were needed for the measurement. 1 ml of diluted Bradford reagent was added to set up blank. Then, 5 BSA standards were set up by adding 10, 15, 20, 25, and 30ul of BSA stock solution up in 1ml Bradford reagent. For each one of the protein samples, between 1-20ul protein sample was added in the Bradford reagent. It was important to add enough protein sample to the Bradford reagent, so that the OD measured would be into the range that the BSA protein standards gave, for accurate estimation of protein concentration. Finally, 250ul of each reaction mix was transferred in a 96-well plate in triplicates and absorbance at 595nm was measured, using a plate reader. Based on the linear part of the curve that the BSA standards were giving, the protein concentration of each sample could be estimated.

SDS-PAGE, Coomassie staining and Western blot

A. SDS-PAGE gel electrophoresis

The first step was to setup the casting frames (clamp two glass plates in the casting frames) on the casting stands, making sure there was not any leakage. For the preparation of 10%

polyacrylamide separating and 4% stacking gel, the following ingredients and quantities were used (Table 1), noting that APS and TEMED must be added right before each use.

10% Separating Gel	Final volume 10ml/gel	4% Stacking Gel	Final volume 5ml/gel
4x Separating mix	2.5 ml	**4x Stacking mix**	1.25 ml
30% Acrylamide	3.33 ml	**30% Acrylamide**	0.66 ml
H20	4.06 ml	**H20**	3.046 ml
10% APS (ammonium persulfate)	100 ul	**10% APS** (ammonium persulfate)	40 ul
TEMED (Tetramethylethylenediamin)	10 ul	**TEMED** (Tetramethylethylenediamin)	4 ul

Table 1. Recipes for separating and stacking SDS-PAGE gels

Appropriate amount of separating gel solution was pipetted into the gap between the glass plates. To make the top of the separating gel be horizontal, 1ml isopropanol was added on top of the gel, before polymerization of the acrylamide. After polymerization of the acrylamide, isopropanol was discarded, appropriate quantity of stacking gel was added until an overflow, the well-forming comb was inserted without trapping air under the teeth and the gel was left for 20-30min to gelate. After the complete gelation of the stacking gel, the glass plates were taken out of the casting frame and set them in the cell buffer tank. 1x running - electrophoresis buffer (10x stock solution containing dissolved 30 g of Tris base, 144 g of glycine, and 10 g of SDS in 1000 ml of H2O) was poured into the inner chamber and kept pouring after overflow until the buffer surface reached the required level in the outer chamber. Then, the comb was extracted out of the gel and the wells were rinsed with running buffer, using a syringe.

For the sample preparation, mixing of samples with appropriate amounts 5x loading buffer (10ml of 5x stock solution including 5ml Glycerol, 3ml 1 M Tris-HCl pH 6.8, 1ml 10% SDS, 0.005% Bromophenol blue, 750ul H20 and 25 ul β-mercaptoethanol freshly added in each 1ml aliquot) and heating them for 5 min at 95 °C were required. Prepared samples and 3ul of NEB prestained protein ladder were loaded to separate wells. Then, the anodes were connected, appropriate volt was set (80V first and after passing stacking gel raise at 100-110V) and the electrophoresis began, when everything was ready. As for the total running time, SDS-PAGE running was terminated when the downmost sign of sample buffer almost reached the foot line of the glass plate, about 1.5h duration. Afterwards the SDS-PAGE, protein gel was either developed with Coomassie staining or used for Western Blot.

B. Coomassie staining

Staining of protein gels with “Blue Silver” Coomassie Colloidal Blue Stain is a common procedure to visualize proteins resolved by SDS-PAGE. It is highly sensitive and is suitable for

long-term storage of the gels. After the electrophoresis, the gel was placed in a plastic tray containing gel fix solution (30% Methanol, 10% Acetic Acid), incubating at RT with shaking for 30 min to 1 h in order to fix the proteins. After fixation, the gel washed 4 times with distilled water, 15 minutes for each wash. For staining, the staining solution was added, containing 0.12% Coomassie Blue G-250 dye, 10% ammonium sulfate, 10% phosphoric acid and 20% methanol, around 5 times of the volume of the gel for an overnight incubation at RT in continuous motion. Finally, gel destaining with distilled water 3-4 times was necessary, until the desired contrast is achieved.

C. Western Blot

Western Blotting refers to the electrophoretic transfer of proteins from sodium dodecyl sulfate polyacrylamide gels to sheets of PVDF or nitrocellulose membrane, followed by immunodetection of proteins using antibodies with chemiluminescent detection. Firstly, sufficient 1x transfer buffer was prepared (100ml of 10x stock solution containing dissolved 24.7 g of Tris base and 112.6 g of glycine in 1000 ml of H2O is added to 700 ml H2O and 200 ml Methanol) to fill the transfer tank, plus an additional volume to equilibrate the gel and membrane, and wet the filter paper and foam pads. After SDS-PAGE electrophoresis, protein gel was removed from its cassette and any stacking gel and wells were discarded. The gel, filter papers, foam pads and membrane were immersed in ice-cold transfer buffer to equilibrate for a few minutes. It was important to set up properly the transfer cassette, by opening the cassette holder and placing in this specific order a foam pad on one side (black) of the cassette, two sheets of filter paper on top of the pad, the protein gel on top of the filter paper, the membrane on top of the gel, two sheets of filter paper on top of the stack, the second foam pad on top of the filter papers and then closing the cassette. To ensure an even transfer, the removal of air bubbles between layers by rolling a stripette over the surface of each layer in the stack was crucial. The cassette holder was placed in the transfer tank so that the gel side of the cassette holder was facing the cathode (–) and the membrane side was facing the anode (+). Adequate ice-cold transfer buffer was added into the tank to cover the cassette holder, an already-freezed ice holder was inside the tank and whole tank was constantly on ice. The black cathode lead (–) and the red anode lead (+) were connected to the correct positions and transfer procedure was executed at constant 400 mA for 1h. After the transfer was complete, the cassette holder was removed from the tank and the transfer stack using forceps was carefully disassembled in order to extract membrane. To achieve blocking, the membrane was placed in a container with 5% REGILAIT milk in TBST (500ml TBST contained 10 ml 1 M Tris-HCl pH 7.5, 9 ml 5M NaCl, 2.5 ml 10% Tween-20 and H2O up to final volume) and incubated in continuous motion for 1h at RT. Afterwards, milk was removed and 1:5.000 diluted primary antibody a-mCherry (rabbit polyclonal antibody, catalog number 600-401-P16, Rockland) in 1% REGILAIT milk was added for an overnight incubation at 4 °C, constantly shaking. Subsequently, the primary antibody was collected and membrane was washed 3 times for 10 min with TBST buffer. The next step was the incubation with 1:10.000 diluted mouse anti-rabbit secondary antibody (catalog number sc-2357, Santa Cruz) in 1% REGILAIT milk with agitation for 1 hour at RT. After removal of secondary antibody, 3 washes with TBST buffer for 10 min each was required to rinse the membrane. Ultimately, the chemiluminescent detection was conducted by SuperSignal West Pico PLUS Chemiluminescent Substrate (catalog number

34580, Thermo Fisher Scientific), according to manufacturer instructions, and then imaging was performed by ChemiDoc Imaging System (Bio-Rad). Exposure time differed and was determined for each gel until the desired signal intensity was achieved.

Protein purification

For sample preparation, 6h heat shock at 37 ℃ and 9h recovery at RT of 23 medium-size animals was necessary. Afterwards, *Parhyale* adult digestive system was dissected, using tongs and a stereoscope set up, in 60 mm tissue culture Petri dishes coated with SYLGARD and filled with filtered artificial sea water and then digestive glands were transferred in 500ul of extraction buffer containing 150mM NaCl, 100mM Tris-HCl pH 8 and 1x protease inhibitors cocktail mix (catalog number P8465, Sigma). Additionally, the remaining body of each dissected *Parhyale*, without the head, was transferred to 500ul extraction buffer. Proteins were extracted by grinding with pestle and then the samples were centrifuged at 4℃ at 4000 g for 20 min. Supernatant was transferred in a clean microcentrifuge tube. For the protein purification, MagStrep “type3” XT beads (catalog number 2-4090-002, IBA Lifesciences) were utilized, according to their suggested protocol. MagStrep “type3” XT beads were supplied as 5% (v/v) suspension, which meant that 20 µl of the homogenous suspension contained 1 µl magnetic beads. Per µl beads up to 0.85 nmol Strep-tagII or Twin-Strep-tag fusion protein could be purified, which corresponded to 25.5 µg of a 30 kDa protein in which the Twin-Strep-tag led to higher yields due to the higher binding affinity of Strep-TactinXT. First of all, the required amount of beads, 3 ul beads or 60 ul homogenous suspension, was pipetted into a reaction tube (Protein LoBind Tubes, Eppendorf) and, then, placed on the magnetic separator to separate the beads and remove the supernatant. Equilibration of beads was achieved with 600 µl 1x Buffer W (100mM Tris-Cl, 150mM NaCl, 1mM EDTA, pH 8). The beads were separated in the magnetic separator and the supernatant was removed. This step was repeated two times. Immediately prior to affinity purification, the sample was centrifuged for 20 min at 4,000g to remove any cell debris or aggregated protein. The magnetic beads were resuspended with the appropriate volume of the cleared sample containing the target protein and the mix was incubated for 30 minutes on ice. Vortexing three to four times during incubation to resuspend the beads was needed. The reaction tube was placed in the magnetic separator and the supernatant was carefully removed. With the magnet removed, 300 µl 1x Buffer W were added. Vortexing shortly and placing the reaction tube in the Magnetic Separator to collect the beads were the next moves. The supernatant was extracted and these steps were repeated two times. To accomplish native elution, the reaction tube was removed from the Magnetic Separator and 75 µl 1x Buffer BXT were added (100mM Tris-Cl, 150mM NaCl, 1mM EDTA, 50mM biotin, pH 8) and vortex. 10 min incubation and vortexing two to three times during this to resuspend the beads were needed. The reaction tube was placed in the Magnetic Separator and the supernatant was transferred containing the target protein into a new reaction tube. This step was repeated once for higher recovery. At every step of this experimental procedure (grinding-protein extraction, input, supernatant, washes, elution), it was crucial to keep a portion of each reaction mix for further analysis. Finally, the purity was analyzed via SDS-PAGE, Coomassie and Western blot and quantified according to Bradford using BSA as standard.

Biochemical assays

A. 3, 5-dinitrosalicylic acid (DNS) assay - Detection of reducing sugars

For the substrate preparation, Avicel (Avicel® PH-101, catalog number 11365, Sigma) suspension in sodium acetate buffer (pH 4.8) was needed. 0.1 M sodium acetate buffer pH 4.8 was prepared and, then 1.25% (w/v) Avicel was added and mixed well to make the Avicel suspension in sodium acetate buffer. For the preparation of 3, 5-dinitrosalicylic acid (DNS) reagent with total volume 50 mL, 0.315 g of DNS (3,5-Dinitrosalicylic acid, catalog number D0550, Sigma) and 1.048 g of sodium hydroxide were added in a 100 mL conical flask containing 20 mL of distilled water and, then, a homogeneous mixture was created with the help of magnetic steering. Distilled water was added until the volume reached 25 ml. In another 100 mL conical flask, containing 20 mL of distilled water, 9.1 g sodium-potassium tartrate, 0.25 g phenol (2.6 ml if it is liquid) and 0.25 g sodium-metabisulfite were added and a homogeneous mixture was created with the help of magnetic steering. Distilled water was added until the volume reached 25 ml. The reagent solution (from first and second flask) was gently mix into a separate 100 mL glass bottle, wrapped with aluminum foil and stored at 4 °C. Prior to the extraction of digestive glands, if heat-shock was needed, *Parhyale* adults were positioned individually in Petri dishes with pre-heated artificial sea water at 37° C for 6 h and then recovery at room temperature for 12 h was occurred. For the sample preparation, it was necessary to dissect Parhyale adult digestive system, using tongs and a stereoscope set up, in 60 mm tissue culture Petri dishes coated with SYLGARD and filled with filtered artificial sea water and then digestive glands were transferred in 0.1 M sodium acetate buffer pH 4.8 containing 1x protease inhibitors (catalog number P8465, Sigma). For each set of adult digestive glands (corresponding to 1 *Parhyale* adult) 50ul of 0.1 M sodium acetate buffer pH 4.8 were used. Digestive proteins and fluids were extracted by grinding with pestle and then centrifuge sample at 4°C at 1200 g for 12 min. Supernatant was transferred in a clean microcentrifuge tube. For the preparation of the reaction mixture for 5 measurements, it was needed to add 1 mL of Avicel suspension in 0.1 M sodium acetate buffer (pH 4.8) and 0.25 mL of digestive gland lysate samples in 0.1 M sodium acetate buffer pH 4.8. 250ul of the mixture were used for each timepoint measurement. This mixture contained digestive fluids and proteins that corresponded to 1 *Parhyale* adult. The reaction mixture was incubated at 30 °C and 1000 rpm, for 24 h and 250ul were extracted for each timepoint measurement. First, 250ul were extracted and measured for t=0 (basal level) and then for 1h, 3h, 6h, 24h. The reaction was quenched by placing the tubes in ice-cold water for 10 min. The mixture was centrifuged at 13,000 g for 3 min and the supernatant was collected. The supernatant was analyzed by DNS method for reducing sugars. The procedure for DNS method was the following. 250ul of sample supernatant were placed in a thin glass test tube and 750ul of DNS solution were added and mixed well by gently vortex for 3–5s. The tube was placed in boiling water bath for 5 min. The tube was removed from hot water bath and cooled down in ice-cold water (or ice) for 10 min. 2.5 mL of distilled water were added into the tube and mixed properly. The absorbance was measured at 540 nm in spectrophotometer with three replicates for each sample. The absorbance values of t=0 measurement (blank) was subtracted from each other timepoint measurement and the enzyme activity was analyzed.

B. p-Nitrophenyl β-D-cellobioside (pNP) assay - Detection of release of pNP ions

For the substrate preparation, it was needed to add 9.2mg of 4-Nitrophenyl β-D-cellobioside (pNG, catalog number N5759, Sigma) in 1ml RNase/DNase free H2O to create 20mM stock solution. The tube was intensively vortexed for 3-5 min in order to make a homogeneous mixture. This mix could be store at 4 ℃ for a few days. For the preparation of 100ml 0.1 M citrate buffer pH 4.5 (reaction buffer), 1.351 g sodium citrate and 1.038 g citric acid were added in 80ml distilled water. pH was adjusted with 1M HCl until 4.5. Distilled water was added until the volume reached 100 ml and it was stored at RT. For the preparation of 100ml 2.5m NaC03 (termination buffer), it was needed to add 26.49g in 80ml distilled water. Distilled water was added until the volume reached 100 ml and store at RT. For the preparation of the reaction mixture for 4 measurements, 20ul of 20mM stock solution 4-Nitrophenyl β-D-cellobioside -pNG- (final concentration 0.5 mM), 80ul of 5M NaCl (final concentration 500 mM) and appropriate volume of protein sample (crude extract or pure enzyme) were added and volume was adjusted until 800ul with 0.1 M citrate buffer pH 4.5. 200ul of the mixture were used for each timepoint measurement. The reaction mixture was incubated at 30 ℃, for 20 h and 200ul were extracted for each timepoint measurement. First, 200ul were extracted and measured for t=0 (basal level) and then for 1h, 4h, 20h. The reaction was quenched by adding 50ul of 2.5m sodium carbonate -NaC03- (final concentration 500 mM) in each 200ul of each sample in a 96-well plate. Detection of release of pNP ions by measuring the absorbance at 405 nm in a plate reader was conducted. The absorbance values of t=0 measurement (blank) was subtracted from each other timepoint measurement and the enzyme activity was analyzed.

Genotyping of CRISPR NHEJ knock-in transgenic animals

Genomic DNA was isolated by cutting limbs and antennas from both sides of transgenic *Parhyale* adults in order to utilize it in PCR tests with specific primers for the desired region. Cut limbs and antennas were placed in 50 ul of squishing buffer, containing 10mM Tris-HCl pH 8, 1mM EDTA, 25mM NaCl and 200 ng/ul Proteinase K, and smashed using pipette tip. This mix was incubated at 37 ℃ for 20-30 min and, then, Proteinase K was heat inactivated at 85℃ for 1-2 min. This preparation could be stored at 4 ℃ for several days and 1-2 ul was adequate as a template for a PCR reaction.

The primers used for genotyping binds particularly on either side of the CRISPR/Cas9 cleavage site. The sgRNA targets at the coding sequence of GH7a, 180 bp downstream ATG. KI_LJ_Gen forward primer (5’ TGGATTCCAACTGGCGCTGG 3’) binds to 35 bp upstream sgRNA, inside the coding sequence of GH7, and KI_LJ_Gen reverse primer (5’ CCCTTGCTCACCATCTCGAG 3’) binds to 1755 bp downstream sgRNA, inside the coding sequence of fluorescent protein tdTom, in order to achieve specific binding to the CRISPR NHEJ knock-in cassette and no binding to endogenous regions. PCR using these primers was conducted with the aid of Phusion High-Fidelity DNA Polymerase (catalog number M0530, NEB) for 35 cycles at annealing temperature 67 ℃ and extension time 1 min and 10 sec. The 1810bp PCR product was gel extracted and ligated to pCRBlunt II-TOPO vector, using Zero Blunt TOPO PCR Cloning Kit (catalog number 450245, Thermo Fisher Scientific). Specific plasmid clones were sent for sequencing with T7 forward and M13 reverse primers.

Other primer sets used in genotyping were 4 different forward primers specific to 5' UTRs of 4 different GH7 genes (a, b, c, and d), around 150-200 bp upstream ATG of each GH7 gene, keeping the same reverse primer KI_LJ_Gen that binds inside the coding sequence of tdTom (specific binding to the CRISPR NHEJ knock-in cassette). PCR using these primers was conducted with the aid of Phusion High-Fidelity DNA Polymerase for 35 cycles at annealing temperature 67 ℃ and extension time 1 min and 10 sec. The 2000bp PCR products was gel extracted and ligated to pCRBlunt II-TOPO vector, using Zero Blunt TOPO PCR Cloning Kit. Specific plasmid clones were sent for sequencing with T7 forward and M13 reverse primers.

Results

A Gain-of-function genetic approach for investigation of glycosyl hydrolase PhGH7

In order to study the role of GH7, we opted to use a gain of function strategy. The Minos transposable element was used with a heat-inducible promoter in *Parhyale* to conditionally overexpress tagged PhGH7 gene in all cells of transgenic animals. PhGH7 gene was tagged at the C-terminus with two tagging cassettes, a removable long tag and a non-removable short tag. Long tag consists of the cleavage peptide for the sequence specific protease TEV (Tobacco Etch Virus endopeptidase), the mCherry red fluorescent protein, followed by a double StrepTagII and the FLAG epitopes for affinity purification (Fig. 10A). On the other hand, short tag includes only the double StrepTagII and the FLAG epitopes (Fig. 10B). The long tagging cassette allowed us to fluorescently label and visualize the GH7 protein *in vivo*, as well to overexpress, purify and test its enzymatic activity under native conditions.

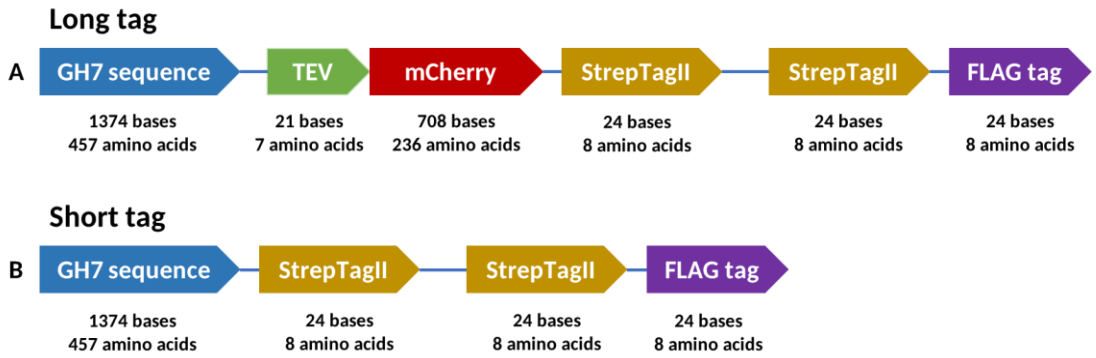

Figure 10. PhGH7 tagging cassettes. *Schematic representation of the heat-inducible (A) removable long-tagged and (B) non-removable short-tagged versions of PhGH7.*

To achieve this purpose, the initial step was the generation of two plasmid constructs, pMi-PhHS-GH7long-OpsinEGFP and pMi-PhHS-GH7short-OpsinEGFP, which contain Minos-transposase inverted repeats for random integration in the genome, a heat-inducible cis-regulatory sequence for conditional overexpression only after heat-shock at 37˚C, the coding sequence of the *GH7a* gene (which is one of 9 highly similar *Parhyale GH7* genes), the long or short tag, respectively, and transgenesis marker PhOpsin regulatory element fused with EGFP. Subsequently, microinjections of 1-cell *Parhyale* embryos were performed for both constructs.

A total of 1327 and 1656 *Parhyale* embryos were injected with pMi-PhHS-GH7long-OpsinEGFP and pMi-PhHS-GH7short-OpsinEGFP constructs, respectively (Table 2). With a 14.5% survival ratio after microinjections of long tagged GH7 and a 19.2% survival ratio after microinjections of short tagged GH7, we could recover 193 and 319 G0 individuals, respectively. Out of these G0 individuals, 29 for GH7long and 28 for GH7short were expressing the transformation marker in the eye, unilaterally or bilaterally. Subsequently, these animals were reared to adulthood. Only 6 G0 animals injected with long tagged GH7 and 13 G0 animals with short tagged GH7 reached adulthood. They were crossed with wild-type *Parhyale* individuals, in order to screen their G1 progeny for transgenic animals. A large screening of over 510 offsprings of GH7long G0s and 1360 offsprings of GH7short G0s was needed to identify transgenic G1s, which is a significantly higher number compared to standard transgenesis experiments in *Parhyale*. Finally, I recovered three distinct viable transgenic lines: one

transgenic line containing the removable C-terminal long tag and two transgenic lines with the non-removable C-terminal short tag.

A. Long tagged GH7		
Number of injected fertilized eggs	1327	
Screened late stage G0 embryos	193	
Number of Opsin-EGFP negative G0s	164	
Number of Opsin-EGFP positive G0s	29	
Number of Opsin-EGFP positive G0 crossed with WT	6	
Total number of G1s screened	519	
Number of G0s producing transgenic G1s	G0#2 ♀	20 wildtype G1s/ 90 transgenic G1s

B. Short tagged GH7		
Number of injected fertilized eggs	1656	
Screened late stage G0 embryos	319	
Number of Opsin-EGFP negative G0s	291	
Number of Opsin-EGFP positive G0s	28	
Number of Opsin-EGFP positive G0 crossed with WT	13	
Total number of G1s screened	1361	
Number of G0s producing transgenic G1s	G0#2 ♂	224 wildtype G1s/ 5 transgenic G1s
	G0#7 ♂	217 wildtype G1s/ 9 transgenic G1s

Table 2. Overview of transgenesis experiments with the long-tagged (A) and short-tagged (B) GH7 constructs.

It is important to note that in all the ensuing experiments described below, the transgenic *Parhyale* line encoding the long-tagged GH7 was used, due to limited number of transgenic short-tagged GH7 animals available. Transgenic G1 embryos with long tagged GH7 could be distinguished from other wild-type ones, from the 10th day onwards, owing to the EGFP fluorescence in their eyes and mCherry fluorescence across their whole body after 1 h of heat shock at 37˚C (Fig. 11).

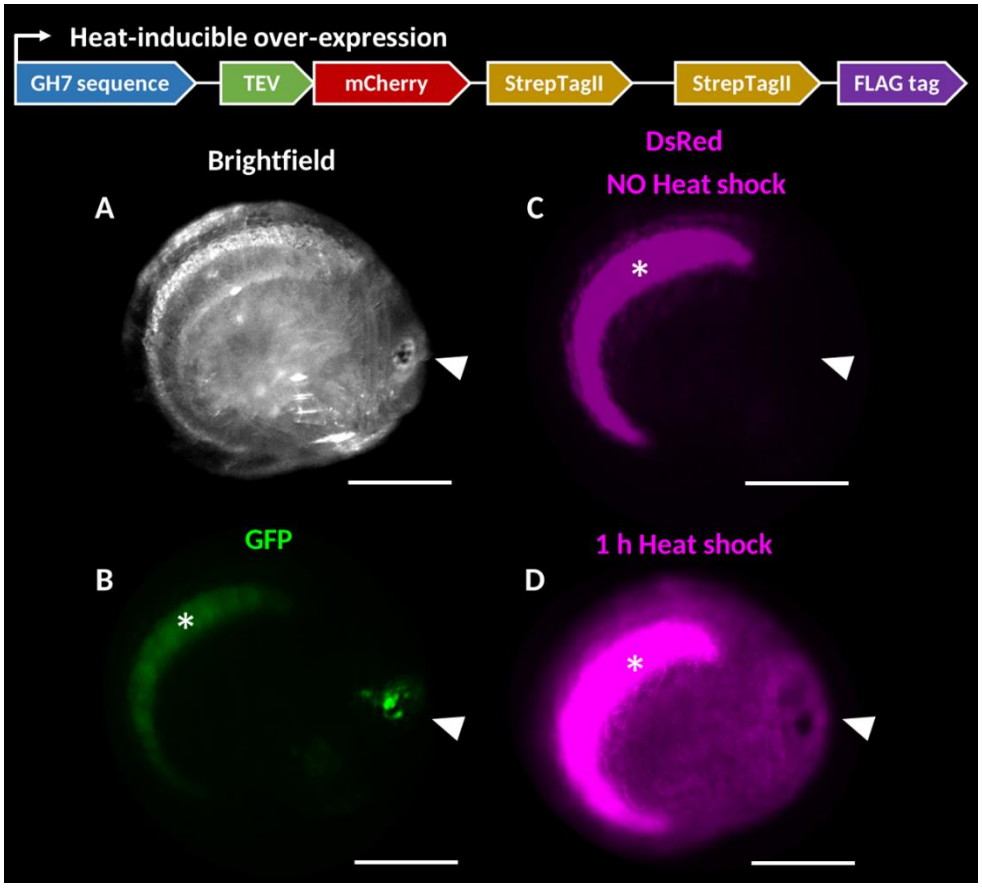

Figure 11. Imaging of transgenic Parhyale embryos encoding the long-tagged GH7 by fluorescence stereomicroscopy. *(A) Brightfield image depicting a 10th-day old Parhyale embryo. (B) Fluorescent image in the GFP channel (green) showing expression of the transgenesis marker in the compound eyes (arrowheads) and autofluorescence of the yolk in the midgut (asterisks). (C) Fluorescent image in the DsRed channel (magenta) before heat-shock exhibiting only the yolk autofluorescence and (D) after 1 h heat-shock exhibiting expression of the mCherry-tagged GH7 throughout the embryo, too. All panels show lateral views with anterior to the right. Scale bars 200um.*

Determination of heat-shock experiment parameters

To exploit the heat-inducible system in *Parhyale* for conditional overexpression of tagged PhGH7 gene, it was crucial to determine the exact duration of heat-shock and subsequent recovery time of the animals. These parameters were previously optimized in embryos (REF Pavlopoulos et al., PNAS, 2009), but in these investigations I used adult transgenic *Parhyale* to isolate larger quantities of expressed long-tagged GH7. I used a transgenic line containing the same heat-inducible cis-regulatory element driving expression of a histone fused red fluorescent protein (H2B-RFP) to determine the best heat-shock conditions. H2B-RFP *Parhyale* adult individuals were transferred to pre-warmed filtered artificial sea water (FASW) and left in the 37 ℃ incubator for different durations of heat-shock 3, 6, 9 and 12 hours, after which they had 3-hour recovery time at room temperature. H2B-RFP *Parhyale* adults without any heat-shock treatment were used as negative control. After recovery, the digestive glands of both treated and untreated adults were dissected, and midgut extracts were analyzed H2B-RFP protein levels (molecular weight of H2B-RFP fusion is 41 kDa and for plain RFP is 27 kDa) by Western blotting using a polyclonal Rabbit anti-mCherry antibody (Rockland, catalog no. 600-401-P16). According to the results depicted in Figure 12A, the 6 and 9 hours of heat-shock demonstrated the higher levels of overexpressed protein. Surprisingly, the molecular weight

of about 27kDa of the overexpressed protein in midgut extracts corresponds to that of plain RFP, while the expected molecular weight of 41kDa of the overexpressed H2B-RFP was detected in the rest non-midgut tissues. This finding was not essential per se for these pilot experiments, but raised concerns about the cleavage of fusion proteins in midgut cells.

Next, in order to define the appropriate recovery time after heat shock at 37 ℃, we subjected H2B-RFP *Parhyale* adults to a 9-hour heat-shock and recovery at room temperature for 6, 9 and 12 hours. Again, H2B-RFP Parhyale adults without any treatment were used as a negative control. Following the same experimental procedure mentioned above, Western blotting showed that H2B-RFP levels were quite similar between all tested recovery times (Fig. 12B).

Based on these results, a Western blotting analysis was applied to GH7long transgenic animals. Hence, transgenic *Parhyale* adults were subjected to a 6-hour heat-shock and a 12-hour recovery time. The protein produced from the GH7long cassette has an expected molecular weight of 82.155 kDa. Unlike H2B-RFP, a band of expected size was observed with the same anti-mCherry antibody detecting the mCherry fluorescent protein in GH7long (Fig. 12C). This time, protein extracts derived from both midgut and the remaining body tissues were analyzed separately by Western blotting. Interestingly, the levels of GH7long in hepatopancreas were remarkably higher than in the remaining tissues (see Discussion).

It is important to mention that all protein extracts used for Western blotting were quantified with the Bradford assay to ensure that all lanes are loaded with equal amounts of total protein. Notably, anti-actin, anti-tubulin and anti-GAPDH antibodies were also used as loading controls on the same membranes. Although proteins of expected size were detectable in non-midgut tissues, they were never detected in the midgut extracts.

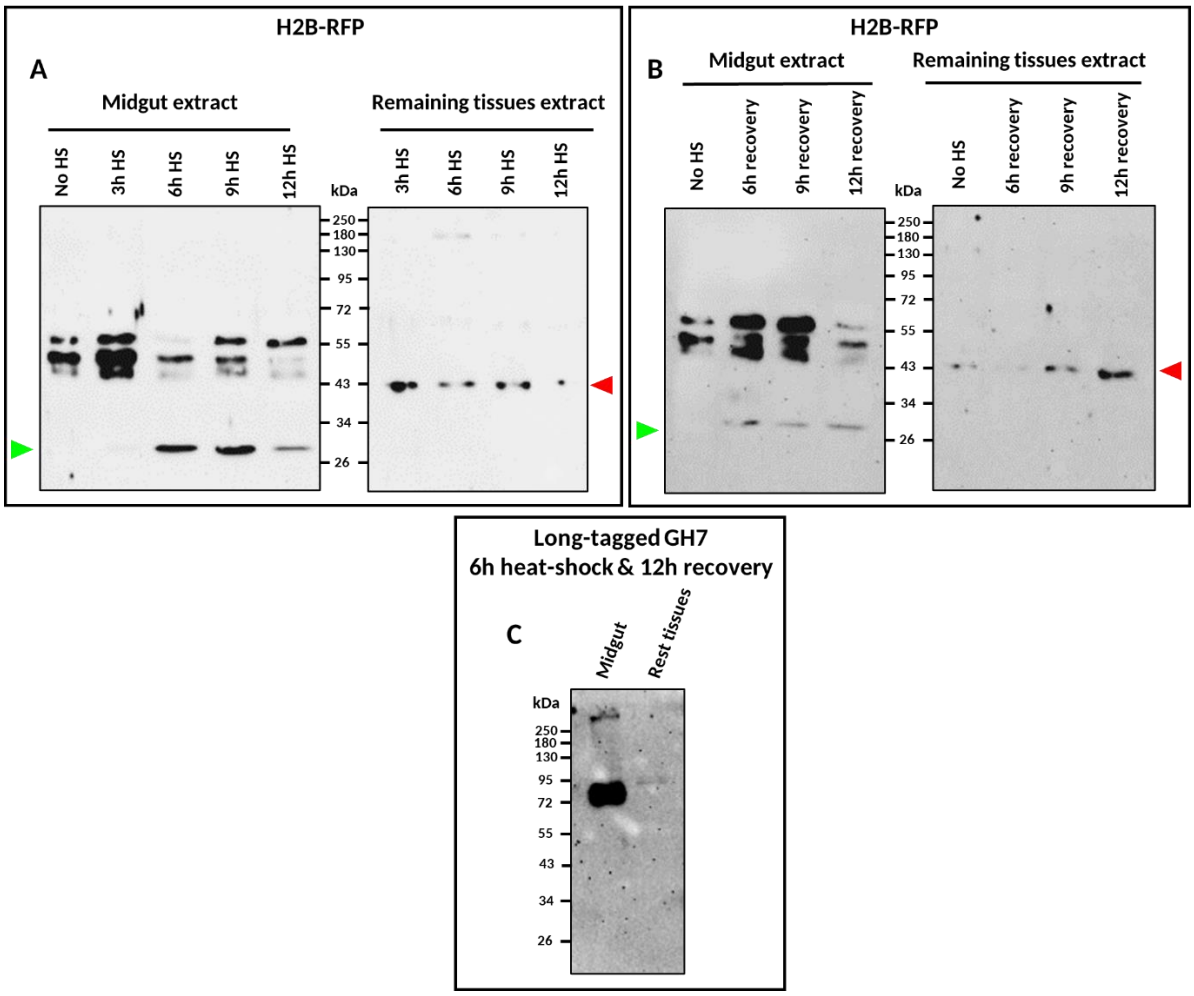

Figure 12. Western blot analyses for determination of heat-shock experimental parameters. *Midgut extracts and remaining tissues extracts from transgenic adult Parhyale encoding HS-H2B-RFP (A) subjected to 3-, 6-, 9- or 12-hour heat-shock at 37˚C followed by a 3-hour recovery at room temperature, and (B) subjected to a 9-hour heat-shock followed by a 6-,9- or 12-hour recovery. No HS corresponds to samples collected from transgenic adults kept at room temperature. (C) Midgut and rest tissue extracts from transgenic adult Parhyale overexpressing GH7long after a 6-hour heat-shock and a 12-hour recovery. All experiments were carried out with an anti-mCherry polyclonal antibody (Rockland, catalog no. 600-401-P16). Red arrowhead depicts the expected molecular weight of H2B-RFP fusion (41 kDa) and green arrowhead displays the molecular weight of plain RFP (27 kDa).*

In vitro analyses of glycosyl hydrolase activity in wildtype and mutant *Parhyale*

In order to investigate the glycosyl hydrolase activity of *Parhyale*'s digestive glands extracts, I employed a hydrolysis assay of the microcrystalline cellulosic substrate Avicel. The activity of cellulolytic enzymes is estimated by measuring the concentration of released sugars following hydrolysis as an increase in optical density at 540nm after treating the samples with dinitrosalicylic acid (DNS). Initially, the DNS assay was applied to samples obtained from wild-type adult *Parhyale* to explore the conditions of endogenous enzymes. This analysis revealed that extracts from dissected wild-type digestive glands displayed a temperature-dependent cellulolytic activity, which was significantly reduced in extracts heat-inactivated or pre-treated with Proteinase K (Fig. 13A). These results suggest that the *Parhyale* cellulolytic capacity depends on enzymes derived from the midgut digestive glands.

In a next experiment, extracts from hepatopancreatic glands were obtained from transgenic adult *Parhyale* encoding the long-tagged GH7, with or without heat-shock treatment, and were measured with the DNS assay. Transgenic Parhyale adults were heat-shocked for 6 h at 37 ˚C and allowed to recover at room temperature for 12 h. As depicted in figure 13B, equal amounts of extracted proteins from heat-shocked and non-heat-shocked adults displayed a higher activity after GH7long overexpression. This increase suggested that the overexpressed long-tagged GH7 enhances the cellulolytic capacity of the endogenous enzymatic cocktail in the *Parhyale* digestive glands.

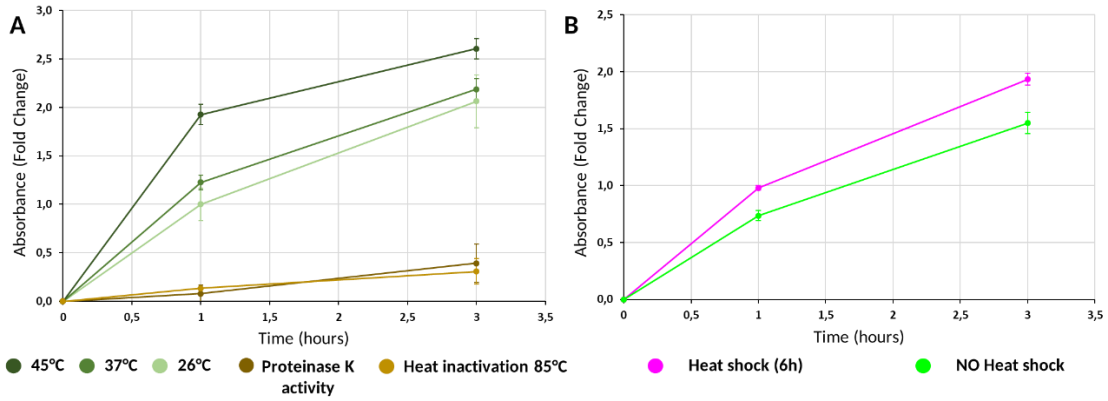

Figure 13. Avicel-DNS hydrolysis assays with midgut extracts. *(A) Line graphs showing the normalized fold change in DNS absorbance over time at 26°C, 37°C, and 45°C (shades of green) using midgut extracts from wild-type adult Parhyale. Hydrolysis reactions with midgut extracts pretreated with Proteinase K (PK) or heat-inactivated at 85°C are indicated with dark and light brown lines, respectively. (B) Line graphs showing the normalized fold change in DNS absorbance over time at 30°C using midgut extracts from transgenic Parhyale encoding the heat-inducible long-tagged GH7, with (magenta) or without (green) heat-shock. Points show mean values from 3 technical replicates and error bars show standard deviations.*

Purification of overexpressed *Parhyale* GH7

With the ultimate goal of purification and biochemical characterization of the overexpressed *Parhyale GH7*, based on the results from the previous experiment, I isolated a large amount of digestive gland extract as follows. 23 transgenic GH7long adults and 23 transgenic H2B-RFP adults were subjected to a 6-hour heat-shock treatment at 37°C. After a 12-hour recovery, proteins were extracted, separately, from the dissected midguts and the remaining tissues. The soluble protein lysates were, then, exposed to streptavidin beads (MagStrep “type3” XT beads) for the purpose of isolating only the overexpressed GH7long protein via the double StrepTagII. Aliquots from the input samples and all the steps of the purification process were analyzed with Coomassie-stained SDS-PAGE and Western blotting with the mCherry antibody (Fig. 14). A band corresponding to the size of the overexpressed GH7long fusion protein was present in the insoluble fractions of both the midgut and rest body tissues (Fig. 14 A’,B’ lanes 1). However, analysis of the soluble fractions showed that it was detected exclusively in the midgut extract (Fig. 14 A’,B’ lanes 2) suggesting that only midgut cells have the appropriate machinery to process the soluble, functional GH7 protein. Consequently, no GH7 protein was detectable in the soluble lysate and purification fractions from rest tissue extracts (Fig. 14A’ lanes 2-5). Analysis of the purification fractions of the midgut soluble extract showed that the majority of the long-tagged GH7 protein was not bound by the streptavidin beads and only a

small fraction was affinity-purified and eluted from the streptavidin beads (Fig. 14B' lanes 2-5).

Analysis of the control H2B-RFP samples revealed the presence of a shorter than expected band both in the soluble and insoluble lysates, in the unbound and just detectable in the wash fractions of the streptavidin beads, and not detected in the eluate from the streptavidin beads (Fig. 14C' lanes 1-5). This is expected, considering the absence of a Strep-tag in H2B-RFP, but was used as a control in the follow-up biochemical assays to test for the presence of any untagged endogenous glycosyl hydrolases in the affinity-purified eluted fraction.

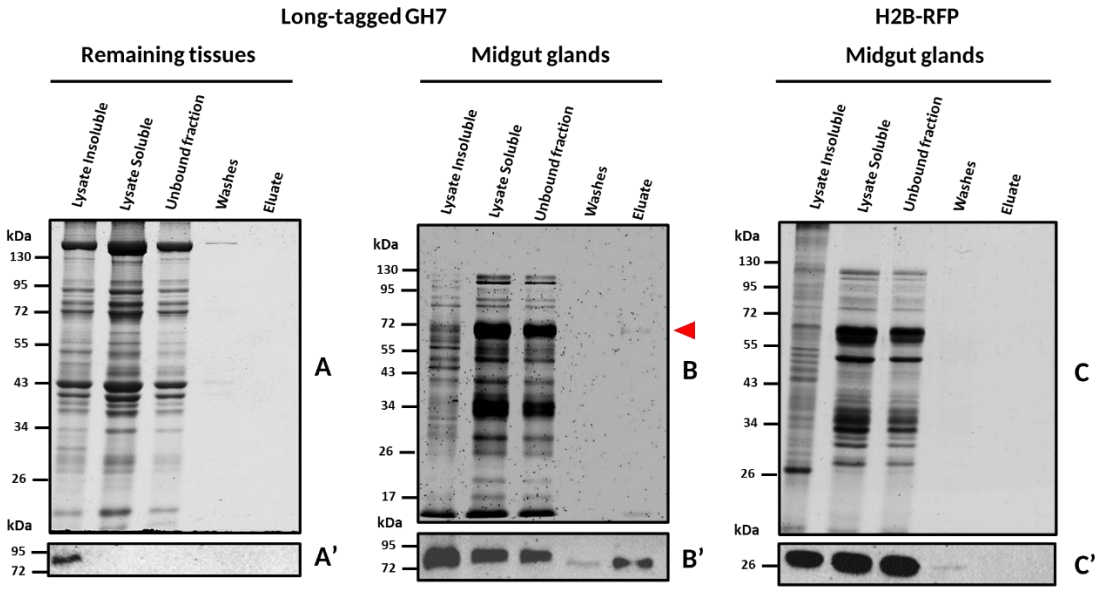

Figure 14. Protein purification analysis. *(A-C) Coomassie-stained SDS-PAGE gels and (A'-C') Western blot analyses of the same samples using an anti-mCherry polyclonal antibody (Rockland, catalog no. 600-401-P16). In all panels, lanes indicate from left to right: i) Insoluble fraction and ii) Soluble fractions of tissue lysate, iii) Unbound, iv) Wash and v) Eluted affinity-purified fractions from streptavidin beads. (A-A') Rest tissue samples and (B-B') midgut samples from transgenic adult Parhyale overexpressing the heat-inducible long-tagged GH7. (C-C') Midgut samples from transgenic adult Parhyale overexpressing the heat-inducible H2B-RFP.*

Biochemical characterization of *Parhyale* glycosyl hydrolases

To investigate the enzymatic activity of the soluble lysates and purification fractions in vitro, we used the chromogenic substrate p-nitrophenyl-β-D-cellobioside (pNP-G2) which is specific to detect cellobiohydrolase activity by measuring absorbance at 405 nm. Endoglucanases / cellobiohydrolases hydrolyze the pNP-G2 cellobiose analog releasing the chromogenic p-nitrophenol (PNP). Absorbance was first measured on soluble lysates and the unbound supernatants following streptavidin bead incubation from the heat-shocked GH7long midgut samples (Fig. 15A) and the heat-shocked H2B-RFP midgut samples (Fig. 15B) that exhibited comparable cellulolytic activity, presumably from the endogenous *Parhyale* cellobiohydrolases. No cellobiohydrolase activity was detected in the GH7long rest tissue samples (Fig. 15C) demonstrating that these enzymes are only functional in the midgut. More importantly, after the streptavidin affinity purification, only the eluate from the GH7long midgut exhibited a cellobiohydrolase activity, in contrast to eluates collected from either the GH7long rest body tissues or the H2B-RFP midgut (Fig. 15D).

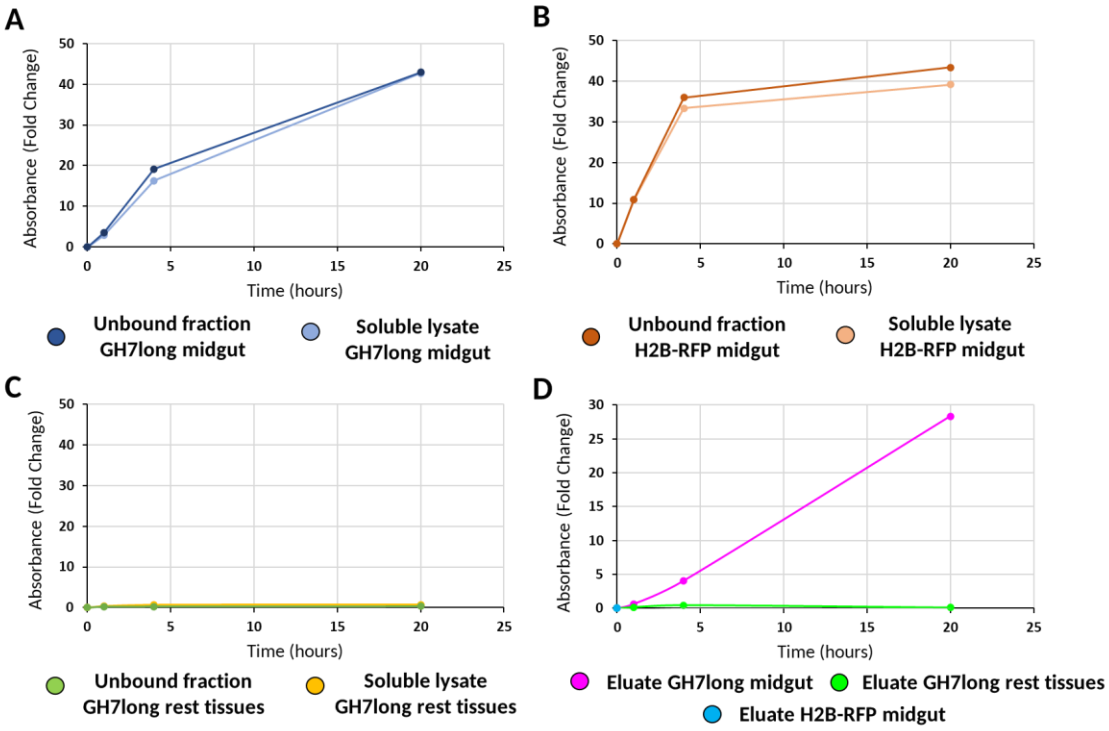

Figure 15. pNP-G2 cellobiohydrolase assays. *Line graphs showing the normalized fold change in pNP-G2 absorbance over time at 30˚C using the soluble fraction of tissue lysate and unbound fraction from streptavidin beads of (A) midgut extracts from transgenic Parhyale overexpressing the long-tagged GH7, (B) midgut extracts from transgenic Parhyale overexpressing H2B-RFP, and (C) rest-tissue extracts from transgenic Parhyale overexpressing the long-tagged GH7. (D) Similar line graphs showing the normalized fold change in pNP-G2 absorbance over time of the eluted affinity-purified fractions obtained from midgut extracts (magenta) and rest-tissue extracts (green) from transgenic Parhyale overexpressing the long-tagged GH7, as well as midgut extracts from transgenic Parhyale overexpressing H2B-RFP (blue). Points show values from single measurements.*

In a last experiment, the purified *Parhyale* GH7long protein was compared to the commercially available fungal cellobiohydrolase HjCel7A (Cellobiohydrolase I from *Hypocrea jecorina*, catalog number E6412, Sigma, 0.13U/mg stock concentration), using the same cellobiohydrolase-specific (pNP-G2) assay. In detail, two different concentrations of HjCel7A enzyme, 769µg (0.1U) and 192µg (0.025U) were compared to 0.5µg of purified PhGH7 at two different temperatures, 45°C, which is the optimal temperature for HjCel7A, and 30°C. Absorbance was measured at 405 nm after 4h incubation of pNp-G2 substrate with each enzyme. This comparison revealed that 0.5µg of *Parhyale* GH7long exhibited a similar activity to 769µg (0.1U) of HjCel7A at 30°C (Fig. 16A) and to 192µg (0.025U) of Cel7A at 45°C (Fig. 16B).

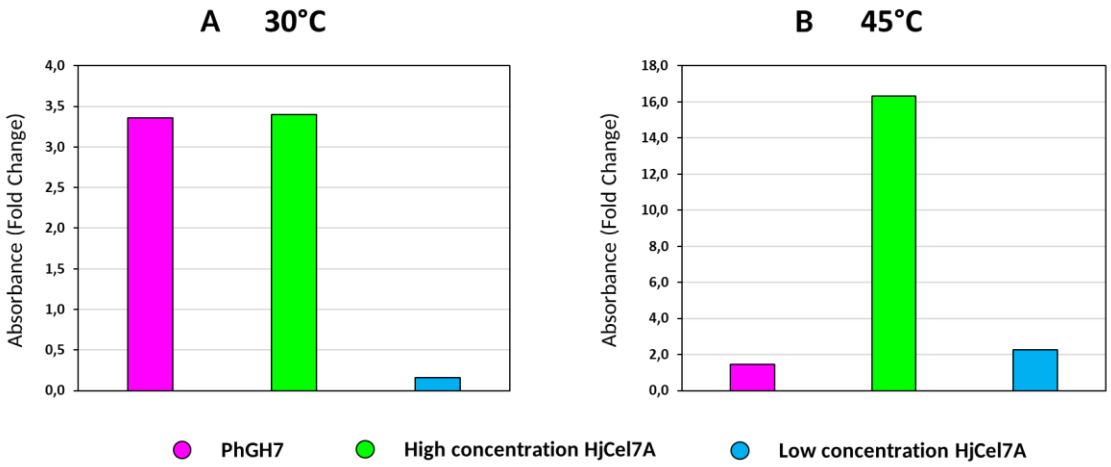

Figure 16. Comparative analysis of HjCel7A and PhGH7. *Bar graphs displaying the normalized fold change in pNP-G2 absorbance at (A) 30°C and (B) 45°C after a 4-hour incubation with 0.5µg of the affinity-purified long-tagged Parhyale GH7 (PhGH7 in magenta) or the recombinant fungal cellobiohydrolase from Hypocrea jecorina HjCel7A at a higher concentration (0.1U/769µg in green) and a lower concentration (0.025U/192µg in blue).*

Expression and monitoring of *Parhyale* GH7 in native conditions

In parallel to creating heat-shock overexpression lines, which may have been toxic, I took a second approach and generated a tagged version of GH7 in the native locus using the CRISPR/Cas9 system. The main purpose of this strategy was the introduction of affinity and fluorescent tags into the endogenous GH7 genes of *Parhyale*, thus enabling protein visualization and purification. In contrast to the over expression transgenic lines, the generated transgenic animals would produce the tagged GH7 protein at physiological relevant amounts and only in midgut cells where it is normally expressed. To achieve this, the plasmid pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP was injected in 1-cell *Parhyale* embryos with Cas 9 protein and the appropriate sgRNA. This plasmid consisted of the GH7 coding sequence, the double StrepTagII and FLAG epitopes (referred to as short tag), the T2A self-cleaving peptide, the *Parhyale* histone H2B fused with the bright red fluorescent protein tandem Tomato (tdTom) and the transgenesis marker PhOpsin regulatory element fused with EGFP (Fig. 17A). In this homology-independent process, double-strand breaks were induced using a specific sgRNA, which is targeting, simultaneously, both the endogenous GH7 genes and the co-injected plasmid carrying the GH7 coding sequence. We expected that all knock-in events of the linearized plasmid in endogenous GH7 genes would result in PhOpsin-EGFP eye fluorescence, with those having the appropriate orientation and open reading frame resulting also in nuclear tdTom fluorescence in midgut cells.

As shown in Table 3, 3311 *Parhyale* fertilized eggs were injected using pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP construct. With a 38.5% surviving percentage after microinjections, I discovered 63 G0 late embryos with Opsin-EGFP expression in the eye and 1 with midgut fluorescence, but without Opsin-EGFP expression in the eye. Among these 64 G0 animals, 60 possessed only eye fluorescence, 3 exhibited both eye and midgut fluorescence and 1 only midgut fluorescence. After reaching sexual maturity, 45 G0 animals, 42 with only eye fluorescence, 2 with eye and midgut fluorescence and 1 with only midgut fluorescence were crossed with wild-type *Parhyale* individuals, in order to screen their G1 progeny. After screening over 2300 offspring, an integration event of the linearized plasmid in the correct

orientation and open reading frame within a GH7 locus was successfully identified. This outcome led to nuclear tdTom fluorescence, presumably in the GH7-expressing midgut cells, that could be easily detected in both knocked-in embryos and hatchlings (Fig. 17 C). Since tdTom is downstream of the GH7 open reading frame, nuclear fluorescence is indicative of short-tagged GH7 protein expression in the same cells.

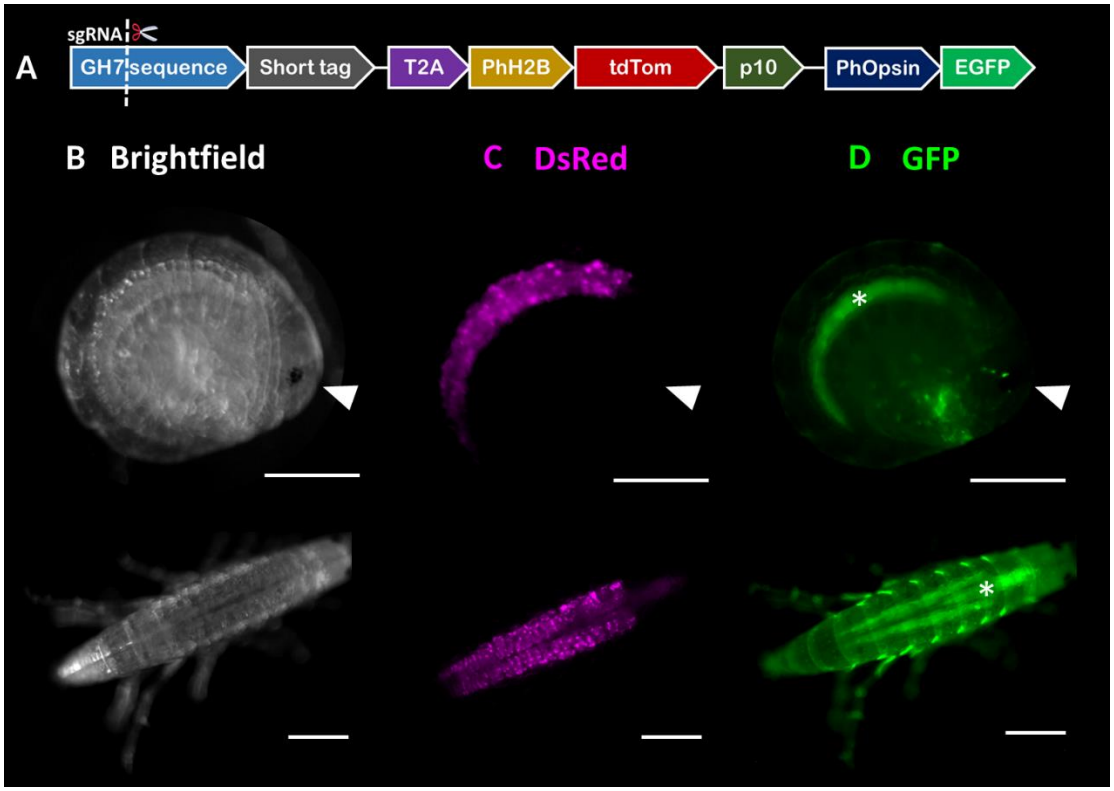

Figure 17. Parhyale GH7 genome editing by CRISPR knock-in. *(A) Schematic illustration of the plasmid construct used in the homology-independent CRISPR/Cas9-based knock-in experiments. Late-stage embryo and hatchling from a CRISPR/Cas9 Parhyale knock-in line with the modified GH7b locus: (B) brightfield image, (C) fluorescent image in the DsRed channel (magenta) exhibiting a nuclear tandem Tomato (tdTom) signal in midgut cells, and (D) fluorescent image in the EGFP channel (green) displaying PhOpsin driven expression in the compound eyes (arrowheads), as well as autofluorescence signal from the midgut yolk (asterisks), the gnathal appendages in the late embryo and the exoskeleton in the hatchling. The late embryo is shown in lateral view and the hatchling in dorsal view with anterior to the right. Scale bars 200µm.*

A complementary approach involved the co-injection of a plasmid construct, pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed, and complexes of Cas9 protein / sgRNA. This plasmid contained the GH7 coding sequence, the double StrepTagII and FLAG epitopes, the T2A self-cleaving peptide, the *Parhyale* histone H2B, the bright green fluorescent protein mEmerald (instead of tdTom), and the transgenesis marker PhOpsin regulatory element fused with DsRed (instead of EGFP). The rationale for this strategy was the comparison of the nuclear fluorescence signals produced by tdTom and mEmerald in midgut cells against the red and green yolk autofluorescence, respectively.

As summarized in Table 3, 620 *Parhyale* fertilized eggs were injected using pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed construct. With a 44% surviving percentage after microinjections, I identified 30 G0 individuals with expression of DsRed in the eye. These

G0 animals possessed only eye fluorescence. After reaching sexual maturity, 17 G0s with only eye fluorescence were crossed with wild-type *Parhyale* individuals, in order to investigate their G1 progeny. A screening of over 630 offsprings did not lead to the discovery of nuclear mEmerald fluorescence in the GH7-expressing midgut cells. It should be noted that this screening was more difficult due to the high autofluorescence of the *Parhyale* exoskeleton and, especially, the yolk in the digestive glands using the GFP filter set (compared to the DsRed filter set) under the epifluorescent stereomicroscope (data not shown).

A. CRISPR KI GH7 with tdTom		
Number of injected fertilized eggs	3311	
Screened late stage G0 embryos	1278	
Number of Opsin-EGFP negative G0s w/o midgut tdTom fluorescence	1214	
Number of Opsin-EGFP positive G0s w/o midgut tdTom fluorescence	60	
Number of Opsin-EGFP negative G0s with midgut tdTom fluorescence	1	
Number of Opsin-EGFP positive G0s with midgut tdTom fluorescence	3	
Number of Opsin-EGFP positive and/or tdTom positive G0 crossed with WT	45	
Total number of G1s screened	2301	
Number of G0s producing Opsin-EGFP & tdTom positive G1s	G0#3 ♀	78 wildtype G1s/ 3 transgenic G1s
Number of G0s producing Opsin-EGFP positive G1s	G0#2 ♂	81 wildtype G1s/ 71 transgenic G1s
	G0#13 ♂	36 wildtype G1s/ 40 transgenic G1s
	G0#16 ♂	30 wildtype G1s/ 44 transgenic G1s
	G0#3 ♂	102 wildtype G1s/ 5 transgenic G1s

B. CRISPR KI GH7 with mEmerald		
Number of injected fertilized eggs	620	
Screened late stage G0 embryos	273	
Number of Opsin-EGFP negative G0s w/o midgut mEmerald fluorescence	243	
Number of Opsin-EGFP positive G0s w/o midgut mEmerald fluorescence	30	
Number of Opsin-EGFP negative G0s with midgut mEmerald fluorescence	0	
Number of Opsin-EGFP positive G0s with midgut mEmerald fluorescence	0	
Number of Opsin-EGFP positive G0 crossed with WT	17	
Total number of G1s screened	631	
Number of G0s producing Opsin-EGFP positive G1s	G0#2 ♀	21 wildtype G1s/ 7 transgenic G1s
	G0#13 ♀	11 wildtype G1s/ 11 transgenic G1s

Table 3. Overview of* CRISPR knock-in *experiments with pCRISPR_NHEJ_KI-GH7short-T2A-H2B-tdTom-p10-OpsinEGFP (left) and pCRISPR_NHEJ_KI-GH7short-T2A-H2B-mEmerald-p10-OpsinDsRed (right)

It is noteworthy that the recovery of knock-in events of the desired construct inside GH7 locus, after the induced double-strand breaks, was a challenging process. Non-homologous end joining (NHEJ) is an error-prone method, often resulting in the accumulation of small insertions or deletions at the cleavage site. Therefore, it was crucial to thoroughly investigate and confirm the correct insertion of the linearized plasmid in the desired orientation and open reading frame, as well as the targeted GH7 gene before proceeding with additional experimental procedures. I conducted genotyping with the aid of specialized primer sets and sequencing of the PCR products amplified from genomic DNA of selected G0s and G2s. The presence of EGFP fluorescence in the eyes served as an indicator of all knock-in events. It was also expected that only integrations in the correct orientation and without disrupting the GH7 open reading frame would also result in tdTom fluorescence in midgut cells. Therefore, genotyping was performed on both categories of knocked-in *Parhyale*, those with only eye fluorescence and

those with both eye and digestive gland fluorescence. Genomic DNA was extracted from limbs dissected from G0 *Parhyale* individuals, without killing the animals. This was used as template to specifically amplify the targeted regions using a forward primer hybridizing upstream of the CRISPR/Cas9 cleavage site in a conserved stretch of nucleotides in the GH7 coding sequences and a reverse primer positioned within the coding sequence of tdTom (Fig. 18A).

Sequencing two individuals that exhibited only eye fluorescence revealed a 25bp and an 8bp deletion at the sgRNA targeted site (Fig. 18B). These deletions caused a frame shift and a premature stop codon in the GH7 coding sequence, thus preventing the expression of functional GH7 and tdTom proteins. In contrast, sequencing of a G0 individual with both midgut and eye fluorescence showed both the correct orientation and an intact GH7 open reading frame compared to the reference sequence (Fig. 18C) resulting in detectable tdTom fluorescence in midgut cells. Unfortunately, this G0 animal did not produce any knocked-in G1 offspring with midgut fluorescence.

Thankfully, knocked-in individuals with strong fluorescent signals in their midgut glands were identified from another G0 founder. Surprisingly, some of these animals exhibited eye fluorescence while some of their siblings did not exhibit fluorescence in their eyes (Fig. 17C,D, lower images). Three of these G2 animals underwent genotyping using four different forward primers (provided by Evagelia Stamataki) that bind specifically to the 5' UTRs of four different GH7 genes (a, b, c, and d) combined with the same reverse primer in tdTom. Only the primer specific to the GH7b locus yielded a PCR amplicon with sequencing data revealing a 9bp deletion (a multiple of 3), preserving the GH7 open reading frame (Fig. 18D). Despite the missing three amino acids (in a region unrelated to the GH7 active site), this deletion apparently leads to GH7 and nuclear tdTom expression and fluorescence in the midgut cells. Finally, the few nucleotide changes upstream of the sgRNA site between the reference GH7a sequence in the knock-in plasmid and the genotyped GH7b coding sequence correspond to synonymous single nucleotide polymorphisms between the GH7a and GH7b genes.

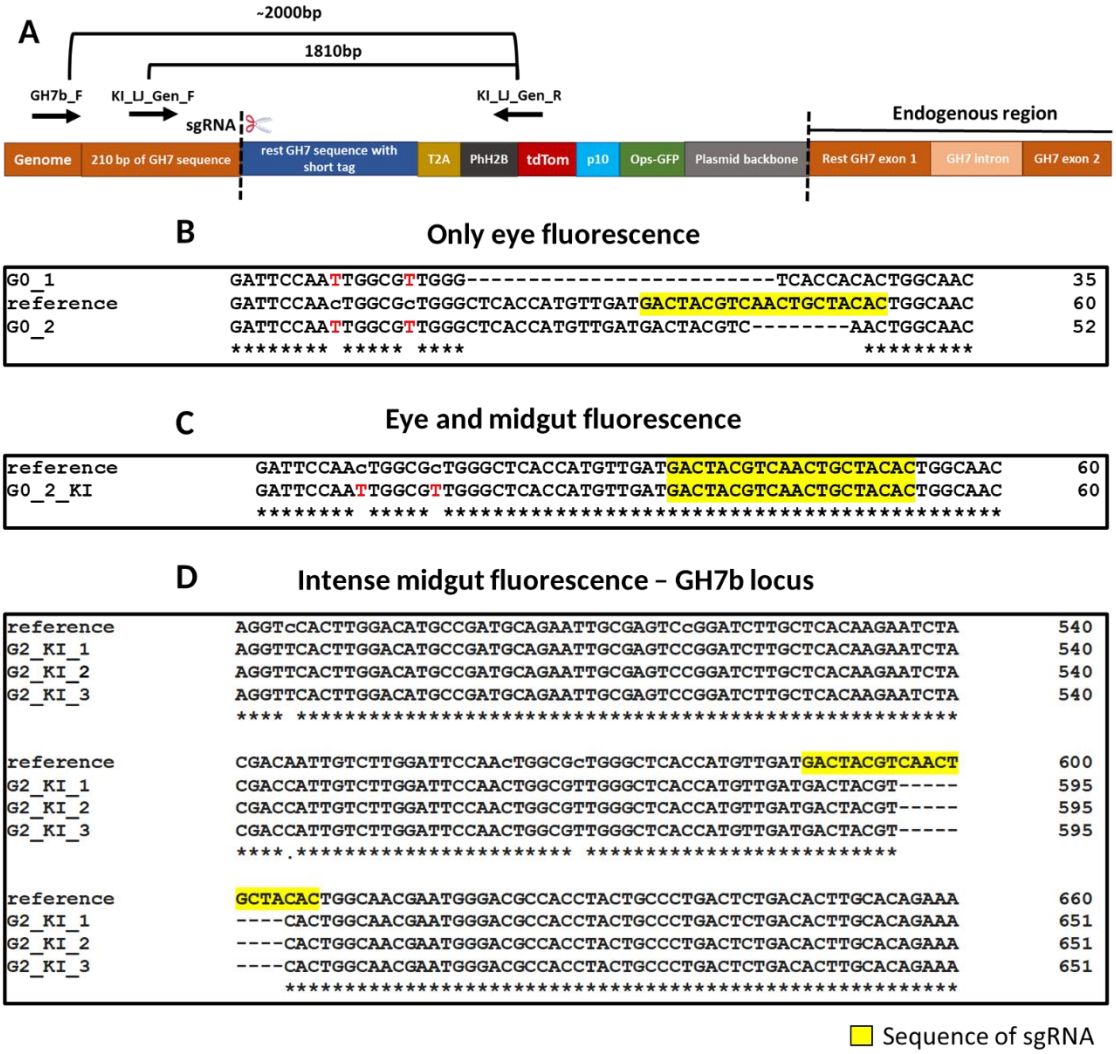

Figure 18. Genotyping of CRISPR knock-in events. *(A) Schematic representation of a successful integration event within a GH7 locus and the specific primers used in genotyping. (B) Knock-in events with out-of-frame deletions in G0s exhibiting only Opsin-EGFP eye fluorescence. (C) Knock-in event with in-frame insertion, probably in the GH7b locus, in G0#2 exhibiting both Opsin-EGFP eye fluorescence and tdTom midgut fluorescence. (D) Knock-in events with in-frame deletions in the GH7b locus in G2s from G0#3 founder exhibiting tdTom midgut fluorescence but not Opsin-EGFP eye fluorescence. The sequence targeted by the sgRNA is highlighted in yellow. The reference sequence corresponds to GH7 sequence in the plasmid used in the CRISPR knock-in experiments.*

Discussion

While model organisms like *Drosophila* and mice have long been instrumental in advancing our knowledge of genetic and developmental processes, researchers are increasingly turning to novel models, such as *Parhyale hawaiensis*, to explore unique biological phenomena and bridge gaps in phylogenetic diversity. This organism is highly adaptable and easy to cultivate in laboratories, making it an excellent model for studying development, evolution, and genetics.

The study of lignocellulose digestion in marine crustaceans, particularly in the genetic model organism *Parhyale hawaiensis*, represents a fascinating and multi-dimensional field of research that holds great promise for both scientific understanding and practical applications. Moreover, its digestive system lacking symbiotic microorganisms offers an opportunity to investigate *Parhyale* enzymatic toolkit for lignocellulose digestion. Researchers have found enzymes, including GH7 family enzymes, which were previously only known in fungi and protozoans but are now discovered in *Parhyale* and other crustaceans. This research lays the groundwork for understanding the genetic basis of lignocellulose digestion in *Parhyale* and potentially other organisms.

In addition to that, this study has significant implications for biotechnological applications, particularly in the field of lignocellulose hydrolysis for biofuel production. By exploring the genetics and biochemistry of these enzymes, it may contribute to more efficient and sustainable methods for converting lignocellulosic materials into biofuels, addressing the global need for alternative and environmentally friendly energy sources. In summary, the research of *Parhyale hawaiensis* has the potential to advance our understanding of genetic and developmental processes and revolutionize lignocellulose digestion, offering sustainable solutions in the era of biofuels and renewable energy sources.

This work, outlined in previous chapters of this thesis, showcases a journey in genetic research, specifically focused on the investigation of glycosyl hydrolase PhGH7 in the *Parhyale* genome. The initial identification of multiple GH7 genes in the *Parhyale* genome, with significant variations in expression levels, led to the unexpected discovery of additional, unannotated GH7 genes, making the genetic landscape even more complex. Loss-of-function genetic approaches (e.g., by CRISPR/Cas9 knock-out) are challenging for very similar paralogous genes with redundant function. Therefore, a gain-of-function genetic strategy was selected involving conditional overexpression of a tagged PhGH7 gene, using transgenesis vectors and a heat-inducible system. The study explored the use of different tagging cassettes for visualization and purification of the GH7 protein.

However, the conditional, heat-inducible system for overexpression of GH7 with added tags had a potential risk, because of the known leakiness of heat-shock promoters even at low temperatures, which could potentially express the tagged cellobiohydrolase of interest not only in the endodermal midgut cells but also in other tissues lead to cell toxicity and mortality of transgenic individuals. Possibly, due to this fact, it was necessary to screen and examine significantly larger numbers of animals, far exceeding the typical transgenesis experiments in *Parhyale*, to create transgenic lines with the conditionally overexpressed GH7 tagged gene.

The glycosyl hydrolase activity in both wild-type and transgenic *Parhyale* was assessed, with the application of biochemical methods utilizing cellobiohydrolase-specific substrates. In wild-type midgut extracts, cellulolytic activity is effected by multiple glycosyl hydrolases that are

abundant in the *Parhyale* digestome. This baseline cellulolytic activity was enhanced upon overexpression of the long-tagged GH7 protein after heat-shock treatment.

A number of challenges were faced during the process of establishing and amplifying these transgenic lines, resulting in a small number of transgenic individuals available for biochemical experiments in the context of this Master thesis. A single systematic effort for the purification of the long-tagged GH7 revealed the presence of the overexpressed protein only in the soluble fraction of the midgut. However, after heat-inducible overexpression, the fluorescently tagged GH7 can be observed throughout the entire *Parhyale* embryo. The absence of soluble long-tagged GH7 in the non-midgut tissues suggests that this protein can be processed properly only in midgut cells (e.g., post-translational modification, folding and secretion), while in other tissues it forms insoluble aggregates. Consequently, the biochemical investigations were conducted with midgut-derived extracts.

In the Western blot analyses, we tried unsuccessfully anti-actin, anti-tubulin and anti-GAPDH antibodies as loading controls after stripping the membranes incubated with the anti-mCherry antibody. It is speculated that the content or pH of the midgut extract may interfere with the detection of these loading controls, as more recent Western blotting experiments with extensive washes of dissected midguts have solved this problem (Evangelia Stamataki, personal communication). As a result, in my studies the loading of equal amounts of protein extracts in Coomassie-stained SDS-PAGE gels and Western blot experiments were based on protein quantification with the Bradford assay.

It is noteworthy that the H2B-RFP transgenic line used as a control in our experiments, encodes the Drosophila H2B protein fused with the mRFPRuby version of the monomeric RFP protein. The expected molecular weight is 41 kDa for the fusion protein and 27 kDa for plain RFP. Analyzing separately the midgut and rest-tissue derived protein extracts revealed a 27kDa band in the midgut and a 41kDa band in the rest-tissue samples. This result suggests that the H2B-RFP fusion may undergo some kind of cleavage in midgut cells (maybe in the linker region between H2B and RFP), unlike in the rest body tissues where it remains intact. This cleavage of fluorescent protein fusions has been also reported in other research studies (Huang et al., 2014). Despite this caveat, the midgut-derived extracts of the H2B-RFP line served their purpose as negative controls in the purification and biochemical experiments.

Importantly, the single purification experiment of the overexpressed long-tagged GH7 from 23 transgenic adults available during this Master thesis yielded a few micrograms of affinity-purified protein. The purified *Parhyale* long-tagged GH7 displayed a comparable activity to 1500-times higher quantity of the HjCel7A commercial fungal cellobiohydrolase at 30°C and 380-times higher amount of HjCel7A at 45°C. These results demonstrate the potential of *Parhyale* GH7 as an enzyme for future biotechnological applications. Current experiments involve the expansion of the populations of transgenic animals encoding the long-tagged GH7 and the short-tagged GH7 to further optimize the streptavidin purification (e.g., with the use of a column instead of beads), the TEV-mediated tag removal, and the characterization of PhGH7 enzymatic kinetics. Future comparisons of the performance of the purified PhGH7 against other characterized GH7 cellobiohydrolases besides HjCel7A and measuring enzyme's kinetic properties under a variety of conditions and with different substrates, could further support our current findings regarding PhGH7's potential.

Another significant aspect of this research involved the establishment of a CRISPR/Cas9-based knock-in method for introducing tags into the endogenous GH7 genes of *Parhyale*. The

objective of this approach was to express and monitor PhGH7 within its native context. This was a challenging process aided by the phenotypic markers provided by PhOpsin-EGFP for plasmid integration and GH7::tdTom for correct orientation and open reading frame. Genotyping of transgenic animals' genomic DNA were essential for confirming successful integration events and identifying the knocked-in GH7 genes. Expansion of the single recovered knocked-in line is currently underway with the intention of conducting the same purification and biochemical experiments described previously. The comparative analysis between the heat-inducible overexpressed GH7 and the endogenous GH7 expressed under native conditions will shed light on the most appropriate methodology in *Parhyale*. As another future plan, it will be really interesting to employ heterologous bacterial systems and establish *Parhyale* cell lines for protein expression and purification that will be benchmarked against the *in vivo* produced cellulases.

Overall, this study demonstrates the significance of combining various -omics analyses and genetic manipulation techniques to explore complex biological processes. It underlines the unexpected discoveries that can arise during scientific research and the potential applications of these findings in addressing pressing global challenges, such as sustainable energy production. Last but not least, this work emphasizes the importance of interdisciplinary research at the intersection of genomics, proteomics, and biotechnology, for the purpose of an initial assessment of *Parhyale*'s glycosyl hydrolases activity, unleashing their substantial potential for use in wood digestion and other green biotechnological applications.

Bibliography

Arantes, V., & Goodell, B. (2014). Current Understanding of Brown-Rot fungal Biodegradation Mechanisms: A review. In *Acs Symposium Series* (pp. 3–21). https://doi.org/10.1021/bk-2014-1158.ch001

Artal, M. C., Santos, A. D., Henry, T. B., & De Aragão Umbuzeiro, G. (2017). Development of an acute toxicity test with the tropical marine amphipod Parhyale hawaiensis. *Ecotoxicology*, *27*(2), 103–108. https://doi.org/10.1007/s10646-017-1875-3

Averof, M. (2022). The crustacean Parhyale. *Nature Methods*, *19*(9), 1015–1016. https://doi.org/10.1038/s41592-022-01596-y

Bruce, H. S., & Patel, N. H. (2020). Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. *Nature Ecology and Evolution*, *4*(12), 1703–1712. https://doi.org/10.1038/s41559-020-01349-0

Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology*, *12*(3), 168–180. https://doi.org/10.1038/nrmicro3182

Brunet, M., Arnaud, J., & Mazza, J. (1994). Gut structure and digestive cellular processes in marine Crustacea. *Oceanography and Marine Biology*, *32*, 335–367. http://www.vliz.be/en/imis?refid=79580

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., & Rahmanpour, R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports*, *28*(12), 1883. https://doi.org/10.1039/c1np00042j

Bui, T. H. H., & Lee, J. (2015). Endogenous cellulase production in the leaf litter foraging mangrove crab Parasesarma erythodactyla. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *179*, 27–36. https://doi.org/10.1016/j.cbpb.2014.09.004

Chang, W. H., & Lai, A. G. (2018). Mixed evolutionary origins of endogenous biomass-depolymerizing enzymes in animals. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4861-0

Cragg, S. M., Beckham, G. T., Bruce, N. C., Bugg, T. D. H., Distel, D. L., Dupree, P., Etxabe, A. G., Goodell, B., Jellison, J., McGeehan, J., McQueen-Mason, S. J., Schnorr, K., Walton, P. H., Watts, J. E. M., & Zimmer, M. (2015). Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion in Chemical Biology*, *29*, 108–119. https://doi.org/10.1016/j.cbpa.2015.10.018

Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., Blanch, H. W., Clark, D. S., & Robb, F. T. (2011). Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. *Nature Communications*, *2*(1). https://doi.org/10.1038/ncomms1373

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., & Foust, T. D. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, *315*(5813), 804–807. https://doi.org/10.1126/science.1137016

Huang, L., Pike, D. H., Sleat, D. E., Nanda, V., & Lobel, P. (2014). Potential pitfalls and solutions for use of fluorescent fusion proteins to study the lysosome. PLOS ONE, 9(2), e88893. https://doi.org/10.1371/journal.pone.0088893

Kao, D., Lai, A. G., Stamataki, E., Rošić, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A., Pouchkina-Stancheva, N., Sémon, M., Grillo, M., Bruce, H. S., Kumar, S., Siwanowicz, I., Le, A., Lemire, A., Eisen, M. B., Extavour, C. G., Browne, W. E., Wolff, C., . . . Aboobaker, A. A. (2016). The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife*, *5*. https://doi.org/10.7554/elife.20062

Kern, M., McGeehan, J., Streeter, S., Martin, R. N. A., Beßer, K., Elias, L., Eborall, W., Malyon, G. P., Payne, C. M., Himmel, M. E., Schnorr, K., Beckham, G. T., Cragg, S. M., Bruce, N. C., & McQueen-Mason, S. J. (2013). Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(25), 10189–10194. https://doi.org/10.1073/pnas.1301502110

King, A. J., Cragg, S. M., Li, Y., Dymond, J., Guille, M., Bowles, D. J., Bruce, N. C., Graham, I. A., & McQueen-Mason, S. J. (2010). Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), 5345–5350. https://doi.org/10.1073/pnas.0914228107

Konstantinides, N., & Averof, M. (2014). A common cellular basis for muscle regeneration in arthropods and vertebrates. *Science*, *343*(6172), 788–791. https://doi.org/10.1126/science.1243529

Kontarakis, Z., & Pavlopoulos, A. (2014). Transgenesis in non-model organisms: the case of Parhyale. In *Methods in molecular biology* (pp. 145–181). https://doi.org/10.1007/978-1-4939-1242-1_10

Kwiatkowski, E., Schnytzer, Y., Rosenthal, J. J. C., & Emery, P. (2023). Behavioral circatidal rhythms require Bmal1 in Parhyale hawaiensis. *Current Biology*, *33*(10), 1867-1882.e5. https://doi.org/10.1016/j.cub.2023.03.015

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C., & Patel, N. H. (2016). CRISPR/CAS9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Current Biology*, *26*(1), 14–26. https://doi.org/10.1016/j.cub.2015.11.021

Myers, A. A. (1985). Shallow-water, coral reef and mangrove Amphipoda (Gammaridea) of Fiji. *Records of the Australian Museum*, *Supplement*, *5*, 1–143. https://doi.org/10.3853/j.0812-7387.5.1985.99

O'Connor, R. M., Fung, J. M., Sharp, K., Benner, J. S., McClung, C., Cushing, S., Lamkin, E. R., Fomenkov, A., Henrissat, B., Londer, Y. Y., Scholz, M., Pósfai, J., Malfatti, S., Tringe, S. G., Woyke, T., Malmstrom, R. R., Coleman-Derr, D., Altamia, M. A., Dedrick, S., . . . Distel, D. L. (2014). Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(47). https://doi.org/10.1073/pnas.1413110111

Pavlopoulos, A., Kontarakis, Z., Liubicich, D. M., Serano, J. M., Akam, M., Patel, N. H., & Averof, M. (2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13897–13902. https://doi.org/10.1073/pnas.0902804106

Pollegioni, L., Tonin, F., & Rosini, E. (2015). Lignin-degrading enzymes. *FEBS Journal*, *282*(7), 1190–1213. https://doi.org/10.1111/febs.13224

Rallis, J., Kapai, G., & Pavlopoulos, A. (2021). Parhyale hawaiensis, Crustacea. In *CRC Press eBooks* (pp. 289–306). https://doi.org/10.1201/9781003217503-16

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., & Patel, N. H. (2009). The Crustacean *Parhyale hawaiensis*: A New Model for Arthropod Development. *CSH Protocols*, *2009*(1), pdb.emo114. https://doi.org/10.1101/pdb.emo114

Scharf, M. E. (2015). Omic research in termites: an overview and a roadmap. *Frontiers in Genetics*, *6*. https://doi.org/10.3389/fgene.2015.00076

Serano, J. M., Martin, A., Liubicich, D. M., Jarvis, E., Bruce, H. S., La, K., Browne, W. E., Grimwood, J., & Patel, N. H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. *Developmental Biology*, *409*(1), 297–309. https://doi.org/10.1016/j.ydbio.2015.10.029

Simmons, C. W., Reddy, A. P., D'haeseleer, P., Khudyakov, J., Billis, K., Pati, A., Simmons, B. A., Singer, S. W., Thelen, M. P., & VanderGheynst, J. S. (2014). Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnology for Biofuels*, *7*(1). https://doi.org/10.1186/s13068-014-0180-0

Stamataki, E., & Pavlopoulos, A. (2016). Non-insect crustacean models in developmental genetics including an encomium to Parhyale hawaiensis. *Current Opinion in Genetics & Development*, *39*, 149–156. https://doi.org/10.1016/j.gde.2016.07.004

Sun, D. A., & Patel, N. H. (2019). The amphipod crustacean *Parhyale hawaiensis* : An emerging comparative model of arthropod development, evolution, and regeneration. *Wiley Interdisciplinary Reviews-Developmental Biology*, *8*(5). https://doi.org/10.1002/wdev.355

Wolff, C., Tinévez, J., Pietzsch, T., Stamataki, E., Harich, B., Guignard, L., Preibisch, S., Shorte, S., Keller, P. J., Tomančák, P., & Pavlopoulos, A. (2018). Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb. *eLife*, *7*. https://doi.org/10.7554/elife.34410