Exploring the Roles of Horizontal Gene Transfer in Metazoans

By

Dongliang Chen

June, 2016

Director: Dr. Jinling Huang

DEPARTMENT OF BIOLOGY

Horizontal gene transfer (HGT; also known as lateral gene transfer, LGT) refers to the movement of genetic information between distinct species by overcoming normal mating barriers. Historically HGT is only considered to be important in prokaryotes. Some researchers believe that eukaryotes have sexual recombination and HGT is insignificant. However, HGT has also been found to play roles in many aspects of eukaryotic evolution, like parasitism and the colonization of land by plants, although at lower frequencies than in prokaryotes.

In this dissertation, I first estimated the scope of HGT in 16 selected metazoan species by genome screening using AlienG. These species are sampled to represent major lineages of metazoans. Among all the 16 species, *Nematostella vectensis* (4.08%) has the highest percentage of HGT genes, while parasitic *Schistosoma japonicum* (0.47%) ranks the lowest. In order to find out which factors are correlated with HGT rates in different species, living habitat, diet, lineage group and reproductive type were analyzed in a statistical framework.

In Chapter 3 and Chapter 4, *Ciona intestinalis* and *Trichoplax adhaerens* were chosen as models to investigate horizontally acquired genes. Tunicate cellulose synthase was discovered to originate from green algae, instead from bacteria as found in previous studies. 43 genes of 21 families in *T. adhaerens* were found to be horizontally acquired. The functions and impacts of acquired genes on *T. adhaerens* are also discussed.

EXPLORING THE ROLES OF HORIZOTNAL GENE TRANSFER IN

METAZOANS

A Dissertation Presented to

The Faculty of the Interdisciplinary Doctoral Program in Biological Sciences

The Brody School of Medicine, East Carolina University

In Association with the Department of Biology, Thomas Harriot College of Arts and

Sciences

Submitted in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Interdisciplinary Doctoral Program in Biological Sciences (IDPBS)

By

Dongliang Chen

June, 2016

Exploring the Roles of Horizontal Gene Transfer in Metazoans

by

Dongliang Chen

APPROVED BY:

DIRECTOR OF DISSERTATION:_____

Jinling Huang, PhD

COMMITTEE MEMBER:_____

Chris Balakrishnan, PhD

COMMITTEE MEMBER:_____

Qing Ding, PhD

COMMITTEE MEMBER:_____

John Stiller, PhD

COMMITTEE MEMBER:_____

Kyle Summers, PhD

CHAIR OF THE DEPARTMENT OF BIOLOGY:

_____

Jeff McKinnon, PhD

DEAN OF THE GRADUATE SCHOOL: _____

Paul J. Gemperline, PhD

Acknowledgements

I have spent seven years in Biology department of East Carolina University (ECU) in pursuit of a Ph.D. degree. Greenville, NC has literally become my second hometown and I spent much more time in this peaceful little city than any other city back in China, except for my birthplace. Not only did I get my Ph.D. degree from ECU, I also harvested two little ones, who make me cheerful, inspired and energetic every day. It would be an extremely hard time if I had not received so much help from my advisor, committee members, colleges, friends and families.

The first one I would like to appreciate is my advisor, Dr. Jinling Huang. I toke two of his courses before I chose to join his lab, and I found his courses very helpful. He is good at changing abstract evolutionary concepts and programming skills to interesting intercommunications in class. When I joined his lab, I had nearly no background in evolutionary biology as well as programming. When I encountered a problem, he always showed me the shortest way to solve my problem. He also recommended lots of publications and books for me to read through, so that I can finish my dissertation in three and a half years.

I also would like to give my appreciation to my committee members, Dr. John Stiller, Dr. Kyle Summers, Dr. Qin Ding and Dr. Christopher Balakrishnan, for their patience and understanding, for their inspiring guidance and thought-provoking suggestions. I took one or two courses from each of them and their courses help finish my dissertation from different aspects: John and Kyle in basic evolutionary

biology, Qin Ding in data mining and AlienG usage and Chris in next-generation sequencing and programming.

Four postdoc fellows from Dr. Huang's lab, Dr. Guiling Sun, Dr. Jipei Yue, Dr. Zefeng Yang and Dr. Ticao Zhang, showed me how to conduct the research of horizontal gene transfer (HGT). I worked together with Dr. Zefeng Yang on Chapter 4, HGT in *Trichoplax*, and I collaborate with Dr. Ticao Zhang on part of Chapter 3, cellulose synthase (CesA) in tunicates. Qia Wang just joined our lab this year, and I really appreciate his inspiring discussion with me about recent HGT study in tardigrades.

I am extremely grateful to Department of biology and Interdisciplinary Doctoral Program in Biological Sciences (IDPBS) for letting me join this friendly community. Two graduate student supervisors, Dr. Terry West and Dr. Edmund Stellwag, gave me so much advice to study successfully as an international student. Two secretaries, Barbara Beltran and Joyce Beatty helped me to get over all the problems I met during study. I also want to say thanks to Dr. Jeff McKinnon, Dr. Jean-Luc Scemama, Dr. Paul Hager, Jianfen Lu and many others.

Dr. Karl Wuensch, a fun German from Department of psychology, lit up my interests in studying statistics. Dr. Peng Xiao from Department of mathematics made the 6000 course statistical inference less frightening. Dr. William Mallett always kindly reminded me of expiration of I-20 and showed me how to apply for OPT (Optional Practical Training). Former Greenville Mayor Patricia Dunn also provided much convenience for international students, making me join this community quickly

and less homesick. Without their help, my study and life in ECU would not be so smooth.

Last but not least, great thanks go to my parents, my wife Yanqiong Zhang and two little ones, Tony and Simon. My parents sacrificed their retirement life in China and flied to USA to take care of the kids. Twelve hours of flight and jet-lag was painful for them. What is even harder is they nearly have no communication with others, since they cannot speak English. My wife had two babies when she was in pursuit of a Ph.D. degree, and I think I could never make up what she had sacrificed these years. A big smile from Tony and Simon is the best present after a long day of hard working. I will spend every minute of my life to love them!

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Background

Introduction

Unlike vertical gene transfer (VGT), which transmits genetic information from parents to offspring through sexual or asexual reproduction, horizontal gene transfer (HGT; also known as lateral gene transfer, LGT) refers to the movement of genetic information between distinct species through overcoming normal mating barriers (Keeling and Palmer 2008; Syvanen 2012). In his book *On the Origin of Species*, Charles Darwin presented the concept of the Tree of Life and argued that genetic information is passed on from parents to offspring, with mutations accumulating from generation to generation, thus leading to gradual speciation (Darwin 1859). Carl Woese, however, stated that horizontally acquired variation played more important roles in early cellular evolution until vertical inheritance became a dominant force, an evolutionary stage which he referred to as Darwinian threshold (Woese 2002).

The origin of a new beneficial gene through accumulative mutations is rare, given that most mutations are deleterious. Instead, "borrowing" ready-to-use foreign genes from other species is much easier and quicker than evolving from scratch (Vogan and Higgs 2011). Moreover, prokaryotes may lose genes due to incorrect segregation, deletion or deleterious mutations (Koskiniemi, Sun et al. 2012), so HGT could be extremely beneficial since one cell can horizontally obtain lost genes from other species. HGT is significantly efficient for prokaryotes to explore new niches or to utilize new resources. Indeed, HGT has been discovered to be involved in many

aspects of prokaryotic evolution, such as spread of photosynthesis (Raymond 2009), nitrogen fixation (Cantera, Kawasaki et al. 2004; Bolhuis, Severin et al. 2010), bacterial pathogenesis (Ochman, Lawrence et al. 2000) and antibiotic resistance (Koonin, Makarova et al. 2001; Barlow 2009; Palmer, Kos et al. 2010; Andam, Fournier et al. 2011).

While it has been widely accepted that HGT contributes substantially to genome evolution in prokaryotes through three broad mechanisms: transformation, conjugation and transduction (Gogarten, Doolittle et al. 2002), it has also been found to be pivotal in many aspects of eukaryotic evolution, like parasitism (Kishore, Stiller et al. 2013) and plant colonization of land (Yue, Hu et al. 2012). Although presumably at a lower frequency than in prokaryotes, HGT is also important for certain eukaryotic organisms to adapt to specialized niches. For example, *Giardia lamblia,Trichomonas vaginalis,* and *Entamoeba histolytica* obtain bacterial genes to adapt to the niche of anaerobic metabolism (Keeling and Palmer 2008). Due to the influences of HGT in the genome evolution of all three domains of life, some even suggest the use of the term "net of life" (Hilario and Gogarten 1993) or "web of life" (Soucy, Huang et al. 2015) to replace "tree of life", which was first used as a metaphor to describe the tree-like relationship of all life on Earth by Charles Darwin.

This Chapter reviews some important concepts of HGT, such as ancient and recent HGT events, differences between intracellular gene transfer (IGT) and HGT. Advancements in HGT research are also reviewed in prokaryotes, unicellular

eukaryotes and in metazoans. Finally, a case study of HGT in subphylum Tunicata is demonstrated.

Intracellular versus horizontal gene transfer; ancient versus recent HGT

Intracellular gene transfer (IGT; also termed as endosymbiotic gene transfer, EGT) refers to gene flux to nucleus from mitochondria or plastids, which arose from an endosymbiotic event of an α-proteobacterium or a cyanobacterium, respectively (Fig. 1, upper panel). The evolution of mitochondria in eukaryotes only happened once. All eukaryotic species contain mitochondria derived from a single endosymbiotic event (Gray, Lang et al. 2004), with the exception of some parasitic linages, such as *Giardia* (Keeling and Doolittle 1997), *Trichomonas* (Germot, Philippe et al. 1996; Roger, Clark et al. 1996), *Entamoeba* (Cavalier-Smith 1993; Clark and Roger 1995) and *Microsporidia* (Keeling and McFadden 1998; Gray, Lang et al. 2004), which underwent mitochondrial reduction or loss. With the exception of chromatophores in *Paulinella* (Nowack, Melkonian et al. 2008), all plastids were ultimately derived from a single primary endosymbiosis in the ancestor of glaucophytes, red algae and green plants. These plastids from red algae and green algae were also transferred to other lineages by secondary or tertiary endosymbioses (Keeling 2004; Gould, Waller et al. 2008). After their establishment in eukaryotes, genes in endosymbionts are lost or transferred into host nuclear genomes by massive IGT, leaving only a small proportion of their original genes in the organellar genomes (Dunning Hotopp, Clark et al. 2007). The protein products of

those transferred genes are often re-imported to mitochondria or plastids via signal peptides to function in related biochemical activities (Bruce 2000; Zhang and Glaser 2002).

HGT is a dynamic process occurring continuously at different stages of evolution. Ancient HGT events have been discovered by either phylogenetic analyses of individual proteins or large-scale whole-genome comparisons among the three domains of life, i.e., Bacteria, Archaea and eukaryotes. It seems likely that most genes have been horizontally transferred at some time; however, extensive ancient HGT conflicts with the universal tree of life in understanding cellular evolution, since the basic tree topologies still remain the same based on genome data from three domains (Fitz-Gibbon and House 1999; Snel, Bork et al. 1999; Brown, Douady et al. 2001). Meanwhile, recent HGT is defined as HGT events happening to lower taxonomic ranks (species, genus, family etc.). Although recent HGT may offer evolutionary benefits to lineages of lower taxonomic ranks, ancient HGT can affect the long-term evolution of major lineages (Andersson, Sjogren et al. 2003; Kishore, Stiller et al. 2013).

By studying anaerobic protists diplomonads, 15 genes in the genome of *Spironucleus barkhanus* and/or *Giardia lamblia* were shown to be derived from HGT (Andersson, Sjogren et al. 2003). Half of those genes are related to the anaerobic lifestyle of diplomonads, which suggests that the aerobic common ancestor of *S. barkhanus* and *G. lamblia* adapted to an anaerobic niche, partially due to horizontal

acquisition of those genes. Horizontal transferring of Set8 from an animal host and Ashr3 from a plant host, along with horizontal transfer of cytoadhesion domains, contributes to the transition of the free-living ancestor of apicomplexans to an obligate parasite (Kishore, Stiller et al. 2013).

In two recent papers, it was suggested that IGT accounted for most genes of prokaryotic origins in the eukaryotic genomes (Katz 2015; Ku, Nelson-Sathi et al. 2015). In Ku's study, 956,053 protein sequences from 55 eukaryotic genomes, 5,793,897 from 1,847 bacterial genomes and 309,128 from 134 archaean genomes were clustered for comparisons. 2,585 sequence families were identified, in which all protein sequences were homologous and there are at least two eukaryotic genomes and at least five prokaryotic genomes. Among all these sequence families, most were interpreted as derived from plastids or mitochondria. Even if lineage-specific HGT events had happened, transferred genes were secondarily lost in the recipient genome.

Katz's study shows that 1,138 genes were found as HGT by studying 13,465 gene trees using a phylogenomic pipeline. Most of 1,138 interdomain gene transfer events were recent (Katz 2015). Sequences from 487 eukaryotes, 303 bacteria and 118 archaea were used in this study. Except for IGT, few ancient transferred genes were found to be retained in eukaryotic genomes, indicating that most interdomain HGTs were lost in the evolution of eukaryotes unless they were involved in endosymbiotic gene transfer (Wolf and Koonin 2013). Both papers indicated that

5

horizontally acquired genes from other species are not stable in recipient genomes

and are prone to be lost shortly afterwards, which is consistent with the previous

model that gene loss dominates the evolution of eukaryotic genomes (Wolf and

Koonin 2013).

HGT in prokaryotes

The prevalence and significance of HGT in bacteria had never been recognized until

researchers noticed the spread of multidrug resistance among different bacterial

species in 1950's (Davies and Davies 2010). Although accumulation of point

mutations brings innovation to bacterial genomes on an evolutionary timescale,

horizontal acquisition of ready-to-use genes from other species enables bacteria to

adapt to new niches more rapidly. Usually HGT among bacteria is accomplished

through three mechanisms: transformation (i.e., uptake of naked DNA directly from

the environment), transduction mediated through bacteriophages, and conjugation.

These mechanisms can lead to the spread antibiotic resistance, virulence attributes

and metabolic properties (Ochman, Lawrence et al. 2000). In 2008, a fluorescent

protein fusion SeqA-YFP was used to visualize the horizontal transfer of the F

plasmid through conjugation in *Escherichia coli* (Babic, Lindner et al. 2008), which

offers a vivid proof of HGT in prokaryotes.

Extensive HGT events also occur in Archaea. After studying HGT in 17 bacterial

and 7 archaeal genomes, Garcia-Vallve et al. pointed out that the archaea, as well

as nonpathogenic bacteria, had the highest percentages of HGT (Garcia-Vallve,

Romeu et al. 2000). HGT is also shown to correspond to the origin of major archael

lineages (Nelson-Sathi, Dagan et al. 2012; Nelson-Sathi, Sousa et al. 2015),

although this finding has been refuted by other researchers (Groussin, Boussau et al.

2016). HGT events can happen in both directions, although it seems that archaea

are reported more often as donors. Reverse transcriptases (retroids), which have

only been found in eukaryotes and bacteria, were reported to be transferred to

methanogenic archaea *Methanosarcina acetivorans* and *M. mazei* from bacteria

(Rest and Mindell 2003). On the other hand, archaea can also transfer their genes to

facilitate the colonization of bacteria (Gophna, Charlebois et al. 2004) and

eukaryotes (Andersson, Sjogren et al. 2003) to new niches.

HGT in eukaryotes

While HGT has been thought to be a driving force in the evolution of prokaryotes, the

study of HGT in eukaryotes is still under debate. However, as more and more

eukaryotic genomes have been sequenced, there are an increasing number of

reports on HGT events in eukaryotes (Andersson 2005; Huang 2013).

There have been many reports of horizontally transferred genes in protist

genomes, such as in Apicomplexa (Huang, Mullapudi et al. 2004), Amoebozoa

(Loftus, Anderson et al. 2005; Watkins and Gray 2006) and Ciliates (Ricard, McEwan

et al. 2006). The frequency of HGT events in protistan genomes is normally lower

than that in bacterial genomes, although in some extreme cases it is comparable.

For example, it has been estimated that 4% of the genes of rumen-dwelling ciliates

are of bacterial origin (Ricard, McEwan et al. 2006). In the choanoflagellate *Monosiga brevicollis*, approximately 4.4% of the nuclear genome was horizontally acquired from algae, bacteria, and other prey (Sun, Yang et al. 2010; Tucker 2013; Yue, Sun et al. 2013). At least 5% of the genome of extremophilic red alga *Galdieria sulphuraria* was horizontally acquired (Schonknecht, Chen et al. 2013) and hundreds (5.4% - 9.3%) out of 8,355 genes of another red alga, *Porphyridium purpureum,* may be due to EGT or HGT (Bhattacharya, Price et al. 2013). Many protists either consume bacteria or are frequently exposed to bacterial DNA (such as parasitism), which offers them a chance to obtain bacterial genes horizontally.

It has been widely accepted that HGT happens in unicellular eukaryotes, due to their close contact with phytoplankton and/or bacteria and phagotrophic lifestyles of many species. However, some researcher are still skeptical of HGT in multicellular eukaryotes, like land plants and animals. Two potential mechanisms are consistent with this phenomenon. Firstly, multicellular eukaryotes can still acquire genes through feeding activities (Ni, Yue et al. 2012), parasitism (Scholl, Thorne et al. 2003) and endosymbiosis (Dunning Hotopp, Clark et al. 2007; Nikoh, Tanaka et al. 2008), although at low frequency. Secondly, since HGT is a dynamic progress, which could happen at any evolutionary stage, vertically inherited genes could be derived from ancient HGT events to the unicellular ancestors of modern multicellular eukaryotes (Huang and Gogarten 2006; Yue, Hu et al. 2012).

In addition, HGT has been frequently studied in fungi. Functions of those horizontally acquired genes include changes to the mating system (Inderbitzin, Harkness et al. 2005; Paoletti, Buck et al. 2006; Milani, Lawrence et al. 2012), small molecules metabolism (Hall, Brachat et al. 2005; Hall and Dietrich 2007; Slot and Hibbett 2007; Wisecaver, Slot et al. 2014) and virulence factors (Friesen, Stukenbrock et al. 2006; Richards, Lang et al. 2011). While interspecific transfer of virulence factors is quite common in bacterial pathogens, Friesen et al. provided the first report of this event in fungal pathogens. An 11kb region encoding a critical virulence factor *ToxA*, was transferred from a virulent fungal species *Stagonospora nodorum* to a previously non-virulent fungal species *Pyrenophora tritici-repentis* as recent as 70 years ago (Friesen, Stukenbrock et al. 2006). Recent studies review that virulence genes can be horizontally transferred among diverse plant pathogenic fungi and plant-associating bacteria (Gardiner, McDonald et al. 2012; Qiu, Cai et al. 2016).

On the other hand, fungi are also able to obtain genes from prokaryotes for niche adaptation. Bacterial enzyme 1-aminocyclopropane-1-carboxylate (ACC) deaminase is encoded by gene *acdS* and can increase the plant's resistance to stress and promote root growth (Broekaert, Delaure et al. 2006). This gene is normally found in plant-interacting bacteria strains (Blaha, Prigent-Combaret et al. 2006), and can also be found in some fungi (Yao, Ose et al. 2000; Viterbo, Landau et al. 2010). Phylogenomic analysis of *acdS* gene indicates the existence of 61 homologues in fungi and four in oomycetes (Bruto, Prigent-Combaret et al. 2014). The authors of

this study also claimed that *acdS* genes were acquired multiple times from various bacterial phyla to different fungi/oomycetes, implying that the transfer of this gene occurred independently and frequently.

HGT in metazoans

It has been proposed that HGT plays only a minor role in metazoan evolution, due to the sequestered germline in animals. However, as reviewed by Boto (Boto 2014), HGT has been widely studied in sponges, cnidarians, rotifers, nematodes, arthropods and mollusks, in which new traits were acquired to explore new niches.

Only one case of a horizontally derived gene was reported in Porifera (Jackson, Macis et al. 2011). The gene was acquired by the ancestor of *Astroclera willeyana* and *Ampimedon quenslandica* and is involved in biomineralization processes. Numerous HGT events were postulated to occur in Cnidarians (Starcevic, Akthar et al. 2008; Bilewitch and Degnan 2011), especially those in *Hydra magnipapillata* (Chapman, Kirkness et al. 2010).

In one of the most interesting cases in metazoans, there is evidence for massive HGT in bdelloid rotifers. Genes of bacterial, fungal and plant origin account for 10% of the genes in *Adineta ricciae* (Boschetti, Carr et al. 2012) and 8% in *Adineta vaga* (Flot, Hespeels et al. 2013). These acquired genes are usually located in the chromosomal subtelomeric regions (Gladyshev, Meselson et al. 2008). Another intriguing case occurred in *Hypsibius dujardini*, a member of the phylum Tardigrada.

Nearly one sixth of the genome of *H. dujardini* was claimed to be derived from

bacteria, plants, fungi and archaea (Boothby, Tenlen et al. 2015). Those horizontally

transferred genes may help to explain the high tolerance to extreme stress, including

desiccation (Welnicz, Grohme et al. 2011), extreme temperatures (Hengherr,

Worland et al. 2009), intensive radiation (Jonsson, Harms-Ringdahl et al. 2005;

Horikawa, Cumbers et al. 2013; Beltran-Pardo, Jonsson et al. 2015), organic

solvents (Ramløv and Westh 2001) and low gravity (Jonsson, Rabbow et al. 2008) in

this taxon. However, reanalysis of the sequencing data of tardigrades pointed out

that rate of HGT may be exaggerated in the original paper, probably due to

contamination of bacteria (Bemm, Weiss et al. 2016; Koutsovoulos, Kumar et al.

2016).

Plant-parasitic nematodes also acquired foreign genes to help establish their

parasitic lifestyle. These genes encode cellulase (Dieterich, Clifton et al. 2008;

Mayer, Schuster et al. 2011; Schuster and Sommer 2012), xylanases, pectate lyases,

polygalacturonases (Danchin, Rosso et al. 2010), vitamin B biosynthesis related

proteins (Craig, Bekal et al. 2008), and some other proteins involved in plant cell

defense suppression and detoxification (Haegeman, Jones et al. 2011). Those non-

metazoan genes comprised at least 3.34% of nematode genomes (Paganini,

Campan-Fournier et al. 2012).

The study of HGT genes of algal origin in kleptoplastic mollusks is controversial.

Based on evidence from PCR (Rumpho, Worful et al. 2008) or transcriptome studies

(Pierce, Fang et al. 2012), some researchers agreed that genes of algal origin were expressed in sea slugs. On the other hand, some other studies failed to detect transcripts corresponding to algal genes in *Elysia timida*, *Plakobranchus ocellatus* (Wagele, Deusch et al. 2011) or *E. chlorotica* (Pelletreau, Bhattacharya et al. 2011). A recent study showed no evidence of algal gene expression in *E. chlorotica* fertilized eggs, implying that HGT may not have provided genes crucial to the evolution of photosynthesis (Bhattacharya, Pelletreau et al. 2013).

Fragments of many endosymbiotic bacterial genomes were found to be transferred into those of arthropods. However, most of the genes in the transferred fragments became pseudogenes and exhibited low rates of transcriptional expression (Kondo, Nikoh et al. 2002; Nikoh, Tanaka et al. 2008), which obscured the importance of these genes for the host.

HGT in tunicates

Tunicata (once called Urochordata) is part of the phylum Chordata, which is characterized by its dorsal nerve cords and notochords. As an early branching clade of chordates, tunicates are the sister group of vertebrates. The subphylum Tunicata consists of three classes, Appendicularia, Ascidiacea and Thaliacea. *Ciona intestinalis* is an ascidian and has been widely used for molecular genetics and developmental biology due to its small genome size and easy adaptation to laboratory conditions. Occupying a relatively basal position in the clade of chordates, tunicates possess only ~14,000 protein-coding genes (Dehal, Satou et al. 2002;

12

Small, Brudno et al. 2007) relative to vertebrates, which underwent extensive gene duplications in the Cambrian explosion and now contain nearly twice the number of genes.

Tunicates are the only metazoans that express cellulose (Satoh 1994; Hirose, Kimura et al. 1999; Kimura, Ohshima et al. 2001). In ascidians and thaliaceans, cellulose microfibers are incorporated into a unique structure called tunic, which is located on top of the epidermis and has supportive and protective functions. In appendicularians, cellulose microfibers form a mucous "house", which is the primary filter in the filter-feeding process.

The biosynthesis of cellulose microfibrils is conducted by large complexes called cellulose synthase complexes, of which cellulose synthases (CesA) are the main components and key enzymes (Bowling and Brown 2008; Yin, Huang et al. 2009). In animals, the expression of CesA can only be found in tunicates, which is an interesting lineage-specific evolutionary innovation. The expression of Ci-CesA (*Ciona intestinalis* CesA) begins at the early tailbud embryo stage in epidermal cells (Matthysse, Deschet et al. 2004; Nakashima, Yamada et al. 2004), which is spatiotemporally in line with cellulose synthesis observed in vivo (Gianguzza and Dolcemascolo 1980). Insertional mutagenesis caused a drastic reduction in cellulose production and induced a "swimming juvenile" phenotype during metamorphosis (Sasakura, Nakashima et al. 2005).

Due to its absence in other metazoans and close phylogenetic affiliation with prokaryotic cellulose synthase, Ci-CesA was thought to have been acquired from a prokaryote and represented an evolutionary innovation in tunicates (Dehal, Satou et al. 2002; Matthysse, Deschet et al. 2004; Nakashima, Yamada et al. 2004; Sagane, Zech et al. 2010). While only one copy of cellulose synthase was found in *C. intestinalis* and *C.savignyi*, two copies of cellulose synthase genes were identified in another tunicate species *Oikopleura dioica,* belonging to the class Appendicularia (Fig. 2). One of the two copies functions before the metamorphosis, while the other one after the metamorphosis, indicating functional specialization in this species (Sagane, Zech et al. 2010). Copy number of CesA genes in Thaliacea has not been clarified, so two evolutionary scenarios are possible. After the bacterial cellulose synthase was horizontally acquired by the ancestor of the tunicate lineage, it is possible that either appendicularians underwent gene duplication, or that gene duplication occurred at the base of the tunicate lineage, followed by the loss of a copy in ascidians (Fig. 2).

Aside from the case study of cellulose synthase, study of ancient gene transfer from algae to the common ancestor of metazoans was also conducted using *C. intestinalis* (Ni, Yue et al. 2012). In this study, Ni et al. identified 14 gene families, including 92 individual genes using molecular phylogenetic and domain structural analyses. The functions of those HGT genes ranged from molecular transport, cellular regulation to methylation signaling, implying that HGT events occurring in the ancestral animal played an important role in the evolution of metazoans. Algal genes

detected in aplastidic eukaryotes, such as oomycete (Tyler, Tripathy et al. 2006), apicomplexa (Huang, Mullapudi et al. 2004) and ciliates (Reyes-Prieto, Moustafa et al. 2008), are often explained by the historical existence and subsequent loss of plastids in those species. Ni et al. (2012), however, suggest that those algal genes may have been acquired through feeding activities (Doolittle 1998; Yue and Huang 2012).

References

Andam, C. P., G. P. Fournier, et al. (2011). "Multilevel populations and the evolution of antibiotic resistance through horizontal gene transfer." <u>FEMS Microbiol Rev</u> **35**(5): 756-767.

Andersson, J. O. (2005). "Lateral gene transfer in eukaryotes." <u>Cell Mol Life Sci</u> **62**(11): 1182-1197.

Andersson, J. O., A. M. Sjogren, et al. (2003). "Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes." <u>Curr Biol</u> **13**(2): 94-104.

Babic, A., A. B. Lindner, et al. (2008). "Direct visualization of horizontal gene transfer." <u>Science</u> **319**(5869): 1533-1536.

Barlow, M. (2009). "What antimicrobial resistance has taught us about horizontal gene transfer." <u>Methods Mol Biol</u> **532**: 397-411.

Beltran-Pardo, E., K. I. Jonsson, et al. (2015). "Tolerance to Gamma Radiation in the Tardigrade Hypsibius dujardini from Embryo to Adult Correlate Inversely with Cellular Proliferation." <u>PLoS One</u> **10**(7): e0133658.

Bemm, F., C. L. Weiss, et al. (2016). "Genome of a tardigrade: Horizontal gene transfer or bacterial contamination?" <u>Proc Natl Acad Sci U S A</u> **113**(22): E3054-3056.

Bhattacharya, D., K. N. Pelletreau, et al. (2013). "Genome analysis of Elysia chlorotica Egg DNA provides no evidence for horizontal gene transfer into the germ line of this Kleptoplastic Mollusc." <u>Mol Biol Evol</u> **30**(8): 1843-1852.

Bhattacharya, D., D. C. Price, et al. (2013). "Genome of the red alga Porphyridium purpureum." Nat Commun **4**: 1941.

Bilewitch, J. P. and S. M. Degnan (2011). "A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function." BMC Evol Biol **11**: 228.

Blaha, D., C. Prigent-Combaret, et al. (2006). "Phylogeny of the 1-aminocyclopropane-1-carboxylic acid deaminase-encoding gene acdS in phytobeneficial and pathogenic Proteobacteria and relation with strain biogeography." FEMS Microbiol Ecol **56**(3): 455-470.

Bolhuis, H., I. Severin, et al. (2010). "Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes." ISME J **4**(1): 121-130.

Boothby, T. C., J. R. Tenlen, et al. (2015). "Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade." Proc Natl Acad Sci U S A.

Boschetti, C., A. Carr, et al. (2012). "Biochemical diversification through foreign gene expression in bdelloid rotifers." PLoS Genet **8**(11): e1003035.

Boto, L. (2014). "Horizontal gene transfer in the acquisition of novel traits by metazoans." Proc Biol Sci **281**(1777): 20132450.

Bowling, A. J. and R. M. Brown, Jr. (2008). "The cytoplasmic domain of the cellulose-synthesizing complex in vascular plants." Protoplasma **233**(1-2): 115-127.

Broekaert, W. F., S. L. Delaure, et al. (2006). "The role of ethylene in host-pathogen interactions." Annu Rev Phytopathol **44**: 393-416.

Brown, J. R., C. J. Douady, et al. (2001). "Universal trees based on large combined protein sequence data sets." Nat Genet **28**(3): 281-285.

Bruce, B. D. (2000). "Chloroplast transit peptides: structure, function and evolution." Trends Cell Biol **10**(10): 440-447.

Bruto, M., C. Prigent-Combaret, et al. (2014). "Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes." Proc Biol Sci **281**(1789): 20140848.

Cantera, J. J., H. Kawasaki, et al. (2004). "The nitrogen-fixing gene (nifH) of Rhodopseudomonas palustris: a case of lateral gene transfer?" Microbiology **150**(Pt 7): 2237-2246.

Cavalier-Smith, T. (1993). "Kingdom protozoa and its 18 phyla." Microbiol Rev **57**(4): 953-994.

Chapman, J. A., E. F. Kirkness, et al. (2010). "The dynamic genome of Hydra." Nature **464**(7288): 592-596.

Clark, C. G. and A. J. Roger (1995). "Direct evidence for secondary loss of mitochondria in Entamoeba histolytica." Proc Natl Acad Sci U S A **92**(14): 6518-6521.

Craig, J. P., S. Bekal, et al. (2008). "Analysis of a horizontally transferred pathway involved in vitamin B6 biosynthesis from the soybean cyst nematode Heterodera glycines." Mol Biol Evol **25**(10): 2085-2098.

Danchin, E. G., M. N. Rosso, et al. (2010). "Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes." Proc Natl Acad Sci U S A **107**(41): 17651-17656.

Darwin, C. (1859). "On the Origin of Species. 1." London: Murray.

Davies, J. and D. Davies (2010). "Origins and evolution of antibiotic resistance." Microbiol Mol Biol Rev **74**(3): 417-433.

Dehal, P., Y. Satou, et al. (2002). "The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins." Science **298**(5601): 2157-2167.

Dieterich, C., S. W. Clifton, et al. (2008). "The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism." Nat Genet **40**(10): 1193-1198.

Doolittle, W. F. (1998). "You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes." Trends Genet **14**(8): 307-311.

Dunning Hotopp, J. C., M. E. Clark, et al. (2007). "Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes." Science **317**(5845): 1753-1756.

Fitz-Gibbon, S. T. and C. H. House (1999). "Whole genome-based phylogenetic analysis of free-living microorganisms." Nucleic Acids Res **27**(21): 4218-4222.

Flot, J. F., B. Hespeels, et al. (2013). "Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga." Nature **500**(7463): 453-457.

Friesen, T. L., E. H. Stukenbrock, et al. (2006). "Emergence of a new disease as a result of interspecific virulence gene transfer." <u>Nat Genet</u> **38**(8): 953-956.

Garcia-Vallve, S., A. Romeu, et al. (2000). "Horizontal gene transfer in bacterial and archaeal complete genomes." <u>Genome Res</u> **10**(11): 1719-1725.

Gardiner, D. M., M. C. McDonald, et al. (2012). "Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts." <u>PLoS Pathog</u> **8**(9): e1002952.

Germot, A., H. Philippe, et al. (1996). "Presence of a mitochondrial-type 70-kDa heat shock protein in Trichomonas vaginalis suggests a very early mitochondrial endosymbiosis in eukaryotes." <u>Proc Natl Acad Sci U S A</u> **93**(25): 14614-14617.

Gianguzza, M. and G. Dolcemascolo (1980). "Morphological and cytochemical investigations on the formation of the test during the embryonic development of Ciona intestinalis." <u>Acta Embryol Morphol Exp</u> **1**(3): 225-239.

Gladyshev, E. A., M. Meselson, et al. (2008). "Massive horizontal gene transfer in bdelloid rotifers." <u>Science</u> **320**(5880): 1210-1213.

Gogarten, J. P., W. F. Doolittle, et al. (2002). "Prokaryotic evolution in light of gene transfer." <u>Mol Biol Evol</u> **19**(12): 2226-2238.

Gophna, U., R. L. Charlebois, et al. (2004). "Have archaeal genes contributed to bacterial virulence?" <u>Trends Microbiol</u> **12**(5): 213-219.

Gould, S. B., R. F. Waller, et al. (2008). "Plastid evolution." <u>Annu Rev Plant Biol</u> **59**: 491-517.

Gray, M. W., B. F. Lang, et al. (2004). "Mitochondria of protists." <u>Annu Rev Genet</u> **38**: 477-524.

Groussin, M., B. Boussau, et al. (2016). "Gene Acquisitions from Bacteria at the Origins of Major Archaeal Clades Are Vastly Overestimated." <u>Mol Biol Evol</u> **33**(2): 305-310.

Haegeman, A., J. T. Jones, et al. (2011). "Horizontal gene transfer in nematodes: a catalyst for plant parasitism?" <u>Mol Plant Microbe Interact</u> **24**(8): 879-887.

Hall, C., S. Brachat, et al. (2005). "Contribution of horizontal gene transfer to the evolution of Saccharomyces cerevisiae." <u>Eukaryot Cell</u> **4**(6): 1102-1115.

Hall, C. and F. S. Dietrich (2007). "The reacquisition of biotin prototrophy in Saccharomyces cerevisiae involved horizontal gene transfer, gene duplication and gene clustering." <u>Genetics</u> **177**(4): 2293-2307.

Hengherr, S., M. R. Worland, et al. (2009). "High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition." <u>Physiol Biochem Zool</u> **82**(6): 749-755.

Hilario, E. and J. P. Gogarten (1993). "Horizontal transfer of ATPase genes--the tree of life becomes a net of life." <u>Biosystems</u> **31**(2-3): 111-119.

Hirose, E., S. Kimura, et al. (1999). "Tunic morphology and cellulosic components of pyrosomas, doliolids, and salps (thaliacea, urochordata)." <u>Biol Bull</u> **196**(1): 113-120.

Horikawa, D. D., J. Cumbers, et al. (2013). "Analysis of DNA repair and protection in the Tardigrade Ramazzottius varieornatus and Hypsibius dujardini after exposure to UVC radiation." <u>PLoS One</u> **8**(6): e64793.

Huang, J. (2013). "Horizontal gene transfer in eukaryotes: the weak-link model." <u>Bioessays</u> **35**(10): 868-875.

Huang, J. and J. P. Gogarten (2006). "Ancient horizontal gene transfer can benefit phylogenetic reconstruction." <u>Trends Genet</u> **22**(7): 361-366.

Huang, J., N. Mullapudi, et al. (2004). "Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in Cryptosporidium parvum." <u>Genome Biol</u> **5**(11): R88.

Inderbitzin, P., J. Harkness, et al. (2005). "Lateral transfer of mating system in Stemphylium." <u>Proc Natl Acad Sci U S A</u> **102**(32): 11390-11395.

Jackson, D. J., L. Macis, et al. (2011). "A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy." <u>BMC Evol Biol</u> **11**: 238.

Jonsson, K. I., M. Harms-Ringdahl, et al. (2005). "Radiation tolerance in the eutardigrade Richtersius coronifer." <u>Int J Radiat Biol</u> **81**(9): 649-656.

Jonsson, K. I., E. Rabbow, et al. (2008). "Tardigrades survive exposure to space in low Earth orbit." <u>Curr Biol</u> **18**(17): R729-R731.

Katz, L. A. (2015). "Recent events dominate interdomain lateral gene transfers between prokaryotes and eukaryotes and, with the exception of

endosymbiotic gene transfers, few ancient transfer events persist." <u>Philos</u>

<u>Trans R Soc Lond B Biol Sci</u> **370**(1678).

Keeling, P. J. (2004). "Diversity and evolutionary history of plastids and their hosts."

<u>Am J Bot</u> **91**(10): 1481-1493.

Keeling, P. J. and W. F. Doolittle (1997). "Evidence that eukaryotic triosephosphate

isomerase is of alpha-proteobacterial origin." <u>Proc Natl Acad Sci U S A</u> **94**(4):

1270-1275.

Keeling, P. J. and G. I. McFadden (1998). "Origins of microsporidia." <u>Trends</u>

<u>Microbiol</u> **6**(1): 19-23.

Keeling, P. J. and J. D. Palmer (2008). "Horizontal gene transfer in eukaryotic

evolution." <u>Nat Rev Genet</u> **9**(8): 605-618.

Kimura, S., C. Ohshima, et al. (2001). "Cellulose in the house of the appendicularian

Oikopleura rufescens." <u>Protoplasma</u> **216**(1-2): 71-74.

Kishore, S. P., J. W. Stiller, et al. (2013). "Horizontal gene transfer of epigenetic

machinery and evolution of parasitism in the malaria parasite Plasmodium

falciparum and other apicomplexans." <u>BMC Evol Biol</u> **13**: 37.

Kondo, N., N. Nikoh, et al. (2002). "Genome fragment of Wolbachia endosymbiont

transferred to X chromosome of host insect." <u>Proc Natl Acad Sci U S A</u> **99**(22):

14280-14285.

Koonin, E. V., K. S. Makarova, et al. (2001). "Horizontal gene transfer in prokaryotes:

quantification and classification." <u>Annu Rev Microbiol</u> **55**: 709-742.

Koskiniemi, S., S. Sun, et al. (2012). "Selection-driven gene loss in bacteria." PLoS Genet **8**(6): e1002787.

Koutsovoulos, G., S. Kumar, et al. (2016). "No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini." Proc Natl Acad Sci U S A **113**(18): 5053-5058.

Ku, C., S. Nelson-Sathi, et al. (2015). "Endosymbiotic origin and differential loss of eukaryotic genes." Nature **524**(7566): 427-432.

Loftus, B., I. Anderson, et al. (2005). "The genome of the protist parasite Entamoeba histolytica." Nature **433**(7028): 865-868.

Matthysse, A. G., K. Deschet, et al. (2004). "A functional cellulose synthase from ascidian epidermis." Proc Natl Acad Sci U S A **101**(4): 986-991.

Mayer, W. E., L. N. Schuster, et al. (2011). "Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover." BMC Evol Biol **11**: 13.

Milani, N. A., D. P. Lawrence, et al. (2012). "Origin of pisatin demethylase (PDA) in the genus Fusarium." Fungal Genet Biol **49**(11): 933-942.

Nakashima, K., L. Yamada, et al. (2004). "The evolutionary origin of animal cellulose synthase." Dev Genes Evol **214**(2): 81-88.

Nelson-Sathi, S., T. Dagan, et al. (2012). "Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea." Proc Natl Acad Sci U S A **109**(50): 20537-20542.

Nelson-Sathi, S., F. L. Sousa, et al. (2015). "Origins of major archaeal clades correspond to gene acquisitions from bacteria." Nature **517**(7532): 77-80.

Ni, T., J. Yue, et al. (2012). "Ancient gene transfer from algae to animals: mechanisms and evolutionary significance." BMC Evol Biol **12**: 83.

Nikoh, N., K. Tanaka, et al. (2008). "Wolbachia genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes." Genome Res **18**(2): 272-280.

Nowack, E. C., M. Melkonian, et al. (2008). "Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes." Curr Biol **18**(6): 410-418.

Ochman, H., J. G. Lawrence, et al. (2000). "Lateral gene transfer and the nature of bacterial innovation." Nature **405**(6784): 299-304.

Paganini, J., A. Campan-Fournier, et al. (2012). "Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes." PLoS One **7**(11): e50875.

Palmer, K. L., V. N. Kos, et al. (2010). "Horizontal gene transfer and the genomics of enterococcal antibiotic resistance." Curr Opin Microbiol **13**(5): 632-639.

Paoletti, M., K. W. Buck, et al. (2006). "Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote Ophiostoma novo-ulmi." Mol Ecol **15**(1): 249-262.

Pelletreau, K. N., D. Bhattacharya, et al. (2011). "Sea slug kleptoplasty and plastid maintenance in a metazoan." Plant Physiol **155**(4): 1561-1565.

Pierce, S. K., X. Fang, et al. (2012). "Transcriptomic evidence for the expression of horizontally transferred algal nuclear genes in the photosynthetic sea slug, Elysia chlorotica." Mol Biol Evol **29**(6): 1545-1556.

Qiu, H., G. Cai, et al. (2016). "Extensive horizontal gene transfers between plant pathogenic fungi." BMC Biol **14**(1): 41.

Ramløv, H. and P. Westh (2001). "Cryptobiosis in the eutardigrade Adorybiotus (Richtersius) coronifer: tolerance to alcohols, temperature and de novo protein synthesis." Zoologischer Anzeiger-A Journal of Comparative Zoology **240**(3): 517-523.

Raymond, J. (2009). "The role of horizontal gene transfer in photosynthesis, oxygen production, and oxygen tolerance." Methods Mol Biol **532**: 323-338.

Rest, J. S. and D. P. Mindell (2003). "Retroids in archaea: phylogeny and lateral origins." Mol Biol Evol **20**(7): 1134-1142.

Reyes-Prieto, A., A. Moustafa, et al. (2008). "Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic." Curr Biol **18**(13): 956-962.

Ricard, G., N. R. McEwan, et al. (2006). "Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment." BMC Genomics **7**: 22.

Richards, V. P., P. Lang, et al. (2011). "Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae." Infect Genet Evol **11**(6): 1263-1275.

Roger, A. J., C. G. Clark, et al. (1996). "A possible mitochondrial gene in the early-branching amitochondriate protist Trichomonas vaginalis." Proc Natl Acad Sci U S A **93**(25): 14618-14622.

Rumpho, M. E., J. M. Worful, et al. (2008). "Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica." Proc Natl Acad Sci U S A **105**(46): 17867-17871.

Sagane, Y., K. Zech, et al. (2010). "Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans." Development **137**(9): 1483-1492.

Sasakura, Y., K. Nakashima, et al. (2005). "Transposon-mediated insertional mutagenesis revealed the functions of animal cellulose synthase in the ascidian Ciona intestinalis." Proc Natl Acad Sci U S A **102**(42): 15134-15139.

Satoh, N. (1994). "Developmental Biology of AscidiansCambridge Univ." Press, Cambridge.

Scholl, E. H., J. L. Thorne, et al. (2003). "Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach." Genome Biol **4**(6): R39.

Schonknecht, G., W. H. Chen, et al. (2013). "Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote." Science **339**(6124): 1207-1210.

Schuster, L. N. and R. J. Sommer (2012). "Expressional and functional variation of horizontally acquired cellulases in the nematode Pristionchus pacificus." Gene **506**(2): 274-282.

Slot, J. C. and D. S. Hibbett (2007). "Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study." PLoS One **2**(10): e1097.

Small, K. S., M. Brudno, et al. (2007). "A haplome alignment and reference sequence of the highly polymorphic Ciona savignyi genome." Genome Biol **8**(3): R41.

Snel, B., P. Bork, et al. (1999). "Genome phylogeny based on gene content." Nat Genet **21**(1): 108-110.

Soucy, S. M., J. Huang, et al. (2015). "Horizontal gene transfer: building the web of life." Nat Rev Genet **16**(8): 472-482.

Starcevic, A., S. Akthar, et al. (2008). "Enzymes of the shikimic acid pathway encoded in the genome of a basal metazoan, Nematostella vectensis, have microbial origins." Proc Natl Acad Sci U S A **105**(7): 2533-2537.

Sun, G., Z. Yang, et al. (2010). "Algal genes in the closest relatives of animals." Mol Biol Evol **27**(12): 2879-2889.

Syvanen, M. (2012). "Evolutionary implications of horizontal gene transfer." Annu Rev Genet **46**: 341-358.

Tucker, R. P. (2013). "Horizontal gene transfer in choanoflagellates." J Exp Zool B Mol Dev Evol **320**(1): 1-9.

Tyler, B. M., S. Tripathy, et al. (2006). "Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis." Science **313**(5791): 1261-1266.

Viterbo, A., U. Landau, et al. (2010). "Characterization of ACC deaminase from the biocontrol and plant growth-promoting agent Trichoderma asperellum T203." FEMS Microbiol Lett **305**(1): 42-48.

Vogan, A. A. and P. G. Higgs (2011). "The advantages and disadvantages of horizontal gene transfer and the emergence of the first species." Biol Direct **6**: 1.

Wagele, H., O. Deusch, et al. (2011). "Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs Elysia timida and Plakobranchus ocellatus does not entail lateral transfer of algal nuclear genes." Mol Biol Evol **28**(1): 699-706.

Watkins, R. F. and M. W. Gray (2006). "The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa." J Mol Evol **63**(6): 801-814.

Welnicz, W., M. A. Grohme, et al. (2011). "Anhydrobiosis in tardigrades--the last decade." J Insect Physiol **57**(5): 577-583.

Wisecaver, J. H., J. C. Slot, et al. (2014). "The evolution of fungal metabolic pathways." PLoS Genet **10**(12): e1004816.

Woese, C. R. (2002). "On the evolution of cells." Proc Natl Acad Sci U S A **99**(13): 8742-8747.

Wolf, Y. I. and E. V. Koonin (2013). "Genome reduction as the dominant mode of evolution." Bioessays **35**(9): 829-837.

Yao, M., T. Ose, et al. (2000). "Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from Hansenula saturnus." J Biol Chem **275**(44): 34557-34565.

Yin, Y., J. Huang, et al. (2009). "The cellulose synthase superfamily in fully sequenced plants and algae." BMC Plant Biol **9**: 99.

Yue, J., X. Hu, et al. (2012). "Widespread impact of horizontal gene transfer on plant colonization of land." Nat Commun **3**: 1152.

Yue, J. and J. Huang (2012). "Algal genes in aplastidic eukaryotes are not necessarily derived from historical plastids." Mob Genet Elements **2**(4): 193-196.

Yue, J., G. Sun, et al. (2013). "The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate Monosiga brevicollis." BMC Genomics **14**(1): 729.

Zhang, X. P. and E. Glaser (2002). "Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone." Trends Plant Sci **7**(1): 14-21.

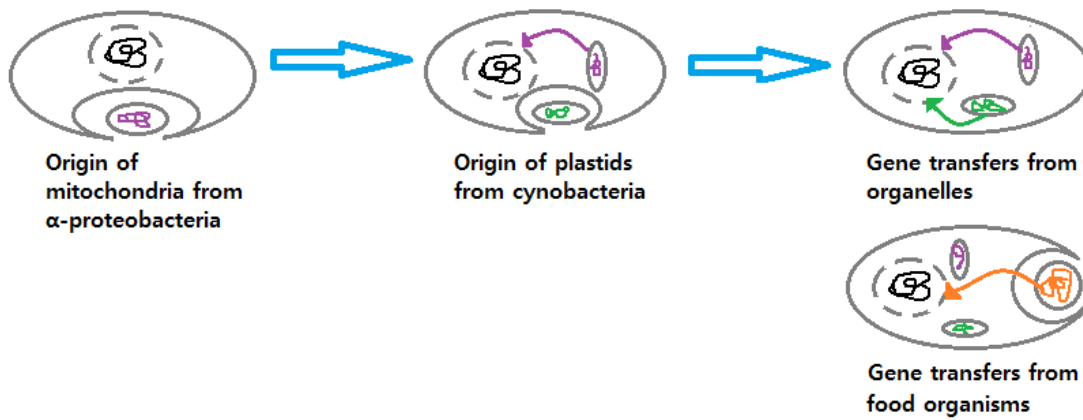**Fig. 1.** Intracellular gene transfer (IGT; or endosymbiotic gene transfer, EGT) and horizontal gene transfer (HGT). Upper panel: gene transfers from organelles to the nuclear genomes. Lower Panel: gene transfers from other species, often due to phagotrophic lifestyles. (Andersson 2005)
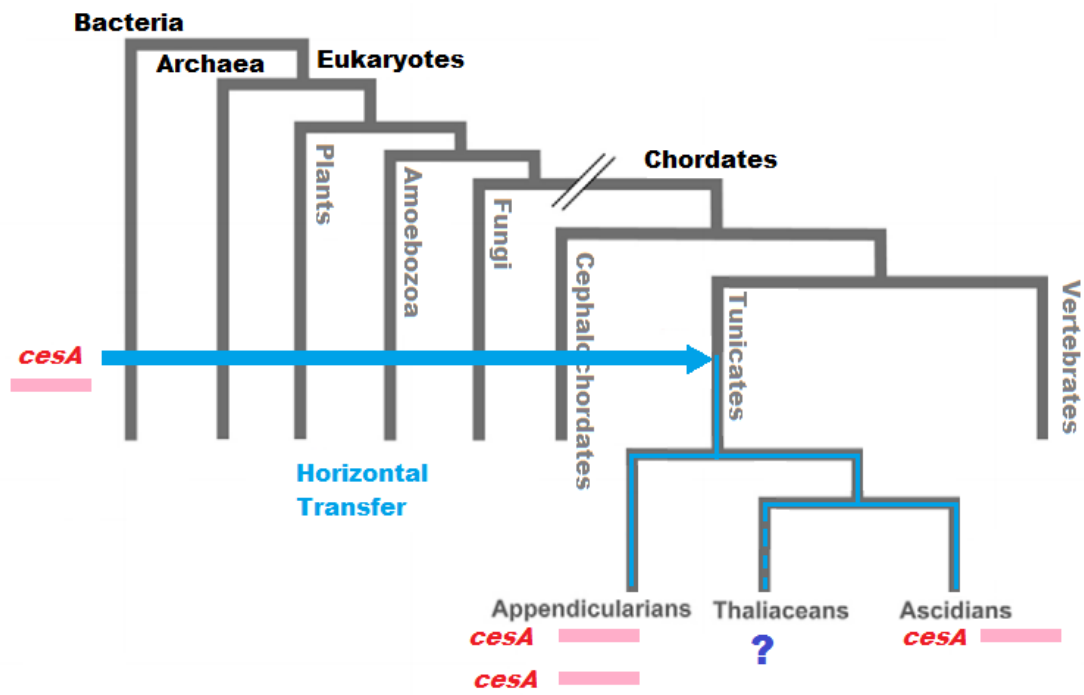
Fig. 2 Origin and evolution of cellulose synthase (CesA) genes in tunicates. Refer to (Sagane, Zech et al. 2010).

Chapter 2: Overall frequency of HGT in metazoans

Introduction

Horizontal gene transfer has long been considered as a driving force in the diversification and adaptation of prokaryotes. In a study of 539,723 genes across 181 sequenced prokaryotic genomes, on average 81±15% of prokaryotic genes were reported to be horizontally transferred at some point in their history (Dagan, Artzy-Randrup et al. 2008). HGT is also believed to play important roles in unicellular eukaryotes (Ricard, McEwan et al. 2006; Yue, Sun et al. 2013; Schonknecht, Chen et al. 2013; Bhattacharya, Price et al. 2013). For multicellular eukaryotes, however, researchers are still reluctant to accept that HGT events frequently occur. Some claim that genes of bacterial origin in multicellular eukaryotes are mainly derived from mitochondria and plastids. However, because genes in mitochondria and plastids were frequently transferred to the nucleus or lost independently during the early stage of eukaryotic evolution, transferrable genes remaining in organellar genomes are very reduced (Wu, Sun et al. 2004; Foster, Ganatra et al. 2005). The limited gene pools of mitochondrial and plastid genomes might not be able to account for large quantities of bacterial genes in eukaryotes.

Although some research has been done on the scale of HGT events in unicellular eukaryotes, like choanoflagellate *M. brevicollis* (Yue, Sun et al. 2013), rumen-dwelling ciliates (Ricard, McEwan et al. 2006), and red algae (Bhattacharya, Price et al. 2013; Schonknecht, Chen et al. 2013), only a few studies have been carried out on the scale of HGT events in multicellular eukaryotes. A recent study on

tardigrades, which survive in extreme environments, revealed that 6,600 genes, nearly one sixth of the whole genome, were horizontally transferred from different sources, like plants, bacteria, fungi and archaeans (Boothby, Tenlen et al. 2015). However, studies in other groups stated that this high HGT rate was overestimated probably due to contamination of bacteria in culture (Bemm, Weiss et al. 2016; Koutsovoulos, Kumar et al. 2016). In order to evaluate the evolutionary impacts of HGT events on recipient species, an overall scale of acquired genes is critical.

In this research, I carefully selected 16 metazoan genomes to study the percentage of horizontally derived genes. The weak-link model suggests that unicellular and early developmental stages may be entry points for foreign DNAs into multicellular eukaryotes (Huang 2013). In order to test this model, living habitat and life style can be analyzed for their effects on HGT rate. Nonparametric statistics were used to compare the HGT percentages between bilateral and non-bilateral animals, aquatic and terrestrial animals, as well as animals with sexual and asexual reproduction.

Materials and Methods

Species selection

Sixteen species were chosen for analyses in this study, namely, *Amphimedon queenslandica*, *Anopheles gambiae*, *Bombyx mori*, *Branchiostoma floridae*, *Caenorhabditis elegans*, *Ciona intestinalis*, *Ciona savignyi*, *Crassostrea gigas*, *Gasterosteus aculeatus*, *Hydra magnipapillata*, *Lottia gigantean, Mus musculus,*

Nematostella vectensis, Schistosoma japonicum, Trichoplax adhaerens, and

Xenopus tropicalis (Table 1). Those species were selected based on the "broad and

balanced" principle, to cover nearly all the major lineages of metazoans. Not only

"popular" and well-studied species such as *C. elegans* and *M. musculus* were

sampled, but also poorly studied ones, such as *B. floridae* and *H. magnipapillata*.

Well annotated genes in these species can be used to refer gene functions of their

homologues in other poorly annotated species. Some early-branching lineages,

which are usually poorly annotated, were also selected, since some ancient HGT

events affect evolution of all the descendant lineages.

Among all the sampled species, *C. intestinalis* and *C. savignyi* are two species of

the same genus. I chose them in order to find out whether recent HGT happens at

the genus level. I also selected *S. japonicum* in order to investigate whether a

parasitic life style affected the HGT rate.

Data sources and local BLAST

Genome sequences of most species were downloaded from Ensembl

(ensembl.org), Joint Genome Institute (JGI, genome.jgi.doe.gov) or Metazome v3.0

(www.metazome.net). Protein sequences of *C. gigas* were downloaded from GigaDB

database (gigadb.org) and *S. japonicum* from Chinese National Human Genome

Center at Shanghai (http://lifecenter.sgst.cn) (Table 1).

Local BLAST was carried out against a database consisting of all sequences

from NCBI non-redundant (*nr*) protein database as well as sequences translated

from the EST (Expressed sequence tags) sequences of the NCBI dbEST database

and the Taxonomically Broad EST Database (TBestDB) (O'Brien, Koski et al. 2007).

EST assembly and translation were carried out using the protocol described

previously (Ni, Yue et al. 2012). Specifically, ESTs were assembled using CAP3

(Huang and Madan 1999) and the resulting consensus sequences were translated in

all six frames using transeq of the EMBOSS package (Rice, Longden et al. 2000).

The formatdb command of the BLAST package was used to create the database.

Genome screening by AlienG

BLAST results were used as input for AlienG to screen for potential HGT-derived

genes in each genome. AlienG is based on the assumption that sequence similarity

is correlated to sequence relatedness. From the BLAST results, bit scores of the

query sequence with homologs from distantly related taxa is defined as Sd, while bit

scores with closely related taxa are defined as Sc. The alien index is defined as the

ratio of the highest bit score of distantly related species (Sd) over the highest bit

score of closely related species (Sc). A higher Sd/Sc (alien index) means that the

query is more similar to sequences from distantly related taxa than those from

closely related ones, thus implies a higher probability that the focal gene is

horizontally derived from those distantly related species. For example, alcohol

dehydrogenase (Query ID: Triad9800) of *T. adhaerens* has highest bit score of 479

and 219 in distantly related species (*Candidatus Nitrosoarchaeum limnia SFB1*) and

in closely related species (*Lingula anatina*), respectively. Alien index is 2.19

according to our definition (479/219). I set the alien index threshold as 1.5 in my

research, so alcohol dehydrogenase falls into the category of HGT candidates.

Previous studies comparing AlienG with other existing programs for HGT detection, such as PhyloGenie and Darkhorse, suggested that AlienG is computationally efficient and can identify more HGT-derived genes with lower numbers of false positive candidates if a proper Sd/Sc cutoff is chosen (Tian, Sun et al. 2011). The Sd/Sc ratio was set to 1.5 in this study, following the criterion used in previous research (Yue, Sun et al. 2013).

In AlienG, Group1 is defined as putative donor group, which is usually distantly related to the query species, while Group2 contains species that are closely related to query species. Some protein sequences can only be found in query species and distantly related species, but not in other groups, so I call them "only-in-group1" sequences. "only-in-group1" sequences could be due to relative recent HGT events, so they may not be found in other groups. It could also be explained by massive gene loss or extremely fast gene evolution in all other species. However, the latter explanation is not parsimonious most of the time. The probability is extremely low since this scenario requires too many losses or rapid sequence changes in too many lineages. Another explanation would be contamination of sequences from distantly related species, especially bacteria and algae. Hence "only-in-group1" data need to be examined carefully.

Statistical analyses

In order to figure out what factors may affect the HGT levels in different species, a contingency table was built. One of the two variables is HGT level. Since proportions, percentages or means cannot be put into a contingency table, I

separated species of different HGT rates into two groups: high HGT group and low

HGT group. Using median value of HGT rates in all 16 species (1.5%) as cutoff point,

species with an HGT rate higher or equal to 1.5% are considered as high HGT group,

otherwise as low HGT group. The other variable is primary food source, living habitat,

taxonomy ranks or reproduction types. If the primary food source of a species is

bacteria or protists, I marked it as "Yes"; otherwise "No". For living habitat, if a

species is aquatic, it is labeled as "Yes". For lineage group, if a species is bilateral, a

"Yes" tag is put on it. Species can be divided into sexual and asexual reproduction.

However, some species can conduct both sexual and asexual reproduction at

different stages of their life cycle. For example, *S. japonicum* undergoes asexual

reproduction when it is in the tissue of snails, but will mate and lay eggs when it is in

the circulation of its primate hosts. Hence, only species that undergo sexual

reproduction throughout their life cycle are considered as sexual in this study.

After two variables were selected, a two-tailed Fisher's exact test was

conducted to calculate a *p* value. If *p* value is less than 0.05, the association

between two variables was considered to be statistically significant, which means

that two variables are related to each other, with a probability of error less than 5%.

N50 length means the length of contig or scaffold for which half of the

assembled sequences are longer than that length. Both contig and scaffold N50

lengths are recorded and analyzed in this study. Some genomes followed protocols

of Sanger sequencing, so N50 lengths are either not available or too large to be

meaningful. For *H. magnipapillata*, different assembly methods give distinct N50

lengths (Chapman, Kirkness et al. 2010), so arithmetic means are used for analyses. Linear regression analyses between N50 lengths and HGT rates were conducted based on available data.

Results

HGT events are widespread in Metazoans

Since 16 metazoan species have been examined in this study, it is nearly impossible to conduct detailed analyses for each protein sequence. The candidate HGT genes after AlienG filtration were used to calculate scale of HGT events, instead of HGT genes validated by some time-consuming methods such as phylogenetic or shared domain analyses. Although using HGT candidates to estimate HGT rate is not very accurate, it saves much time and makes this work possible.

The results of the AlienG analyses are listed in Table 2. Latin names of 16 selected species are listed in the first column. The total number of protein-coding sequences in each genome is also listed. Among all 16 genomes, *B. floridae* has the largest genome, with more than 50,000 protein-coding sequences, while *T. adhaerens* has the smallest genome, with only 11,520 protein-coding sequences. After AlienG analysis, number of candidate HGT sequences and only-in-group1 sequences were reported. The percentage of candidate HGT sequences over protein-coding sequences in the whole genome was recorded as the HGT rate.

Table 2 shows that HGT happens in all investigated metazoan species. It shows that HGT events are likely not rare in metazoans, but happen ubiquitously. Among 16 sampled species, HGT rates range from 0.47% (*S. japonicum*) to 4.08 % (*N. vectensis*). In order to figure out what factors affect this ten-fold difference in HGT rate, correlations between HGT rates and living habitat, food sources, parasitism, life cycle and data quality were investigated.

Living habitat and food source affect HGT rate

According to the "you are what you eat" hypothesis of W. Ford Doolittle, bacterial and algal genes in eukaryotic genomes may be acquired through feeding activities (Doolittle 1998). In order to test whether food sources have affected the HGT rate, a 2X2 contingency table was used to calculate a two-tailed *p* value with Fisher's exact test. One of the two variables is HGT level (high or low), and the other one is food source. Since aquatic metazoans have closer contact with bacteria and protists than terrestrial ones, I hypothesized that aquatic animals are prone to obtain foreign genes, thus having a higher HGT rate. Hence a Fisher's test was also conducted between HGT level and living habitat (aquatic/terrestrial). Percentages of HGT events in 16 sampled species range from 0.47% (*S. japonicum*) to 4.08 % (*N. vectensis*).

Xenopus tropicalis is amphibian, so it is labeled as undetermined and excluded from aquatic/terrestrial analysis. Among all 16 metazoan species sampled, ten are aquatic and five are terrestrial metazoans. Eight out of ten aquatic metazoans have high HGT level, while all five terrestrial metazoans have low HGT level. Two-tailed

Fisher's test gives a *p* value of 0.007, which means that HGT rates in aquatic metazoans are significantly higher than those in terrestrial metazoans.

Similarly, food source analysis was conducted using two-tailed Fisher's test. Metazoans whose primary food sources are bacteria and/or protists were labeled "yes", otherwise "no" (Table 3). Statistical analysis results indicate that food sources significantly affect HGT rates in metazoans ($p = 0.0256$).

Early-branching metazoans may not have higher HGT rate

There are three non-bilateral metazoans with relatively high HGT rates: *A. queenslandica* (1.69%)*, T. adhaerens* (1.88%) *and N. vectensis* (4.08%). I hypothesize that early-branching metazoans may have higher HGT rates than bilateral metazoans, as most of them have aquatic living habitats, sequestered germline cells and feed primarily on bacteria and algae, thus have a higher chance of contacting foreign genes. However, when we conducted two-tailed Fisher's test, the result is insignificant ($p = 0.5692$). Two explanations may account for this lack of significance. First, although *H. magnipapillata* falls into "low HGT" category, in fact it has a relatively high percentage of HGT (1.32%), which is close to 1.5%. Second, in bilateral metazoans, species with high HGT rates are usually aquatic metazoans, which implies that living habitat, instead of lineage group, may be the main cause of the acquisition of foreign genes.

In 12 bilateral metazoans sampled in my analyses, six are aquatic, five are terrestrial and one is amphibian. A Fisher's test between living habitat and HGT level gave a *p* value 0.0152. This indicates that habitats (aquatic/terrestrial) are more

likely to be a deciding factor concerning whether a species is prone to foreign genes.

The finding that non-bilateral metazoans have higher HGT rates is probably due to

the fact that early-branching metazoans are usually aquatic. Therefore, early-

branching metazoans may not have a higher HGT rate, but aquatic metazoans do.

Asexual reproduction does not determine rates of HGT

Sequestered germline cells in metazoans are often believed to be an

insurmountable barrier (Jensen, Grant et al. 2016), which is one of the reasons why

many researchers are skeptical of the occurrence of HGT in animals. Hence, I split

16 species into "sexual" and "asexual" groups. Sexual group means the species only

conducts sexual reproduction in its life cycle, while asexual group means asexual

reproduction happens at some stage of its life cycle. Again Fisher's test gives a *p*

value 0.5692, which implies that the barrier of germline cells in an animal may not be

insurmountable, especially in aquatic metazoans.

Life styles and HGT rates

There is only one parasitic species in sampled species, so I cannot conduct

statistical analysis to test whether parasitism affects HGT level. *Schistosoma*

japonicum has only 0.47% of HGT rate, the lowest among the 16 sampled species.

At first glance, it conflicts with massive acquisition of foreign genes from bacteria and

fungi in plant-parasitic nematodes (Scholl, Thorne et al. 2003; Jones, Furlanetto et al.

2005; Mitreva, Smant et al. 2009; Danchin and Rosso 2012). However, a previous

report of HGT rates in different plant-parasitic nematode species claims that

percentages of HGT candidates range from 0% to 0.667% (Scholl, Thorne et al.

2003), which is very close to 0.47% in my study. However, another study on root-knot parasitic nematodes *Meloidogyne incognita* and *Meloidogyne hapla* shows that up to 3.34% of genes are horizontally obtained from non-metazoan origins (Paganini, Campan-Fournier et al. 2012).

Lower HGT rates in *S. japonicum* could be caused by two factors. First, a parasitic life cycle allows parasites to occupy specific niches without being exposed to high selective pressure from the environments, so superfluous genes tend to mutate and be removed from the genome (Mira, Ochman et al. 2001; Wolf and Koonin 2013). This phenomenon is called deletional bias (Kuo and Ochman 2009). Even if foreign genes have been acquired by *S. japonicum*, they are very likely to mutate and be deleted from the genome of *S. japonicum*. Second, except for being a free-swimming larva for one or two days, *S. japonicum* spends most of its life cycle in the tissue of snails or the circulation of its primate hosts, so it has an extremely low chance to come into contact with bacteria, fungi or protists.

Discussion

More and more cases have been reported of HGT events in different metazoans. However, genome-wide studies are still rare on the scale of HGT events. What is even rarer are comparisons among the scales of HGT events in different species and the factors that cause differences in HGT rates. When this dissertation was prepared, a similar study was published by Crisp A. et al. (Crisp, Boschetti et al. 2015). Compared with my research, this research examined more genomes,

including ten primates, twelve flies and four nematodes for a detailed analysis and fourteen vertebrate genomes for a simplified analysis. All of these genome and transcriptome data are of high quality. Crisp et al. also studied the origin of horizontally transferred genes and their functionalities. Finally the authors concluded that HGT was not just a trait belonging to a narrow group of animals, but instead a general feature of all chordate genomes.

My research is similar to, but not identical with that of Crisp et al. First, not only is a general picture of HGT scales in different metazoan lineages depicted, but also factors that may cause differences in HGT rates in various metazoan species are analyzed and discussed. Association between HGT rates and primary food sources, living habitat (aquatic or terrestrial), taxonomy ranks (bilateral or non-bilateral) or reproduction type has been analyzed based on non-parametric statistics. Second, sample selection is not restricted in model animals (e.g., flies and nematodes) in my research. Four early-branching animals (*A. queenslandica*, *T. adhaerens, H. magnipapillata* and *N. vectensis*) were also selected. Our goals were to address the question of whether a gene was horizontally transferred to the last common ancestor of all metazoans (see discussion of PAM genes in *T. adhaerens*), or to a recently evolved clade of metazoans and increased the adaptation of recipient species to a new niche (see discussion of cellulose in tunicates or PBLD genes in *T. adhaerens*). It would be difficult to accomplish such research goals if genomes of early-branching animals were excluded. Two species of tunicates (*C. intestinalis* and *C. savignyi*) were chosen to study the effects of recent HGT events.

The genome of parasite *S. japonicum* was selected to determine the effect of

parasitism on HGT rates in metazoans.

Most importantly, different parameters were used in these two studies. In

Crisp's article, an *h* index was calculated as the difference between the bitscores of

the best non-metazoan and the best metazoan hits. In my research, an alien index is

calculated as the ratio of the highest bitscore of Group1 (distantly related species)

over that of Group2 (close related species). For conserved genes, *h* index may

cause high false positive results. For example, if a sequence has the highest bit

score 1,000 for metazoan hit and 1,100 for non-metazoan hit, it would be considered

as a HGT candidate, since *h* index is 100 (1,100-1,000) and meets the most

stringent criteria. However, with alien index, this sequence will not be grouped as a

HGT candidate, since the ratio (1,100/1,000) is only 1.1, less than our cutoff value

1.5.

I hypothesize that early-branching metazoans may have higher HGT rates

than bilateral metazoans. However, according to the results of my analyses,

metazoan lineages may not be a decisive factor affecting HGT rates (p = 0.5692).

Aquatic bilateral metazoans have relatively high HGT rates. Primary food sources

and living habitat show a strong correlation with HGT rates (p = 0.007 and p =

0.0256, respectively). Indeed, active feeding of bacteria and algae are found to

account for acquisition of foreign genes in *Monosiga* (Yue, Sun et al. 2013),

Entamoeba (Grant and Katz 2014) and other protists (Doolittle 1998), which is

consistent with the mechanism of "you are what you eat" proposed by W. Ford Doolittle (Doolittle 1998).

However, feeding activities cannot explain all the HGT events; especially those in plants (no feeding activity) or terrestrial metazoans (do not feed on bacteria or algae). The "weak-link model" (Huang 2013) suggests that foreign genes may be acquired by a recipient at weakly protected stages of its life cycle. For bilateral metazoans, germline cells are always sequestered by other cells in reproductive systems, leaving rare chance to encounter foreign genes. However, if eggs are laid and hatched in water, chances of foreign DNA entry are much higher than on land. This might be the reason why aquatic bilateral metazoans still have high HGT rates in my study. An interesting case is *X. tropicalis* in this study, which is amphibian and usually lays eggs in moist places. It has a moderate HGT rate (0.96%), which is between that of aquatic and terrestrial metazoans.

In my study, reproduction type is also taken into consideration of its effects on HGT rate. Coincidentally, all bilateral metazoans undergo sexual reproduction and non-bilateral metazoans undergo asexual reproduction in the 16 sampled species. Our Fisher's test produced the same p value 0.5692 in both reproduction test and taxonomy group test. In nature, there might be some exceptions, but the percentage is not high enough to reject the conclusion made.

Two species of the same genus were chosen in this study, namely, *C. intestinalis* and *C. savignyi*. Although HGT rates are very different in these two species (2.35% and 1.56%), the numbers of candidate HGT genes are very close

(329 and 314). Differences in HGT rates are solely due to genome sizes, since the

number of proteins in *C. savignyi* is 44% more than that in *C. intestinalis*. One

possible explanation would be different annotation quality of these two genomes.

Differences in candidate numbers could be due to recent HGT events in individual

species after they diverged from their most recent common ancestor.

The effects of parasitism on HGT rates are also considered in this study.

Since there is only one parasitic species being selected, no statistical analysis can

be conducted. *Schistosoma japonicum* has the lowest HGT rate (0.47%), but

whether the same low HGT rate occurs in other parasitic species needs to be further

investigated. There have been a large number of reports of HGT events in metazoan

and plant parasites. For example, the ferrochelatase gene, which is essential for

heme biosynthesis but is absent in most nematodes, was acquired from bacteria by

some human-parasitic nematodes (Wu, Novelli et al. 2013). A plant-parasitic

nematode, *Globodera pallida*, also obtained Glycosyl Hydrolase Family 32 (GH32)

sequences from bacteria, which enables this species to metabolize host-derived

sucrose (Danchin, Guzeeva et al. 2016). Sun has reported transfer of a strictosidine

synthase-like (SSL) gene from a host plant in Brassicaceae to its root parasitic plant

Orobanche aegyptiaca and shoot parasitic plant *Cuscuta australis* (Zhang, Qi et al.

2014).

There are also some reports of HGT events in schistosomes, including *S.

japonicum* and *S. mansoni*, from mouse (Imase, Kobayashi et al. 2000; Imase,

Kobayashi et al. 2001; Imase, Ohmae et al. 2004), fish (Leaver 2001), anthropoids

(Okada, Imase et al. 2006; Zhao, Li et al. 2009), human (Yu, Li et al. 2008) and other mammals (Hu, Yan et al. 2003; Inal 2005). However, after scrutinizing the genes mentioned in 13 published studies, none of them could be substantiated in a recent study using molecular, bioinformatic and phylogenetic approaches (Wijayawardena, Minchella et al. 2015). This implies that HGT events may not be as widespread in *S. japonicum* as in other parasitic metazoans, probably because *S. japonicum* spends most of its life in a closed environment (circulatory system).

Quality of the genome annotation is extremely important for the prediction of HGT events, since contaminated sequences cannot be ruled out in AlienG filtration. In order to ensure that high rates of HGT in some species were not due to poor quality of genome data, linear regression between N50 lengths and HGT rates was analyzed. Half of 16 sampled genomes (8 out of 16) had N50 lengths available and are listed in Table 3. Correlation coefficients are $-3*10^{-4}$ and -10^{-7} for contig N50 and scaffold N50 respectively, indicating that the prediction of HGT events is not significantly affected by the genome quality in this study.

However, this study has some pitfalls. Since phylogenetic tree analyses are not possible for all 16 selected genomes, I use HGT candidates for analyses in this chapter. Plenty of predicted candidates may be false positive, so detailed analyses should be conducted in future research. In the following two chapters, HGT candidate genes of *C. intestinalis* and *N. vectensis* are scrutinized in details.

Overall, in spite of some pitfalls listed above, this study is not only one of the first studies on the scale of HGT in all metazoans, but also the first to investigate

factors that may affect HGT rates based on rigorous statistical analyses. Detailed

phylogenetic validation of transferred genes has also been conducted and will be

discussed in the following two chapters.

References

Bemm, F., C. L. Weiss, et al. (2016). "Genome of a tardigrade: Horizontal gene transfer or bacterial contamination?" Proc Natl Acad Sci U S A **113**(22): E3054-3056.

Bhattacharya, D., D. C. Price, et al. (2013). "Genome of the red alga Porphyridium purpureum." Nat Commun **4**: 1941.

Boothby, T. C., J. R. Tenlen, et al. (2015). "Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade." Proc Natl Acad Sci U S A **112**(52): 15976-15981.

Chapman, J. A., E. F. Kirkness, et al. (2010). "The dynamic genome of Hydra." Nature **464**(7288): 592-596.

Crisp, A., C. Boschetti, et al. (2015). "Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes." Genome Biol **16**: 50.

Dagan, T., Y. Artzy-Randrup, et al. (2008). "Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution." Proc Natl Acad Sci U S A **105**(29): 10039-10044.

Danchin, E. G., E. A. Guzeeva, et al. (2016). "Horizontal Gene Transfer from Bacteria Has Enabled the Plant-Parasitic Nematode Globodera pallida to Feed on Host-Derived Sucrose." Mol Biol Evol.

Danchin, E. G. and M. N. Rosso (2012). "Lateral gene transfers have polished animal genomes: lessons from nematodes." Front Cell Infect Microbiol **2**: 27.

Doolittle, W. F. (1998). "You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes." Trends Genet **14**(8): 307-311.

Foster, J., M. Ganatra, et al. (2005). "The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode." PLoS Biol **3**(4): e121.

Grant, J. R. and L. A. Katz (2014). "Phylogenomic study indicates widespread lateral gene transfer in Entamoeba and suggests a past intimate relationship with parabasalids." Genome Biol Evol **6**(9): 2350-2360.

Hu, W., Q. Yan, et al. (2003). "Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource." Nat Genet **35**(2): 139-147.

Huang, J. (2013). "Horizontal gene transfer in eukaryotes: the weak-link model." Bioessays **35**(10): 868-875.

Huang, X. and A. Madan (1999). "CAP3: A DNA sequence assembly program." Genome Res **9**(9): 868-877.

Imase, A., K. Kobayashi, et al. (2000). "Horizontal and vertical transfer of mouse endogenous retroviral DNA sequences in schistosomes." Parasitology **121 (Pt 2)**: 155-162.

Imase, A., K. Kobayashi, et al. (2001). "Horizontal and vertical transmission of mouse class I MHC sequence in Schistosoma mansoni." Parasitology **123**(Pt 2): 163-168.

Imase, A., H. Ohmae, et al. (2004). "A comparative study on mouse MHC class I

sequences detected in Schistosoma japonicum recovered from BALB/c (H-2d)

and C57BL/6 (H-2b) mice." <u>Southeast Asian J Trop Med Public Health</u> **35**(1):

10-18.

Inal, J. M. (2005). "Complement C2 receptor inhibitor trispanning: from man to

schistosome." <u>Springer Semin Immunopathol</u> **27**(3): 320-331.

Jensen, L., J. R. Grant, et al. (2016). "Assessing the effects of a sequestered

germline on interdomain lateral gene transfer in Metazoa." <u>Evolution</u> **70**(6):

1322-1333.

Jones, J. T., C. Furlanetto, et al. (2005). "Horizontal gene transfer from bacteria and

fungi as a driving force in the evolution of plant parasitism in nematodes."

<u>Nematology</u> **7**(5): 641-646.

Koutsovoulos, G., S. Kumar, et al. (2016). "No evidence for extensive horizontal

gene transfer in the genome of the tardigrade Hypsibius dujardini." <u>Proc Natl

Acad Sci U S A</u> **113**(18): 5053-5058.

Kuo, C. H. and H. Ochman (2009). "Deletional bias across the three domains of life."

<u>Genome Biol Evol</u> **1**: 145-152.

Leaver, M. J. (2001). "A family of Tc1-like transposons from the genomes of fishes

and frogs: evidence for horizontal transmission." <u>Gene</u> **271**(2): 203-214.

Mira, A., H. Ochman, et al. (2001). "Deletional bias and the evolution of bacterial

genomes." <u>Trends Genet</u> **17**(10): 589-596.

Mitreva, M., G. Smant, et al. (2009). "Role of horizontal gene transfer in the evolution of plant parasitism among nematodes." <u>Methods Mol Biol</u> **532**: 517-535.

Ni, T., J. Yue, et al. (2012). "Ancient gene transfer from algae to animals: mechanisms and evolutionary significance." <u>BMC Evol Biol</u> **12**: 83.

O'Brien, E. A., L. B. Koski, et al. (2007). "TBestDB: a taxonomically broad database of expressed sequence tags (ESTs)." <u>Nucleic Acids Res</u> **35**(Database issue): D445-451.

Okada, A., A. Imase, et al. (2006). "Heterogeneity of class I and class II MHC sequences in Schistosoma mansoni." <u>Parasitol Res</u> **99**(1): 21-27.

Paganini, J., A. Campan-Fournier, et al. (2012). "Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes." <u>PLoS One</u> **7**(11): e50875.

Ricard, G., N. R. McEwan, et al. (2006). "Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment." <u>BMC Genomics</u> **7**: 22.

Rice, P., I. Longden, et al. (2000). "EMBOSS: the European Molecular Biology Open Software Suite." <u>Trends Genet</u> **16**(6): 276-277.

Scholl, E. H., J. L. Thorne, et al. (2003). "Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach." <u>Genome Biol</u> **4**(6): R39.

Schonknecht, G., W. H. Chen, et al. (2013). "Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote." Science **339**(6124): 1207-1210.

Tian, J., G. Sun, et al. (2011). AlienG: An Effective Computational Tool for Phylogenomic Identification of Horizontally Transferred Genes. BICoB.

Wijayawardena, B. K., D. J. Minchella, et al. (2015). "Horizontal gene transfer in schistosomes: A critical assessment." Mol Biochem Parasitol **201**(1): 57-65.

Wolf, Y. I. and E. V. Koonin (2013). "Genome reduction as the dominant mode of evolution." Bioessays **35**(9): 829-837.

Wu, B., J. Novelli, et al. (2013). "Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes." Proc Natl Acad Sci U S A **110**(19): 7748-7753.

Wu, M., L. V. Sun, et al. (2004). "Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements." PLoS Biol **2**(3): E69.

Yu, F., Y. Li, et al. (2008). "Comparative genomics of human-like Schistosoma japonicum genes indicates a putative mechanism for host-parasite relationship." Genomics **91**(2): 152-157.

Yue, J., G. Sun, et al. (2013). "The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate Monosiga brevicollis." BMC Genomics **14**: 729.

Zhang, D., J. Qi, et al. (2014). "Root parasitic plant Orobanche aegyptiaca and shoot

parasitic plant Cuscuta australis obtained Brassicaceae-specific strictosidine

synthase-like genes by horizontal gene transfer." <u>BMC Plant Biol</u> **14**: 19.

Zhao, G. H., J. Li, et al. (2009). "Heterogeneity of class I and class II MHC

sequences in Schistosoma japonicum from different endemic regions in

mainland China." <u>Parasitol Res</u> **106**(1): 201-206.

Table 1. Genome information of 16 sampled species. Common name, species

name (Latin) and date source are listed.

Common name	Species name	Data source
Demosponge	*Amphimedon queenslandica*	ensembl.org
Mosquito	*Anopheles gambiae*	ensembl.org
Silkworm	*Bombyx mori*	Joint Genome Institute
Lancelet	*Branchiostoma floridae*	Joint Genome Institute
Nematode	*Caenorhabditis elegans*	Metazome v3.0
Sea squirt	*Ciona intestinalis*	Metazome v3.0
Sea squirt	*Ciona savignyi*	Metazome v3.0
Oyster	*Crassostrea gigas*	gigadb.org
Fish	*Gasterosteus aculeatus*	Joint Genome Institute

Polyp	*Hydra magnipapillata*	Metazome v3.0
Sea snail	*Lottia gigantean*	Joint Genome Institute
Mouse	*Mus musculus*	Metazome v3.0
Sea anemone	*Nematostella vectensis*	Metazome v3.0
Schistosome	*Schistosoma japonicum*	http://lifecenter.sgst.cn/schistosoma/cn/schistosomaCnIndexPage.do
Placozoan	*Trichoplax adhaerens*	Joint Genome Institute
Frog	*Xenopus tropicalis*	Joint Genome Institute

Table 2. List of 16 metazoan species names, numbers of protein sequences, candidate HGT genes, only_in_group1 genes as well as the percentage of HGT genes in the whole genomes.

Species name	No. of sequences	Candidates	only_in_group1	HGT rate(%)
Amphimedon queenslandica	29,883	505	2566	1.69
Anopheles gambiae	14,324	176	742	1.23
Bombyx mori	21,302	132	950	0.62
Branchiostoma floridae	50,817	1389	3984	2.73
Caenorhabditis elegans	27,001	154	883	0.57
Ciona intestinalis	14,002	329	688	2.35
Ciona savignyi	20,150	314	783	1.56
Crassostrea gigas	28,027	413	1739	1.47
Gasterosteus aculeatus	27,658	538	1182	1.95
Hydra magnipapillata	32,338	427	3,026	1.32
Lottia gigantean	23,851	367	1417	1.54
Mus musculus	31,289	188	476	0.60

Nematostella vectensis	27,273	1,113	3,080	4.08
Schistosoma japonicum	12,657	59	388	0.47
Trichoplax adhaerens	11,520	216	574	1.88
Xenopus tropicalis	27,916	268	827	0.96

Table 3. Factors that may affect HGT rates in 16 sampled metazoans. In the food column, "Yes" means primary food sources are bacteria or algae. In the habitat column, "Yes" indicates aquatic living habitats. Only species that undergo sexual reproduction throughout its life cycle are considered as sexual in this study.

Species name	Percentage(%)	Food	Habitat	Bilateral (Yes/No)	Reproduction	Parasitic	Contig N50 (kb)	Scalfold N50 (kb)
Amphimedon queenslandica	1.69	Yes	Yes	No	Non-sexual	No	11.20	120.00
Anopheles gambiae	1.23	No	No	Yes	Sexual	No	NA	NA
Bombyx mori	0.62	No	No	Yes	Sexual	No	NA	NA
Branchiostoma floridae	2.73	Yes	Yes	Yes	Sexual	No	25.67	38.00
Caenorhabditis elegans	0.57	Yes	No	Yes	Sexual	No	NA	NA
Ciona intestinalis	2.35	Yes	Yes	Yes	Sexual	No	33.20	190.00
Ciona savignyi	1.63	Yes	Yes	Yes	Sexual	No	NA	NA
Crassostrea gigas	1.47	Yes	Yes	Yes	Sexual	No	0.91	137.40
Gasterosteus aculeatus	1.95	Yes	Yes	Yes	Sexual	No	NA	NA
Hydra magnipapillata	1.32	Yes	Yes	No	Non-sexual	No	11.25	77.95

Lottia gigantean	1.54	Yes	Yes	Yes	Sexual	No	NA	NA
Mus musculus	0.6	No	No	Yes	Sexual	No	NA	NA
Nematostella vectensis	4.08	Yes	Yes	No	Non-sexual	No	1.98	472.00
Schistosoma japonicum	0.47	No	No	Yes	Sexual	Yes	6.12	174.00
Trichoplax adhaerens	1.88	Yes	Yes	No	Non-sexual	No	204.20	5490.00
Xenopus tropicalis	0.96	No	ND	Yes	Sexual	No	NA	NA

Chapter 3: HGT events in urochordates

Introduction

Tunicates, commonly called sea squirts, are one of the early branching clades of chordates and are the sister group of vertebrates, including humans. Three classes are included in the clade Tunicata, Thaliacea, Appendicularia and Ascidiacea. *Ciona intestinalis* and *C. savignyi* are two model organisms used for developmental biology studies, and both species belong to the class Ascidiacea. *Oikopleura dioica* of class Appendicularia is another commonly used model organism for molecular genetics studies, due to its small genome size. Full genome sequence of *C. intestinalis* shows that it is less than 1/20 of the human genome (Dehal, Satou et al. 2002). In spite of its small genome size, all the homologues of human proteins can be found in *C. intestinalis*, although most genes only have one copy (Dehal, Satou et al. 2002).

A previous study of HGT events in tunicates demonstrates that 92 genes of 14 gene families are derived from miscellaneous photosynthetic eukaryotes (Ni, Yue et al. 2012). Most horizontally transferred genes are not only found in tunicates, but also are distributed in various animals, which implies that algal genes might be acquired by the common ancestor of all animals. Horizontally acquired genes are involved in molecule transport, cellular regulation and methylation signaling, suggesting their functionalities in intercellular communication in the ancestral animal.

Celluloses and hemicelluloses are the most abundant biopolymers on earth. Biosynthesis of these two types of polysaccharides has been observed in the

extracellular matrices of a number of prokaryotes, plants and fungi. They are the

main components of plant cell walls and stimulate scientific interest due to their

potential for bioenergy production (Mizrachi, Mansfield, and Myburg 2012). Cellulose

synthase (Ces), which forms a large hexagonal structure called cellulose synthase

complexes (CSCs) (Wightman and Turner 2010), is the key enzyme for cellulose

biosynthesis. Since the first plant Ces gene was identified in cotton due to its

similarity to a bacterial Ces (Pear, Kawagoe et al. 1996), many Ces and Ces-like

(Csl) genes have been found and analyzed in *Arabidopsis* (Richmond and

Somerville 2000), rice (Hazen, Scott-Craig et al. 2002), poplar (Djerbi, Lindskog et al.

2005; Suzuki, Li et al. 2006) and moss (Roberts and Bushoven 2007). Furthermore,

detailed studies have been carried on in the phylogenies of Ces genes in diverse

plants and algae by comprehensive data mining of multiple plant genomes and

transcriptomes. Those genes could be further classified into one CesA family and ten

Csl families, which comprise the CesA/Csl superfamily (Yin, Huang et al. 2009; Yin,

Johns et al. 2014).

Early researches on Csl genes located 29 and 37 copies in *Arabidopsis* and rice

genomes, respectively, all of which belongs to the glycosyltransferase family 2 (GT2).

These Csl genes can be divided into eight main families CslA-H (Richmond and

Somerville 2000; Richmond and Somerville 2001; Hazen, Scott-Craig et al. 2002).

Besides the eight main Csl families, there are also some other Csl genes that can

only be found in a small subset of plants. The CslJ family is found in several cereal

genomes (Farrokhi, Burton et al. 2006) and several other genomes of flowering

plants (Yin, Johns et al. 2014), while the CslK family is found in six chlorophyte green algal genomes (Yin, Johns et al. 2014). CslK is phylogenetically close to CslA/C and may be the ancient ortholog before diversification into CslA and CslC (Yin, Huang et al. 2009). Various Csl genes are responsible for the synthesis of different hemicelluloses (Richmond and Somerville 2000; Lerouxel, Cavalier et al. 2006). This is known as the "CSL hypothesis" (Roberts and Gonzalez-Carranza 2008). For example, CslA encodes for mannan synthases (Dhugga, Barreiro et al. 2004; Liepman, Wilkerson et al. 2005), CslC for xyloglucan synthases (Cocuron, Lerouxel et al. 2007), CslF for mixed-linkage glucan synthases (Burton, Wilson et al. 2006) and CslH for mixed-linkage glucan synthases (Doblin, Pettolino et al. 2009). Whether other Csl families are involved in the syntheses of other hemicelluloses needs further clarification.

Yin et al. proposed a model to depict an evolutionary scenario for the CesA/Csl superfamily (Yin, Huang et al. 2009; Yin, Johns et al. 2014) (Fig. 3). In the earliest plant cells, there were two copies of the genes. One was the ancestor of extant CslA/C/K families and the other gave rise to all the other families. The former ancestor gene remained a single copy in chlorophytes (CslK) and underwent gene duplication to CslA and CslC in charophycean green algae (CGA). Gene duplication also occurs in the latter ancestor gene and results in two main clades: CesA/CslD/F family and CslB/H/E/J/G. Due to a lack of genome data, this research was carried on with a combination of genome sequences and transcriptome data. As more whole

genome data become available, it will be necessary to revisit the origin of Ces and Csl families in land plants and algae.

The only case of cellulose biosynthesis in animals is found in the tunicate (urochordate) lineage (Kimura, Ohshima et al. 2001). Biosynthesis of cellulose microfibrils was observed in all three classes of urochordates, Ascidiacea, Thaliacea and Appendicularia. In ascidians and thaliaceans, these cellulose microfibrils are incorporated into a protecting structure called tunic, while in appendicularians, they form a jelly-like "house" to help filter-feed on marine plankton. Correspondingly, Ces genes were also identified in ascidians (Dehal, Satou et al. 2002) and appendicularians (Sagane, Zech et al. 2010), but no research has been carried out in thaliaceans. While only one copy of Ces was discovered in *C. savignyi* (Dehal, Satou et al. 2002), two Ces genes exist in *Oikopleura dioica*, one acting in the pre-metamorphic stage and the other functioning in the post-metamorphic stage (Sagane, Zech et al. 2010), displaying functional specialization of Ces genes.

According to my conclusions from Chapter one, the tunicate is prone to high rate of HGT events, since it is a marine organism and feed on bacteria and algae. Preliminary data also show that potentially 2.35% of *C. intestinalis* genome is derived from HGT events, so it would be interesting to see what genes are horizontally transferred, especially from bacteria.

Although most previous studies proposed that the tunicates Ces was acquired more than 500 million years ago from a prokaryote (Dehal, Satou et al. 2002; Nakashima, Yamada et al. 2004; Nobles and Brown 2004; Sasakura, Nakashima et

al. 2005; Sagane, Zech et al. 2010), the origin of these genes required further study

in view of the availability of new genome data. Therefore, we conducted the

phylogenetic analysis of Ces and Csl proteins using updated data sets, to study the

origin of Ces and Csl proteins in plants and tunicates.

Materials and methods

Data sources

14,002 protein sequences of *C. intestinalis* were downloaded from Metazome

v3.0 (www.metazome.net/). Customized database construction, local BLAST and

AlienG analysis were carried out as described in Chapter 2.

Cellulose synthases (Ces) and cellulose synthases-like (Csl) proteins of plants

and urochordates were used as queries to search against the NCBI nr protein

sequence database (E-value < 1e-10) using blastp, in order to identify and retrieve

their homologs respectively. Specifically, we use Ces1, CslD3, CslA1, and CslC8

sequences of *Arabidopsis thaliana* as well as the Ci-CesA (*Ciona intestinalis* CesA)

sequence of *Ciona intestinalis* to search against the nr database based on previous

researches (Yue, Hu et al. 2012). The top hits in representative taxa from each blast

search result were sampled for further analysis. A total of 46 sequences were

sampled for cellulose synthases and related proteins for phylogenetic analyses in

plants and urochordates.

Phylogenetic analysis

The default parameters were used for web-based BLAST except that the maximum number of hit sequences was changed from 100 to 10,000. The number of hit sequences was increased in order to include as many lineages as possible for further analyses. Expected value threshold (E-value cutoff) was also changed from 10 to a much more stringent value 0.001.

Three phylogenetic trees were constructed: a plant Ces and Csl tree, a tunicate Ces tree, and an integrated tree including both clades. Sequences were aligned manually in MEGA6. Phylogenetic analyses were performed using maximum likelihood (ML) in PhyML 3.0 (Guindon, Dufayard et al. 2010) and Bayesian inference in MrBayes v. 3.1. For ML analysis, the appropriate model of amino acid substitution, LG model, was selected. The LG model was recently developed and generally outperforms the JTT and WAG models (Le and Gascuel 2008). Relative clade support was estimated by ML bootstrap analysis of 100 replicates of a heuristic search, with settings as above. Bayesian analysis was performed with MrBayes 3.1 using the model (GTR+I+G) suggested by MrModeltest v2.2. The settings for the Metropolis-coupled Markov chain Monte Carlo process were: three runs with four chains each were run simultaneously for $1*10^7$ generations, which were logged every 1000 generations. Convergence was considered to have been reached when the variance of split frequencies was 0.01. The first 2500 generations were discarded as the transient burn-in period. The 50%-majority-rule consensus of trees sampled in the Bayesian phylogenetic analysis was used to construct a phylogram.

Results and discussion

HGT-derived genes in C. intestinalis

Among all 14,002 annotated protein sequences in *C. intestinalis*, 329 were identified as candidates of HGT genes, which take up 2.35% of the whole genome. After analyses of taxonomic distributions, multiple sequence alignments and molecular phylogenies, 46 genes of 32 families were identified as horizontally transferred genes from green plants, bacteria or archaea (Table 4). Along with phylogenetic trees, shared indels between *C. intestinalis* and other species also indicate a common origin of those genes in many cases, which suggests horizontal acquisition from donor species to *C. intestinalis*. All 46 genes are located on large genomic scaffolds, which minimizes the probability of bacterial or algal contamination.

Origin of cellulose synthases proteins in tunicates

Most previous studies found that the tunicate Ces proteins have been horizontally derived from a prokaryote (Dehal, Satou et al. 2002; Nakashima, Yamada et al. 2004; Nobles and Brown 2004; Sagane, Zech et al. 2010). Surprisingly, in contrast to previous research, our phylogenetic results show that the cellulose synthases-like protein in tunicates is most closely related to Ces proteins from two green algae species (*Auxenochlorella protothecoides* and *Coccomyxa subellipsoidea* C-169) and one haptophyte (*Emiliania huxleyi*) (Fig.3). This clade clustered with red algae (*Chondrus crispus*) and a Stramenopiles (*Nannochloropsis gaditana*) lineage. All of these eukaryotic clades composed one monophyletic group,

while all bacteria shared another clade. The close relationship of Csl proteins

between green algae and tunicates was also confirmed by the motif composition

analysis (Fig. 4B). Csl proteins in green algae and tunicates groups contain

glycosyltransferase like family 2 (GT-2) and C-terminal glycosylhydrolase family (GH)

domains. These mosaic domain organizations of Csl genes are only found in

tunicates and two green algae in eukaryotes. Interestingly, the GT-2 and GH genes

are also contained within a single operon in some bacteria (Römling 2002; Xu,

Chater et al. 2008). This similarity suggests the possibility of multiple HGT events

from bacteria to green algae, and then from algae to tunicates. However, it is difficult

to conclude which bacteria donated this transferred gene on the basis of the existing

evidence. These kinds of HGT events from algae to animals, in association with

phagotrophic lifestyles have also been found in previous studies (Ni, Yue et al. 2012).

In addition, the red algae, haptophyte and stramenopiles also clustered with green

algae and tunicates, possibly due to secondary endosymbiosis events. The Csl

genes in red algae, haptophyte and stramenopiles only contain the GT-2 domain,

which suggests the GH domain has been lost in these species. Another question is

why previous studies did not find these closed homologous genes in eukaryotes

except for tunicates. We checked these homologous genes in NCBI, and found all of

them were submitted to NCBI protein database since 2012, which was earlier than

related studies about the origin of Ces genes in tunicates (Dehal, Satou et al. 2002;

Nakashima, Yamada et al. 2004; Nobles and Brown 2004; Sagane, Zech et al. 2010).

We speculated that there may be phylogenetically more closely related genes in green algae or other eukaryotes in these new genome datasets.

Origin of Ces and Csl proteins in plants

In plants, cellulose synthase and cellulose synthase-like proteins are divided into two clades, namely, CesA/CslB/D/E/F/G/H and CslA/C (Fig. 5). Our phylogenetic analyses indicate that CesA/CslB/D/E/F/G/H originated from cyanobacteria, and are most closed to Chroococcales (e.g. *Synechococcus* and *Gloeocapsa*) and Nostocales (e.g. *Nostoc*, *Anabaena*, *Calothrix*, and *Dolichospermum*), which is consistent with previous research (Nobles, Romanovicz et al. 2001; Nobles and Brown 2004; Yin, Huang et al. 2009; Yin, Johns et al. 2014). However, the origin of CslA/C is a little confusing based on our results. The top hits of blastp search in non-plant species contain several taxa groups, proteobacteria, Cytophaga/Flexibacter/Bacteroides (CFB), cyanobacteria, and archaea. The sequence identity and alignments show no significant difference in E-values in all of these different taxa sequences. There are two possible explanations for the origin of CslA/C in plants. One is bacterial (noncyanobacteria) origin, such as d-proteobacteria or CBF according to Fig.4. Another is cyanobacterial origin through primary endosymbiosis, which is the same as Ces and other Csl proteins. It is difficult to determine which explanation is correct according to the phylogenetic tree.

However, it should be noted that the same cyanobacteria strain *Synechococcus sp.* PCC 7502 is present in both Ces and CslA/C clades. The genome sequence of S. sp. PCC 7502 was released in a recent study of about 54 cyanobacteria genomes

(Shih, Wu et al. 2013). This strain was closely clustered with the genus

Pseudanabaena rather than the rest of the species of *Synechococcus* in the

phylogenetic analysis (Shih, Wu et al. 2013). The genus *Pseudanabaena* is also

among the first hits of homologous gene against CslA/C. Therefore, the above

results suggest that the Ces and CslA/C genes may have originated from ancient

common ancestors, such as close relatives of *Pseudanabaena-Synechococcus sp.*

PCC 7502. Shih et al. also does not reject the hypothesis that plastids emerged from

Pseudanabaena-Synechococcus sp. PCC 7502 clade (Shih, Wu et al. 2013). In

addition, our phylogenetic analysis also indicates that many species of *Nostocales*

show high identity to plant Ces proteins. Previous researches tend to agree that the

Ces/Csl genes in plants originate from Nostocales (group IV cyanobacteria) (Nobles,

Romanovicz et al. 2001; Nobles and Brown 2004), which is supported by other

hypothesis that *Nostocales* are most similar to the original endosymbiont (Martin,

Rujan et al. 2002; Deusch, Landan et al. 2008). However, there are more studies

disagreeing with this endosymbiont claim (Reyes-Prieto, Yoon et al. 2010; Criscuolo

and Gribaldo 2011; Shih, Wu et al. 2013). Therefore, the origin of Ces and CslA/C, if

the latter also come from endosymbiont events, probably reflect the complexity of the

formation of the original cyanobacterial endosymbiont.

Cellulose and hemicelluloses play major roles in plant cell wall formation,

organogenesis (xylem, phloem, etc.) and the ability to stand upright on land. The

phylogenetic tree shows that *Ces* gene first appeared in charophyte (*Mesotaenium*

caldariorum), which are thought to be the closest relatives of the land plants(McCourt,

Delwiche et al. 2004). CslC also firstly appeared in charophyte (*Chara globularis*) while CslA firstly appeared in green algae (*Ostreoccus* and *Micromonas*). These results suggest a stepwise origin and evolution of cellulose and hemicelluloses synthase. In both cases, the appearance of cellulose/-like synthases related genes in charophytes is beneficial, enabling land plants to colonize terrestrial environments. Previous research has also found that many cell wall polysaccharides have their evolutionary origins in the charophytes (Mikkelsen, Harholt et al. 2014). These studies speculate that some features of land plant cell walls evolved prior to the transition to land, rather than having evolved as a result of selection pressures inherent in this transition (Mikkelsen, Harholt et al. 2014).

Conclusion

Although there have been some studies involving the evolution of Ces and Csl genes in plants and animals (Nakashima, Yamada et al. 2004; Nobles and Brown 2004; Yin, Huang et al. 2009; Yin, Johns et al. 2014), large gaps exist due to limited genomic data and complex evolutionary scenarios for these genes. We conducted the phylogenetic analysis of Ces and Csl proteins using updated data sets, and proposed several hypotheses about the origin of Ces and Csl proteins in plants and tunicates. Firstly, horizontal transfer events of Ces might have occurred from bacteria to green algae, and then from green algae to tunicates through phagocytosis. Secondly, Ces may have originated from cyanobacteria through endosymbiosis and then was transferred to the nuclear genome in charophytes.

Finally, the origins of CslA/C genes are still unclear according to the existing data,

which might transfer from cyanobacteria or other type of bacteria. In conclusion,

multiple HGT events have played major roles in the origin of cellulose synthases and

cellulose synthases-like proteins in plants and tunicates.

References

Burton, R. A., S. M. Wilson, et al. (2006). "Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans." Science **311**(5769): 1940-1942.

Cocuron, J. C., O. Lerouxel, et al. (2007). "A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase." Proc Natl Acad Sci U S A **104**(20): 8550-8555.

Criscuolo, A. and S. Gribaldo (2011). "Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria." Molecular biology and evolution **28**(11): 3019-3032.

Dehal, P., Y. Satou, et al. (2002). "The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins." Science **298**(5601): 2157-2167.

Dehal, P., Y. Satou, et al. (2002). "The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins." Science **298**(5601): 2157-2167.

Deusch, O., G. Landan, et al. (2008). "Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor." Molecular biology and evolution **25**(4): 748-761.

Dhugga, K. S., R. Barreiro, et al. (2004). "Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family." Science **303**(5656): 363-366.

Djerbi, S., M. Lindskog, et al. (2005). "The genome sequence of black cottonwood

    (Populus trichocarpa) reveals 18 conserved cellulose synthase (CesA)

    genes." <u>Planta</u> **221**(5): 739-746.

Doblin, M. S., F. A. Pettolino, et al. (2009). "A barley cellulose synthase-like CSLH

    gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis."

    <u>Proc Natl Acad Sci U S A</u> **106**(14): 5996-6001.

Farrokhi, N., R. A. Burton, et al. (2006). "Plant cell wall biosynthesis: genetic,

    biochemical and functional genomics approaches to the identification of key

    genes." <u>Plant Biotechnol J</u> **4**(2): 145-167.

Guindon, S., J. F. Dufayard, et al. (2010). "New algorithms and methods to estimate

    maximum-likelihood phylogenies: assessing the performance of PhyML 3.0."

    <u>Syst Biol</u> **59**(3): 307-321.

Hazen, S. P., J. S. Scott-Craig, et al. (2002). "Cellulose synthase-like genes of rice."

    <u>Plant Physiol</u> **128**(2): 336-340.

Kimura, S., C. Ohshima, et al. (2001). "Cellulose in the house of the appendicularian

    Oikopleura rufescens." <u>Protoplasma</u> **216**(1-2): 71-74.

Le, S. Q. and O. Gascuel (2008). "An improved general amino acid replacement

    matrix." <u>Mol Biol Evol</u> **25**(7): 1307-1320.

Lerouxel, O., D. M. Cavalier, et al. (2006). "Biosynthesis of plant cell wall

    polysaccharides - a complex process." <u>Curr Opin Plant Biol</u> **9**(6): 621-630.

Liepman, A. H., C. G. Wilkerson, et al. (2005). "Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases." Proc Natl Acad Sci U S A **102**(6): 2221-2226.

Martin, W., T. Rujan, et al. (2002). "Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus." Proceedings of the National Academy of Sciences **99**(19): 12246-12251.

McCourt, R. M., C. F. Delwiche, et al. (2004). "Charophyte algae and land plant origins." Trends in Ecology & Evolution **19**(12): 661-666.

Mizrachi, E., S. D. Mansfield, and A. A. Myburg. (2012). "Cellulose factories: advancing bioenergy production from forest trees." New Phytol 194 (1):54-62.

Mikkelsen, M. D., J. Harholt, et al. (2014). "Evidence for land plant cell wall biosynthetic mechanisms in charophyte green algae." Annals of botany **114**(6): 1217-1236.

Nakashima, K., L. Yamada, et al. (2004). "The evolutionary origin of animal cellulose synthase." Development genes and evolution **214**(2): 81-88.

Nakashima, K., L. Yamada, et al. (2004). "The evolutionary origin of animal cellulose synthase." Dev Genes Evol **214**(2): 81-88.

Ni, T., J. Yue, et al. (2012). "Ancient gene transfer from algae to animals: mechanisms and evolutionary significance." BMC Evol Biol **12**: 83.

Ni, T., J. Yue, et al. (2012). "Ancient gene transfer from algae to animals: Mechanisms and evolutionary significance." BMC evolutionary biology **12**(1): 83.

Nobles, D. R. and R. M. Brown (2004). "The pivotal role of cyanobacteria in the evolution of cellulose synthases and cellulose synthase-like proteins." Cellulose **11**(3-4): 437-448.

Nobles, D. R., D. K. Romanovicz, et al. (2001). "Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase?" Plant physiology **127**(2): 529-542.

Pear, J. R., Y. Kawagoe, et al. (1996). "Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase." Proc Natl Acad Sci U S A **93**(22): 12637-12642.

Reyes-Prieto, A., H. S. Yoon, et al. (2010). "Differential gene retention in plastids of common recent origin." Molecular biology and evolution **27**(7): 1530-1537.

Richmond, T. A. and C. R. Somerville (2000). "The cellulose synthase superfamily." Plant Physiol **124**(2): 495-498.

Richmond, T. A. and C. R. Somerville (2001). "Integrative approaches to determining Csl function." Plant Mol Biol **47**(1-2): 131-143.

Roberts, A. W. and J. T. Bushoven (2007). "The cellulose synthase (CESA) gene superfamily of the moss Physcomitrella patens." Plant Mol Biol **63**(2): 207-219.

Roberts, J. A. and Z. Gonzalez-Carranza (2008). Annual Plant Reviews, Plant Cell Separation and Adhesion, John Wiley & Sons.

Römling, U. (2002). "Molecular biology of cellulose production in bacteria." <u>Research in microbiology</u> **153**(4): 205-212.

Sagane, Y., K. Zech, et al. (2010). "Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans." <u>Development</u> **137**(9): 1483-1492.

Sagane, Y., K. Zech, et al. (2010). "Functional specialization of cellulose synthase genes of prokaryotic origin in chordate larvaceans." <u>Development</u> **137**(9): 1483-1492.

Sasakura, Y., K. Nakashima, et al. (2005). "Transposon-mediated insertional mutagenesis revealed the functions of animal cellulose synthase in the ascidian Ciona intestinalis." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **102**(42): 15134-15139.

Shih, P. M., D. Wu, et al. (2013). "Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing." <u>Proceedings of the National Academy of Sciences</u> **110**(3): 1053-1058.

Suzuki, S., L. Li, et al. (2006). "The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in Populus trichocarpa." <u>Plant Physiol</u> **142**(3): 1233-1245.

Wightman, R. and S. Turner (2010). "Trafficking of the plant cellulose synthase complex." <u>Plant Physiol</u> **153**(2): 427-432.

Xu, H., K. F. Chater, et al. (2008). "A cellulose synthase-like protein involved in hyphal tip growth and morphological differentiation in Streptomyces." <u>Journal of bacteriology</u> **190**(14): 4971-4978.

Yin, Y., J. Huang, et al. (2009). "The cellulose synthase superfamily in fully sequenced plants and algae." <u>BMC Plant Biol</u> **9**: 99.

Yin, Y., J. Huang, et al. (2009). "The cellulose synthase superfamily in fully sequenced plants and algae." <u>BMC plant biology</u> **9**(1): 99.

Yin, Y., M. A. Johns, et al. (2014). "A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily." <u>BMC Genomics</u> **15**: 260.

Yin, Y., M. A. Johns, et al. (2014). "A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily." <u>BMC Genomics</u> **15**(269).

Yue, J., X. Hu, et al. (2012). "Widespread impact of horizontal gene transfer on plant colonization of land." <u>Nat Commun</u> **3**: 1152.

Table 4. 46 HGT-derived genes in *C. intestinalis.* These genes belong to 32 gene families. Function annotation and gene ID were listed.

HGT Candidate ID	Function Annotation	Gene ID
Cioint8818	glutathione S-transferase	XP_002128369.1
Cioint12112	alpha/beta hydrolase fold protein	XP_002127127.1
Cioint1885,	beta-hexaminidase, putative, hex20B	XP_004225944.1,
Cioint2671		XP_009858403.1
Cioint10242,	xanthine/uracil permease family protein	XP_002124198.1,
Cioint9494,		XP_002124527.1,
Cioint1672		XP_002124908.1
Cioint13648	sodium:neurotransmitter symporter	XP_004225622.1
Cioint832	protein up-regulated by thyroid hormone-putative PQQ-dependent glucose dehydrogenase	XP_002125570.1
Cioint3412	conserved hypothetical protein	XP_002130204.1
Cioint12457	hypothetical protein VspiD_29695	XP_002125364.1
Cioint2728,	sulfotransferase	XP_009860561.1,
Cioint2560		XP_002123028.1
Cioint3050,	hypothetical protein N9414_13215	XP_009858684.1,
Cioint3051,		XP_009861303.1,
Cioint3052		XP_002126405.2
Cioint4310	glycoside hydrolase family 31	XP_009858796.1
Cioint1214,	hypothetical protein	XP_002119185.1,

Cioint10354		XP_007889422.1, XP_007897409.1
Cioint1196	oxidoreductase, short chain dehydrogenase/reductase family	XP_002125705.1
Cioint693	aldehyde dehydrogenase	XP_002128052.1
Cioint1439, Cioint1478	Na+/glucose cotransporter	XP_002131980.1, XP_002120003.3
Cioint8801	peptidase T2, asparaginase 2	XP_002121188.1
Cioint1262	alpha-glucosidase	XP_004225484.1
Cioint5602, Cioint9788	FAD dependent oxidoreductase	XP_009857807.1, XP_002128931.1
Cioint13293	nad dependent epimerase/dehydratase family	XP_009859008.1
Cioint12197	cartilage oligomeric matrix protein	NP_001029015.1
Cioint2634	KR domain superfamily	XP_009859973.1
Cioint9435	alpha amylase catalytic region	XP_002123014.1
Cioint4359, Cioint7040	peroxisomal bifunctional enzyme	XP_002125312.1, XP_002125215.1
Cioint6080	sodium-coupled monocarboxylate transporter	XP_002130613.2
Cioint1575, Cioint1602	putative agmatinase	XP_002122881.1, XP_002122659.1
Cioint1376, Cioint754	Animal heme peroxidase	XP_002123273.1, XP_002127674.1
Cioint12952	Acyl-CoA dehydrogenase	XP_004227063.1

Cioint1006	hexose phosphate uptake regulatory protein UhpC	XP_002130897.1
Cioint352	hydrolase activity, hydrolyzing O-glycosyl compounds;carbohydrate metabolism	NP_001041448.1
Cioint12074, Cioint1438, Cioint8625	homocysteine S-methyltransferase	XP_002131994.1, XP_002125928.2, XP_002131974.1
Cioint51	hypothetical protein PPSIR1_31073	XP_002129943.1
Cioint2635	oxidoreductase, short chain dehydrogenase/reductase family protein	XP_009860042.1
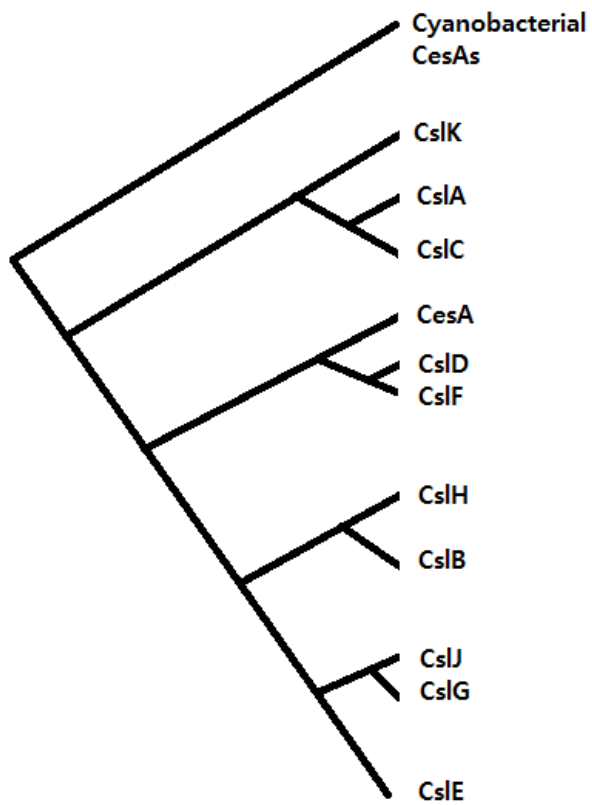
Fig. 3. Evolution of CesA/Csl superfamily with cyanobacteria as the out-group.

Length of branch does not reflect evolutionary distances.

A

B

Fig. 4. Phylogenetic tree (A) and multiple sequence alignment (B) of cellulose synthase (Ces) and cellulose synthase-like (Csl) proteins in urochordates. Boxed columns indicate the amino-acid residues shared by bacterial and *C. intestinalis* Ces sequences. Numbers above branches are bootstrap values from maximum likelihood method. Bootstrap values that were less than 50 were deleted from the tree.

Fig. 5. Phylogenetic tree of cellulose synthase (Ces) and cellulose synthase-like (Csl) proteins in plants.

Chapter 4: Horizontal gene transfer in *Trichoplax*

Introduction

Although the role of HGT has been widely accepted in prokaryotes and unicellular eukaryotes, heated debate continues concerning the scale and cumulative impact of HGT on the evolution of multicel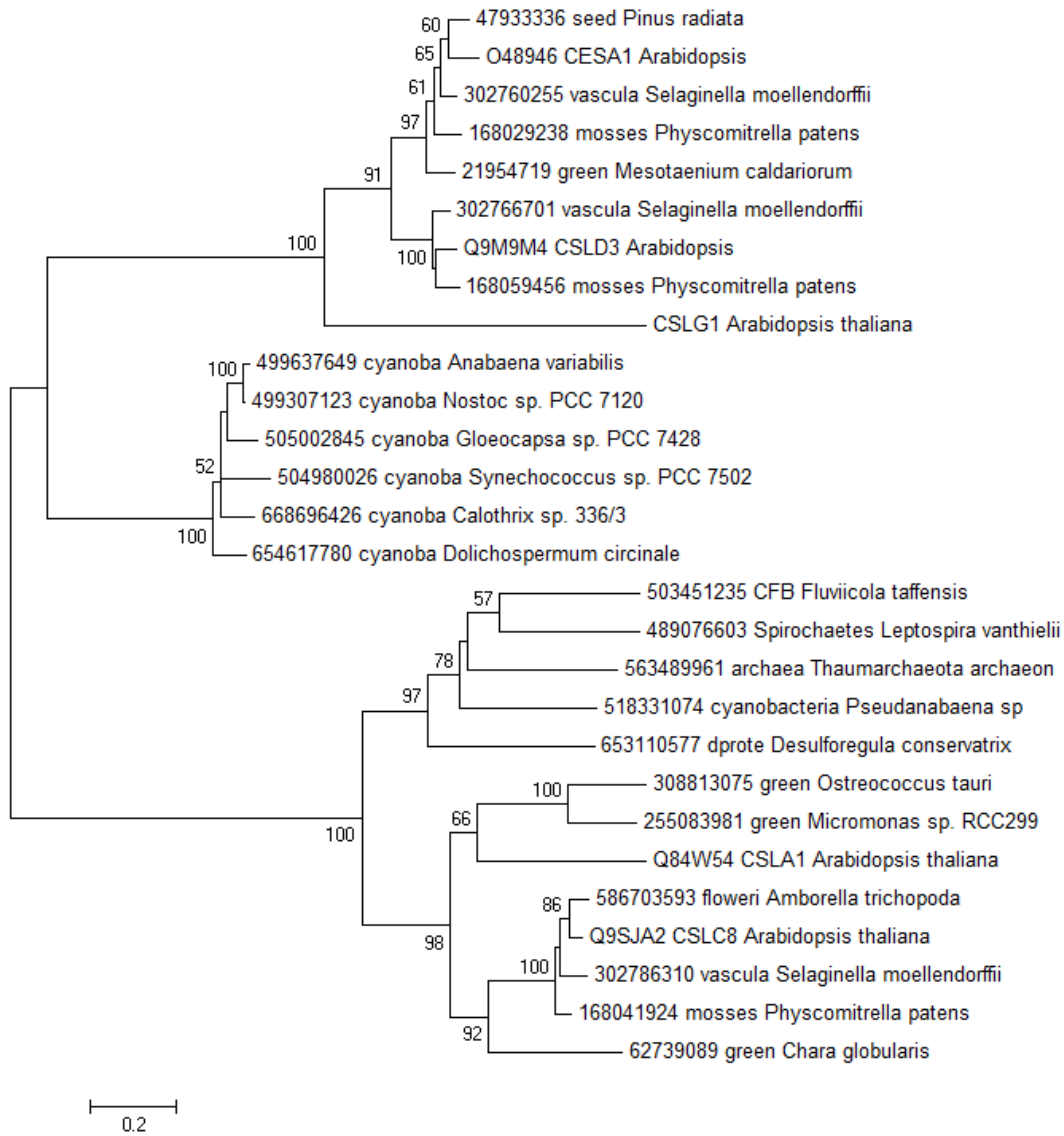lular eukaryotes, especially metazoans and land plants. Although reports of horizontally acquired genes in multicellular eukaryotes are accumulating (Dunning Hotopp, Clark et al. 2007), some researchers still believe that HGT events are rare due to the isolation of the germline cells in animals and the apical meristem in plants (Andersson 2005; Bock 2010). However, since all multicellular eukaryotes are derived from unicellular ancestors, ancient HGT events to the common ancestors may also have impacts on evolution of the whole descendent lineage (Huang and Gogarten 2006). HGT events might happen at different stages of evolution and have cumulative effects on the offspring lineages of HGT recipients (Yue, Hu et al. 2012). Furthermore, HGT events may also have happened in multicellular eukaryotes recently through feeding activities (Ni, Yue et al. 2012) and parasitism (Zhang, Qi et al. 2014). Thus, instead of simply ignoring occurrence of HGT events, it is necessary to study the scale of HGT events in multicellular eukaryotes and to assess the biological functions of individual horizontally derived genes in the diverse biological processes of the recipient species.

Placozoans are the simplest known free-living animals and were thought to consist of two species: *Trichoplax adhaerens* and *Treptoplax reptans*. The latter

species, however, has not been observed since its description in 1896, casting doubts on its existence. Therefore, it is commonly believed that *T. adhaerens* is the only extant species of phylum Placozoa. Both genetic and physiological evidence suggests that Placozoa represents the simplest phylum of animals. The genome size of *T. adhaerens* is only about 98 million base pairs (MB), the smallest among all sequenced animal genomes (Srivastava, Begovic et al. 2008). The body of *T. adhaerens* is made up of a few thousand cells with only four types: cylinder cells, gland cells, cover cells and fiber cells (Schierwater 2005). Obviously there are no more recently evolved cell types found in other animals, such as nerve cells, sensory cells or muscle cells. In contrast, sponges, which belong to the closest but later branched phylum Porifera, have 10 - 20 different types of cells (Ruppert, Fox et al. 2004); flies, a representative of the much later evolved phylum Arthropoda, have roughly 90 types; and mammals have over 200 types (Lodish 2008).

Trichoplax has two distinct sides: a bottom epithelial layer consisting of cylinder cells and gland cells and an upper epithelial layer composed of cover cells. Star-shaped fiber cells reside in the space between the two epithelial layers. Both in nature and in laboratory culture, *Trichoplax* can reproduce asexually, either by binary fission or, less often, by budding, although in some laboratories development of oocytes and sperm cells was observed when the population density was high (Grell 1972).

The relationship of Placozoa to other metazoans (Ctenophora, Porifera, Cnidaria and Bilateria) remains controversial. Four hypotheses have been proposed. The first

hypothesis suggests that Placozoa may be the earliest branch of animals based on comparisons of complete mitochondrial genomes (Dellaporta, Xu et al. 2006; Signorovitch, Buss et al. 2007) (Panel A of Fig. 6.). This hypothesis is consistent with the fact that placozoans have the smallest genome size and fewest cell types, as well as an extremely simple body plan. However, some early studies on a small group of genes (Ruiz-Trillo, Roger et al. 2008) led to the second hypothesis, which postulates that placozoans could be secondarily simplified cnidarians, while Porifera is the sister group of all other animals (Bridge, Cunningham et al. 1995) (Panel B of Fig. 6). These first two hypotheses are both consistent with the fact that nervous system components are absent from Placozoa or Porifera, but are present in other animals (Ryan and Chiodin 2015). The third hypothesis was proposed based on analyses of small subunit ribosomal RNA (18S), which suggested that placozoans might be the sister group to bilaterians (Collins 1998) or the earliest eumetazoan branch (Silva, Muschner et al. 2007) (Panel C of Fig. 6). However, recent phylogenomic analyses led to a fourth hypothesis, posing that Ctenophora is the sister group of all other animals (Dunn, Hejnol et al. 2008; Hejnol, Obst et al. 2009; Ryan, Pang et al. 2013; Moroz 2014; Whelan, Kocot et al. 2015) (Panel D of Fig. 6).

Due to their aquatic habitat, placozoans are in closer physical contact with bacteria and algae, thus increasing the probability of acquiring foreign genes from bacteria and/or algae. In addition, *T. adhaerens* is known to harbor an uncharacterized gram-negative endosymbiont in fiber cells (Grell 1972; Grell and Benwitz 1974; Driscoll, Gillespie et al. 2013), and the genes of those endosymbiotic

bacteria can pass on to developing oocytes via fiber cell extensions (Eitel, Guidi et al. 2011). All of these characteristics suggest that HGT may happen to *T. adhaerens*. Currently, there is only one case study on HGT in *T. adhaerens*. An abundant alkyl sulphatase BDS1 was identified to be horizontally acquired from proteobacteria to baker's yeast *Saccharomyces cerevisiae* (Hall, Brachat et al. 2005) as well as *T. adhaerens* (Ringrose, van den Toorn et al. 2013). However, to our knowledge, there is no systematic study of HGT events in *T. adhaerens*. In this study, phylogenetic analyses were performed using the genome of *T. adhaerens* and 43 genes from 21 families were identified as horizontally transferred from green plants, bacteria or archaea. These genes are related to many aspects of cellular metabolism and development, including phenazine antibiotic activity and neural development, suggesting their potential roles in the evolution of placozoans and metazoans in general.

Materials and methods

Data sources and genome screening

11,520 protein sequences of *T. adhaerens* were downloaded from JGI (Joint Genome Institute, http://genome.jgi-psf.org/Triad1/Triad1.download.ftp.html). Customized database construction, local BLAST and AlienG analysis were carried out as described in Chapter 2.

Phylogenetic analyses

For each HGT candidate identified from AlienG analyses, I performed phylogenetic analyses manually. Taxonomic distribution, gene structure, and shared domains were analyzed as additional evidence to assess the origin of each candidate gene. Protein sequences were sampled from each domain of life (bacteria, archaea and eukaryotes). Sequence sampling is broad and balanced, which means that all major lineages are covered and no particular lineages were sampled in a biased manner. In order for our sampling to be as broad as possible, I not only used the BLAST results against the NCBI nr database, but also performed local BLAST against other available eukaryotic genomes and EST databases that had not been archived by NCBI.

The default parameters were used for web-based BLAST except that the maximum number of target sequences was changed from 100 to 10,000, and the E-value cutoff was changed from 10 to 0.001. The number of target sequences was increased in order to include as many lineages as possible for further analyses. The bit scores were used in AlienG calculations, but the E-values were used for a first-step filtration. The E-value is a parameter that describes the "expected" number of hits with a specific raw score or higher when a query sequence is used to search against a database of a particular size. According to the equation E-value=$mn2^{-S}$ (m is the length of query database, n is the length of query sequence, S is bit score), the higher the bit score, the lower the E-value, given a specific sequence and a specific database. As such, E-value is a measure of the probability that two sequences are

similar only by chance. The E-value cutoff was decreased to 0.001 to reduce the false positive rate.

Usually the number of BLAST results for a query sequence is so large that it is impossible to use all of them for tree construction, so the number of candidate sequences has to be reduced using several criteria. Some lineages are well studied and often over-represented, while others are poorly studied and often under-represented. Based on the "broad and balanced" sampling principle, one to three representative organisms were kept in each lineage in my analyses. If more than one representative organism was found in a lineage, sequences of model organisms would be preferred, because genes are usually better annotated in model organisms (for example, *Caenorhabditis elegans* was preferred to *C. remanei*). If there were many hits in an organism, sequences would be ranked in order of the E-value and the first one was used when two or more sequences had the same level of divergence.

Multiple protein sequence alignments were performed using ClustralX and MUSCLE with the default parameters. Gaps were removed and mismatched sites were manually adjusted. Sequences that were too long, too short or obviously deviated from other sequences and would cause aberrant alignments were deleted from multiple alignments. ModelGenerator was used to select the best-fit model of protein substitution and rate heterogeneity for the data (Keane, Creevey et al. 2006). PhyML 3.0 (maximum likelihood methods) was used to perform phylogenetic analyses (Guindon, Dufayard et al. 2010). One hundred pseudo-replicates were

conducted for bootstrap analyses. The results of the phylogenetic analyses were visualized using TreeGraph 2 (Stover and Muller 2010). Bootstrap values that were less than 50 were deleted from the phylogenetic trees.

Results and discussion

HGT-derived genes in Trichoplax

In this study, we focused on genes in metazoans that were horizontally acquired from other sources, including green plants, bacteria and archaea, mainly based on phylogenomic analyses of *T. adhaerens*. Of 11,520 annotated protein sequences of *T. adhaerens*, 216 were identified as candidates of transferred genes. After analyses of taxonomic distributions, multiple sequence alignments and molecular phylogenies, 43 genes of 21 families were identified as horizontally transferred genes from green plants, bacteria or archaea (Table 5). Among all identified 43 genes, 30 genes of 14 families showed a monophyletic relationship with bacterial sequences; 12 genes of six families are likely derived from green plants and one gene from archaea (Table 5). Along with phylogenetic trees, shared indels between *T. adhaerens* and other species also indicate a common origin of those genes in many cases, which suggests horizontal acquisition from donor species to *T. adhaerens*. All of these 43 genes are located on large genomic scaffolds; therefore they are less likely to be due to bacterial or algal contamination.

GO analyses of HGT-derived genes in Trichoplax

Gene Ontology (GO) analyses of those 43 genes were performed using the

PANTHER (Protein ANalysis THrough Evolutionary Relationships) database

(http://pantherdb.org/) (Thomas, Campbell et al. 2003; Mi, Lazareva-Ulitsky et al.

2005). PANTHER family and three types of GO analyses (Molecular Function, MF;

Biological Process, BP; Cellular Component, CC) are listed (Table 6).

34 out of these 43 genes had hits associated with function in the GO analyses

(Fig. 7). In the "Molecular Function" analysis, more than half of these acquired genes

(21/34, 61.76%) were involved in "catalytic activity". Most of the genes encode

simple enzymes that can act alone, such as serine protease, oxidoreductase and

hydrolase, instead of being a subunit of a large complex. In order for a subunit of an

enzyme complex to fulfill its catalytic functions, all other subunits have to be

available in recipient's genome or horizontally acquired from other species, which is

usually unrealistic. That might explain why most horizontally transferred genes are

simple enzymes. In the "Biological Process" analysis, the largest category is

"metabolic process" (22/34, 64.71%). The genes identified can be divided into two

functionally distinct groups: informational and operational genes (Rivera, Jain et al.

1998). The former group refers to genes which function in transcription, translation,

replication and related processes, while the latter group is usually involved in

"housekeeping" processes such as metabolism and regulatory functions (Rivera,

Jain et al. 1998). Most horizontally transferred genes belong to operational genes,

because they can quickly work together with genes in recipient species and confer

an advantage to adapt to a new niche (Jain, Rivera et al. 1999; Keeling and Palmer 2008).

Since false positive conclusions may be drawn from poor data quality and/or methodological limitations in analyzing HGT events (Huang and Gogarten 2006), a number of precautionary measures were taken to avoid these issues. These include construction of a comprehensive database, broad and balanced taxonomic sampling, careful inspection of alignments, and detection of other molecular characters consistent with the identified relationships. These measures, to a large degree, may help avoid most of the artifacts in HGT identification. Although HGT events can always be explained alternatively by differential gene loss, the former is the most parsimonious interpretation for all cases listed in Table 5. Furthermore, some identified HGT genes (Triad8318, Triad10384, Triad10385 and Triad11269) are supported by independent evidence such as the sharing of indels and/or amino acid residues, which makes the HGT scenario more likely.

Phenazine biosynthesis-like domain-containing protein (PBLD)

Phenazines are secondary metabolites of bacteria, and natural phenazines are heterocyclic pigments containing nitrogen (Laursen and Nielsen 2004). Phenazines can be used to produce derivatives including many dyestuffs used in industry. Phenazine compounds found in nature are produced by diverse bacteria of the genera *Pseudomonas* (Mavrodi, Blankenfeldt et al. 2006), *Streptomyces* (McDonald, Wilkinson et al. 1999), *Pantoea* (Galbraith, Giddens et al. 2004), *Brevibacterium* (Choi, Kwon et al. 2009), *Burkholderia* and *Pectobacterium* (Mavrodi, Peever et al.

2010), most of which are soil-dwelling or plant-associated strains. These natural

phenazine products have been implicated in the virulence and competitive fitness of

producing organisms, as they are involved in broad-spectrum antibiotic activities

towards bacteria, fungi and other eukaryotes (Blankenfeldt, Kuzin et al. 2004). For

example, phenazine pyocyanin produced by *Pseudomonas aeruginosa* contributes

to its ability to colonize the lungs of cystic fibrosis patients by killing other competing

bacteria. Similarly, phenazine-1-carboxylic acid, produced by a number of

Pseudomonas species, increases bacterial survival in soil environments and has

been shown to be essential for the biological control of certain strains (McDonald,

Mavrodi et al. 2001; Dietrich, Okegbe et al. 2013).

Previous research indicates that HGT of PBLD happens among different

bacterial species such as *Streptomyces cinnamonensis*, *Pantoea agglomerans*,

Burkholderia cepacia, *Pectobacterium atrosepticum*, *Brevibacterium linens*, and

Mycobacterium abscessus (Fitzpatrick 2009). Interkingdom gene transfer of PBLD

also occurs from proteobacteria to the fungus *Candida parapsilosis* (Fitzpatrick,

Logue et al. 2008). In this study, we found a family of four genes (Triad8318,

Triad10384, Triad10385 and Triad11269) that are homologues of phenazine

biosynthetic genes. When we searched against the Panther database, we found that

all four of them contained a phenazine biosynthesis-like domain. PBLD genes can

only be found in metazoans, fungi, bacteria and archaea, but not in green plants.

The fungal PBLD gene grouped with two beta-proteobacterial sequences with a

bootstrap value 100 (Fig. 8A), suggesting that PBLD gene was horizontally

transferred from proteobacteria to fungi; this relationship is also consistent with previous research (Fitzpatrick, Logue et al. 2008). All four metazoan PBLD genes are nested together with two chlamydial and one firmicute sequences, with a bootstrap value 100. Three archaeal sequences were used as the outgroup (Fig. 8. Panel B). This finding implies that phenazine may also be utilized in *T. adhaerens* as an antibiotic to increase its competitive fitness, although no similar cases were reported.

The phenazine biosynthetic pathway involves key enzymes (*phz* gene family) and housekeeping genes (*rrs*, *recA*, *rpoB*, *atpD*, and *gyrB*) (Mavrodi, Peever et al. 2010). Although some steps remain unclear, much progress has been made towards understanding the enzymes and mechanisms concerning phenazine biosynthesis, especially in *Pseudomonas fluorescens* (McDonald, Mavrodi et al. 2001). Seven members of *phz* family (*phzA-G*) form a gene cluster and work together to produce phenazine-1-carboxylic acid (PCA), the specific form the phenazine in *P. fluorescens*. Among these seven *phz* genes, *phz*F and *phz*G catalyze the formation of phenazine ring and hence are the critical enzymes for the phenazine biosynthesis (McDonald, Mavrodi et al. 2001). PBLD genes in *T. adhaerens* are homologues of *phzG*, which is similar to bacterial pyridoxamine 5′-phosphate oxidases, while no homologue of *phzA-F* can be found in metazoans.

Incorrect identification of HGT may be caused by multiple factors, including poor genome data quality or improper methodology. However, independent evidence such as shared domain structures between recipient species and potential donors can

reduce most of these artifacts. Aside from the gene tree topology that shows a

placozoans/bacteria clade with strong bootstrap support (100%), shared gene

structures like indels or conserved amino acid residues can also be used to support

the common origin of placozoans/bacteria PBLD genes (Fig. 8. Panel A). Six boxed

amino acid residues are shared by four placozoan sequences and three bacterial

homologues, but these shared amino acid residues can hardly be found in other

species. Shared amino acid residues between placozoan and bacterial PBLD genes

further support the hypothesis that placozoans have acquired them from chlamydias

or firmicutes.

Peptidylglycine alpha-amidating monooxygenase (PAM)

Although *T. adhaerens* does not have a nervous system, including nerves or

specialized sensory cells, genes coding for components of animal nervous systems

can be found in this species (Srivastava, Begovic et al. 2008). Those "nervous

system" genes include transcription factors with pro-neural activity, neurotransmitter

biosynthetic genes, pre- and postsynaptic genes, subunits of potassium, sodium and

calcium channels, GPCR family involved in signal transduction and genes

associated with axon guidance and neural migration (Srivastava, Begovic et al.

2008). Interestingly, although placozoans have no sensory cells, they do contain

some neuro-secretory cells dispersed around their outer edge (Grell and Ruthmann

1991), which can express neural proteins like syntaxin, synapsin and synaptobrevin

(Smith, Varoqueaux et al. 2014). Those cells also express FMRFamide-like peptide

(FLPs) and are sensitive to antibodies against FLPs (Schuchert 1993). The FLPs

neuropeptide family is usually present in the nervous system of later animals (Price

and Greenberg 1977) and is often used to specifically visualize the nervous system

of whole-mounted coelenterates such as *Hydra* (Grimmelikhuijzen 1985).

Secreted peptides play important regulatory roles as neurotransmitter and

hormones in the animal kingdom. Usually a carboxy-terminal modification from a

glycine residue to an amide is needed to activate secreted peptides to perform their

biological activities (Driscoll, Mueller et al. 1999). Two enzymes are involved in this

activation process, namely, a monooxygenase (peptidylglycine alpha-hydroxylating

monooxygenase or PHM) and an amidating lyase (peptidyl-alpha-hydroxyglycine

alpha-amidating lyase or PAL). In vertebrates and some invertebrates, the enzymatic

activities of monooxygenase and amidating lyase are performed by a single

bifunctional enzyme known as peptidylglycine alpha-amidating monooxygenase

(PAM), which sequentially catalyzes neuroendocrine peptides to their active final

products (Eipper, Milgram et al. 1993). Surprisingly, PAMs are also present in green

algae, but not in green plants or fungi, suggesting that horizontal gene transfer may

account for the intriguing phylogenetic distribution (Attenborough, Hayward et al.

2012).

Indeed, our analyses support the hypothesis that PAMs were horizontally

transferred from bacteria to the most recent common ancestor of all animals, with the

exception of Ctenophora (Fig. 9.). In lower animals such as *T. adhaerens* and *A.

queenslandica*, PAMs cannot activate neurotransmitters or hormones, since no

nervous or endocrine system is present in those animals. Instead, in nonbilaterian

animals PAMs work as epitheliopeptide amidase. (Attenborough, Hayward et al. 2012). Expression of PAM genes in *A. queenslandica* was also reported in previous research (Sakarya, Armstrong et al. 2007). Except for PAM, there are also a substantial number of "neural" genes that have pre-metazoan origins. For example, members of the membrane-associated guanylate kinase protein family are present in protists such as the choanoflagellate *Monosiga ovata* and the amoeba *Capsaspora owczarzaki* (Ruiz-Trillo, Roger et al. 2008) and MAGUK protein gene family in choanoflagellates (de Mendoza, Suga et al. 2010). Based on the finding that so many neural genes predate the origins of nervous system, some researchers have proposed a "two-step model" for the evolution of the eumetazoan nervous system (Galliot and Quiquand 2011). In the first step, ancient genes that had various roles in other process were recruited to perform new functions. These genes usually predate the origin of eumetazoan. In the second step, novel genes/gene families emerged, which usually happened in the most recent common ancestor of eumetazoans (Galliot and Quiquand 2011).

When placozoan PAM protein sequence was BLASTed against each of five major linages of animals, no hits were found in the Ctenophora. This result implies that Ctenophora may be the sister group of all other animals and is consistent with recent findings (Ryan, Pang et al. 2013; Whelan, Kocot et al. 2015). However, nervous systems can be found in Ctenophora, Cnidaria and Bilateria, but are absent in Placozoa and Porifera, which can be explained by two possible scenarios: either (1) nervous systems evolved independently in Ctenophora and in the clade of

Cnidaria and Bilateria, or (2) nervous systems arose in the most recent common ancestor of all animals but were lost in Placozoa and Porifera (Ryan and Chiodin 2015). A recent review paper states that scenario two seems to be more likely (Ryan and Chiodin 2015). Given our result that PAM protein cannot be found in Ctenophora, however, it seems likely that nervous systems evolved independently in Ctenophora without PAM proteins. In fact, besides PAM proteins, many critical components involved in development of neurons and neural signaling pathways do not have homologues in the genomes of Ctenophora (Moroz, Kocot et al. 2014; Moroz 2015).

In summary, a possible scenario of nervous system evolution in animals is given below. After the split of Ctenophora and the most recent common ancestor of all other animals, nervous systems evolved independently in Ctenophora, but its components and signaling pathways are distinct from those in other animals. PAMs was horizontally transferred from bacteria to the common ancestor of other animals and passed on to the descendent lineages. They belong to the first category in the "two-step model", and function as epitheliopeptide amidases in *T. adhaerens* and *A. queenslandica*. These PAM genes were later recruited to function as a component of nervous systems. As more and more neuronal fate-determining genes and signal pathway components evolved over time, PAMs were co-opted to interact with these genes and work as activators of neurotransmitters or hormones, leading to more sophisticated nervous systems in Cnidaria and Bilateria.

Conclusions

This research is the first systematic study of HGT events in *T. adhaerens* to our knowledge. A total of 43 genes in 21 families were detected to be horizontally acquired from bacteria, archaea or green plants. More than half of these HGT genes are operational genes or encode simple enzymes, probably increasing the competitive fitness of placozoans in the context of exploring new niches. Two specific genes (PBLD and PAM) were carefully examined and presented here. Placozoan PBLD genes are homologues of bacterial *phzG* gene, one of the two critical enzymes in phenazine biosynthesis. However, from our data, it is hard to determine whether phenazines are really produced in *T. adhaerens* or PBLD genes are involved in other activities. In vertebrates and some invertebrates, PAM genes function to activate neuropeptides. However, PAM genes are also detected in *T. adhaerens*, in which a nervous system is absent. It remains a fundamental question how PAM genes and the nervous system evolved in metazoans. The relationship between PAM genes and the nervous system is also intriguing.

References

Andersson, J. O. (2005). "Lateral gene transfer in eukaryotes." <u>Cell Mol Life Sci</u> **62**(11): 1182-1197.

Attenborough, R. M., D. C. Hayward, et al. (2012). "A "neural" enzyme in nonbilaterian animals and algae: preneural origins for peptidylglycine alpha-amidating monooxygenase." <u>Mol Biol Evol</u> **29**(10): 3095-3109.

Blankenfeldt, W., A. P. Kuzin, et al. (2004). "Structure and function of the phenazine biosynthetic protein PhzF from Pseudomonas fluorescens." <u>Proc Natl Acad Sci U S A</u> **101**(47): 16431-16436.

Bock, R. (2010). "The give-and-take of DNA: horizontal gene transfer in plants." <u>Trends Plant Sci</u> **15**(1): 11-22.

Bridge, D., C. W. Cunningham, et al. (1995). "Class-level relationships in the phylum Cnidaria: molecular and morphological evidence." <u>Mol Biol Evol</u> **12**(4): 679-689.

Choi, E. J., H. C. Kwon, et al. (2009). "6-Hydroxymethyl-1-phenazine-carboxamide and 1,6-phenazinedimethanol from a marine bacterium, Brevibacterium sp. KMD 003, associated with marine purple vase sponge." <u>J Antibiot (Tokyo)</u> **62**(11): 621-624.

Collins, A. G. (1998). "Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence." <u>Proc Natl Acad Sci U S A</u> **95**(26): 15458-15463.

de Mendoza, A., H. Suga, et al. (2010). "Evolution of the MAGUK protein gene family in premetazoan lineages." BMC Evol Biol **10**: 93.

Dellaporta, S. L., A. Xu, et al. (2006). "Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum." Proc Natl Acad Sci U S A **103**(23): 8751-8756.

Dietrich, L. E., C. Okegbe, et al. (2013). "Bacterial community morphogenesis is intimately linked to the intracellular redox state." Journal of bacteriology **195**(7): 1371-1380.

Driscoll, T., J. J. Gillespie, et al. (2013). "Bacterial DNA sifted from the Trichoplax adhaerens (Animalia: Placozoa) genome project reveals a putative rickettsial endosymbiont." Genome Biol Evol **5**(4): 621-645.

Driscoll, W. J., S. A. Mueller, et al. (1999). "Differential regulation of peptide alpha-amidation by dexamethasone and disulfiram." Mol Pharmacol **55**(6): 1067-1076.

Dunn, C. W., A. Hejnol, et al. (2008). "Broad phylogenomic sampling improves resolution of the animal tree of life." Nature **452**(7188): 745-749.

Dunning Hotopp, J. C., M. E. Clark, et al. (2007). "Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes." Science **317**(5845): 1753-1756.

Eipper, B. A., S. L. Milgram, et al. (1993). "Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains." Protein Sci **2**(4): 489-497.

Eitel, M., L. Guidi, et al. (2011). "New insights into placozoan sexual reproduction and development." PLoS One **6**(5): e19639.

Fitzpatrick, D. A. (2009). "Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species." J Mol Evol **68**(2): 171-185.

Fitzpatrick, D. A., M. E. Logue, et al. (2008). "Evidence of recent interkingdom horizontal gene transfer between bacteria and Candida parapsilosis." BMC Evol Biol **8**: 181.

Galbraith, M. D., S. R. Giddens, et al. (2004). "Role of glutamine synthetase in phenazine antibiotic production by Pantoea agglomerans Eh1087." Can J Microbiol **50**(10): 877-881.

Galliot, B. and M. Quiquand (2011). "A two-step process in the emergence of neurogenesis." Eur J Neurosci **34**(6): 847-862.

Grell, K. and G. Benwitz (1974). "Spezifische Verbindungsstrukturen der Faserzellen von Trichoplax adhaerens FE Schulze." Zeitschrift für Naturforschung C **29**: 790.

Grell, K. G. (1972). "Eibildung und Furchung von Trichoplax adhaerens FE Schulze (Placozoa)." Zoomorphology **73**(4): 297-314.

Grell, K. G. and A. Ruthmann (1991). "Placozoa." Microscopic anatomy of invertebrates **2**: 13-27.

Grimmelikhuijzen, C. (1985). "Antisera to the sequence Arg-Phe-amide visualize neuronal centralization in hydroid polyps." Cell and tissue research **241**(1): 171-182.

Guindon, S., J. F. Dufayard, et al. (2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." Syst Biol **59**(3): 307-321.

Hall, C., S. Brachat, et al. (2005). "Contribution of horizontal gene transfer to the evolution of Saccharomyces cerevisiae." Eukaryot Cell **4**(6): 1102-1115.

Hejnol, A., M. Obst, et al. (2009). "Assessing the root of bilaterian animals with scalable phylogenomic methods." Proc Biol Sci **276**(1677): 4261-4270.

Huang, J. and J. P. Gogarten (2006). "Ancient horizontal gene transfer can benefit phylogenetic reconstruction." Trends Genet **22**(7): 361-366.

Jain, R., M. C. Rivera, et al. (1999). "Horizontal gene transfer among genomes: the complexity hypothesis." Proc Natl Acad Sci U S A **96**(7): 3801-3806.

Keane, T. M., C. J. Creevey, et al. (2006). "Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified." BMC Evol Biol **6**: 29.

Keeling, P. J. and J. D. Palmer (2008). "Horizontal gene transfer in eukaryotic evolution." Nat Rev Genet **9**(8): 605-618.

Laursen, J. B. and J. Nielsen (2004). "Phenazine natural products: biosynthesis, synthetic analogues, and biological activity." Chem Rev **104**(3): 1663-1686.

Lodish, H. (2008). Molecular cell biology, Macmillan.

Mavrodi, D. V., W. Blankenfeldt, et al. (2006). "Phenazine compounds in fluorescent Pseudomonas spp. biosynthesis and regulation." <u>Annu Rev Phytopathol</u> **44**: 417-445.

Mavrodi, D. V., T. L. Peever, et al. (2010). "Diversity and evolution of the phenazine biosynthesis pathway." <u>Appl Environ Microbiol</u> **76**(3): 866-879.

McDonald, M., D. V. Mavrodi, et al. (2001). "Phenazine biosynthesis in Pseudomonas fluorescens: branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid." <u>J Am Chem Soc</u> **123**(38): 9459-9460.

McDonald, M., B. Wilkinson, et al. (1999). "Biosynthesis of phenazine antibiotics in Streptomyces antibioticus: stereochemistry of methyl transfer from carbon-2 of acetate." <u>Journal of the American Chemical Society</u> **121**(24): 5619-5624.

Mi, H., B. Lazareva-Ulitsky, et al. (2005). "The PANTHER database of protein families, subfamilies, functions and pathways." <u>Nucleic Acids Res</u> **33**(Database issue): D284-288.

Moroz, L. L. (2014). "The genealogy of genealogy of neurons." <u>Commun Integr Biol</u> **7**(6): e993269.

Moroz, L. L. (2015). "Convergent evolution of neural systems in ctenophores." <u>J Exp Biol</u> **218**(Pt 4): 598-611.

Moroz, L. L., K. M. Kocot, et al. (2014). "The ctenophore genome and the evolutionary origins of neural systems." <u>Nature</u> **510**(7503): 109-114.

Ni, T., J. Yue, et al. (2012). "Ancient gene transfer from algae to animals: mechanisms and evolutionary significance." BMC Evol Biol **12**: 83.

Price, D. A. and M. J. Greenberg (1977). "Structure of a molluscan cardioexcitatory neuropeptide." Science **197**(4304): 670-671.

Ringrose, J. H., H. W. van den Toorn, et al. (2013). "Deep proteome profiling of Trichoplax adhaerens reveals remarkable features at the origin of metazoan multicellularity." Nat Commun **4**: 1408.

Rivera, M. C., R. Jain, et al. (1998). "Genomic evidence for two functionally distinct gene classes." Proc Natl Acad Sci U S A **95**(11): 6239-6244.

Ruiz-Trillo, I., A. J. Roger, et al. (2008). "A phylogenomic investigation into the origin of metazoa." Mol Biol Evol **25**(4): 664-672.

Ruppert, E., R. Fox, et al. (2004). "Invertebrate Zoology . Brooks/Cole." CA, USA.

Ryan, J. F. and M. Chiodin (2015). "Where is my mind? How sponges and placozoans may have lost neural cell types." Philos Trans R Soc Lond B Biol Sci **370**(1684).

Ryan, J. F., K. Pang, et al. (2013). "The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution." Science **342**(6164): 1242592.

Sakarya, O., K. A. Armstrong, et al. (2007). "A post-synaptic scaffold at the origin of the animal kingdom." PLoS One **2**(6): e506.

Schierwater, B. (2005). "My favorite animal, Trichoplax adhaerens." Bioessays **27**(12): 1294-1302.

Schmidt, H. A., K. Strimmer, et al. (2002). "TREE-PUZZLE: maximum likelihood

phylogenetic analysis using quartets and parallel computing." <u>Bioinformatics</u>

18(3): 502-504.

Schuchert, P. (1993). "Trichoplax adhaerens (phylum Placozoa) has cells that react

with antibodies against the neuropeptide RFamide." <u>Acta Zoologica</u> **74**(2):

115-117.

Signorovitch, A. Y., L. W. Buss, et al. (2007). "Comparative genomics of large

mitochondria in placozoans." <u>PLoS Genet</u> **3**(1): e13.

Silva, F. B. d., V. C. Muschner, et al. (2007). "Phylogenetic position of Placozoa

based on large subunit (LSU) and small subunit (SSU) rRNA genes."

<u>Genetics and Molecular Biology</u> **30**(1): 127-132.

Smith, C. L., F. Varoqueaux, et al. (2014). "Novel cell types, neurosecretory cells,

and body plan of the early-diverging metazoan Trichoplax adhaerens." <u>Curr

Biol</u> **24**(14): 1565-1572.

Srivastava, M., E. Begovic, et al. (2008). "The Trichoplax genome and the nature of

placozoans." <u>Nature</u> **454**(7207): 955-960.

Stover, B. C. and K. F. Muller (2010). "TreeGraph 2: combining and visualizing

evidence from different phylogenetic analyses." <u>BMC Bioinformatics</u> **11**: 7.

Thomas, P. D., M. J. Campbell, et al. (2003). "PANTHER: a library of protein families

and subfamilies indexed by function." <u>Genome Res</u> **13**(9): 2129-2141.

Whelan, N. V., K. M. Kocot, et al. (2015). "Error, signal, and the placement of Ctenophora sister to all other animals." Proc Natl Acad Sci U S A **112**(18): 5773-5778.

Yue, J., X. Hu, et al. (2012). "Widespread impact of horizontal gene transfer on plant colonization of land." Nat Commun **3**: 1152.

Zhang, D., J. Qi, et al. (2014). "Root parasitic plant Orobanche aegyptiaca and shoot parasitic plant Cuscuta australis obtained Brassicaceae-specific strictosidine synthase-like genes by horizontal gene transfer." BMC Plant Biol **14**: 19.

Table 5. Horizontally transferred genes identified in *T. adhaerens*.

Gene ID	Annotation	Putative donor	Functional category
Triad9800	Alcohol dehydrogenase 3-like	Archaea	Oxidoreductase
Triad8145	Glutathione synthetase	Bacteria	Oxidative stress resistance
Triad8146	Glutathione synthetase	Bacteria	Oxidative stress resistance
Triad8072	Potassium channel protein	Bacteria	Potassium transport
Triad8071	Potassium channel protein	Bacteria	Potassium transport
Triad8078	Potassium channel protein	Bacteria	Potassium transport
Triad8075	Potassium channel protein	Bacteria	Potassium transport
Triad8077	Potassium channel protein	Bacteria	Potassium transport
Triad8076	Potassium channel protein	Bacteria	Potassium transport
Triad8073	Potassium channel protein	Bacteria	Potassium transport
Triad6224	Putative transcriptional regulator domain protein	Bacteria	Transcriptional regulator
Triad6223	Putative transcriptional regulator domain	Bacteria	Transcriptional regulator

	protein		
Triad6214	Putative transcriptional regulator domain protein	Bacteria	Transcriptional regulator
Triad6222	Putative transcriptional regulator domain protein	Bacteria	Transcriptional regulator
Triad6221	Putative transcriptional regulator domain protein	Bacteria	Transcriptional regulator
Triad5587	Amidohydrolase	Bacteria	Metabolism of purine and pyrimidine ring
Triad5585	Amidohydrolase	Bacteria	Metabolism of purine and pyrimidine ring
Triad2364	Putative adolase/adducing	Bacteria	Carbohydrate transport and metabolism
Triad1567	Carbamoyl transferase	Bacteria	Pyrimidine biosynthesis
Triad1430	Poly(ADP-ribose) glycohydrolase ARH3	Bacteria	Cellular response to superoxide

Triad1431	Poly(ADP-ribose) glycohydrolase ARH3	Bacteria	Cellular response to superoxide
Triad476	Peptidyl-glycine alpha-amidating monooxygenase isoform X2	Bacteria	Neuropeptide bioactivation
Triad10012	Alpha/beta hydrolase domain-containing protein 11-like	Bacteria	Hydrolase
Triad10098	Tetratricopeptide repeat protein 28 isoform X1	Bacteria	Cell division
Triad9802	Beta-carbonic anhydrase	Bacteria	Acid-base balance
Triad8202	Ribosome biogenesis protein	Bacteria	Ribosome biogenesis
Triad11269	Phenazine biosynthesis-like domain-containing protein	Chlamydias	Isomerase
Triad10385	Phenazine biosynthesis-like domain-containing protein	Chlamydias	Isomerase
Triad10384	Phenazine biosynthesis-like domain-containing protein	Chlamydias	Isomerase

Triad8318	Phenazine biosynthesis-like domain-containing protein	Chlamydias	Isomerase
Triad7572	Procollagen galactosyltransferase	High GC Gram+	Transferase
Triad8243	Speckle-type POZ protein variant	Viridiplantae	Transcriptional repression
Triad8266	Speckle-type POZ protein variant	Viridiplantae	Transcriptional repression
Triad7945	Voltage-dependent T-type calcium channel subunit alpha-1I	Viridiplantae	Calcium transport
Triad7684	Cytochrome P450 74A	Viridiplantae	Bioactivation and detoxication
Triad7682	Cytochrome P450 75A	Viridiplantae	Bioactivation and detoxication
Triad7683	Cytochrome P450 76A	Viridiplantae	Bioactivation and detoxication
Triad7482	Retinoid-inducible serine carboxypeptidase	Viridiplantae	Proliferation repression
Triad7481	Retinoid-inducible serine carboxypeptidase	Viridiplantae	Proliferation repression
Triad5101	Retinoid-inducible serine	Viridiplantae	Proliferation repression

carboxypeptidase

Triad5100	Retinoid-inducible serine	Viridiplantae	Proliferation repression
	carboxypeptidase		
Triad5102	Predicted protein	Viridiplantae	Unknown
Triad6614	Selenium-binding protein 1-A	Viridiplantae	Respond to xenobiotics

Table 6. GO analyses of 43 HGT genes using PANTHER classification system.

Gene ID	PANTHER Family/Subfamily	GO database MF	GO database BP	GO database CC Complete
Triad8071	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810); ion transport(GO:0006811);potassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane transport(GO:0071805)	membrane(GO:0016020);integral component of membrane(GO:0016021)

Triad6221	NO VEIN-LIKE PROTEIN-RELATED (PTHR32387:SF0)		
Triad1567	SUBFAMILY NOT NAMED (PTHR34847:SF1)	catalytic activity(GO:0003824)	biosynthetic process(GO:0009058)
Triad9802	BETA CARBONIC ANHYDRASE 1 (PTHR11002:SF13)	carbonate dehydratase activity(GO:0004089);zinc ion binding(GO:0008270)	metabolic process(GO:0008152)
Triad8266	PROTEIN ROADKILL (PTHR24413:SF77)		
Triad6614	SELENIUM-BINDING PROTEIN 1 (PTHR23300:SF0)	selenium binding(GO:0008430)	
Triad9800	PROTEIN ZK829.7 (PTHR11695:SF553)	zinc ion binding(GO:0008270);oxidoreductase activity(GO:0016491);metal ion binding(GO:0046872)	oxidation-reduction process(GO:0055114)

Triad8075	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810);ion transport(GO:0006811);potassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane transport(GO:0071805)	membrane(GO:0016020);integral component of membrane(GO:0016021)
Triad6214	NO VEIN-LIKE PROTEIN-RELATED (PTHR32387:SF0)			
Triad10012	ESTERASE YJFP (PTHR10992:SF732)			
Triad7572	GLYCOSYLTRANSFERASE			

25 FAMILY MEMBER

(PTHR10730:SF27)

Triad7682	ABSCISIC ACID 8'-HYDROXYLASE 4 (PTHR24286:SF77)	monooxygenase activity(GO:0004497);iron ion binding(GO:0005506);oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen(GO:0016705);heme binding(GO:0020037);metal ion binding(GO:0046872)	oxidation-reduction process(GO:0055114)
Triad7482	RETINOID-INDUCIBLE SERINE CARBOXYPEPTIDASE (PTHR11802:SF3)	serine-type carboxypeptidase activity(GO:0004185)	proteolysis(GO:0006508)
Triad7481	RETINOID-INDUCIBLE	serine-type carboxypeptidase	proteolysis(GO:0006508)

	SERINE CARBOXYPEPTIDASE (PTHR11802:SF3)	activity(GO:0004185)		
Triad8145	SUBFAMILY NOT NAMED (PTHR39217:SF1)	glutathione synthase activity(GO:0004363);ATP binding(GO:0005524)	glutathione biosynthetic process(GO:0006750)	
Triad8146	SUBFAMILY NOT NAMED (PTHR39217:SF1)	glutathione synthase activity(GO:0004363);ATP binding(GO:0005524)	glutathione biosynthetic process(GO:0006750)	
Triad8078	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810);ion transport(GO:0006811);potassium ion transport(GO:0006813);regulation of ion transmembrane	membrane(GO:0016020);integral component of membrane(GO:0016021)

			transport(GO:0034765);po tassium ion transmembrane transport(GO:0071805)	
Triad2364	SUBFAMILY NOT NAMED (PTHR10672:SF21)			
Triad10098	43 KDA RECEPTOR-ASSOCIATED PROTEIN OF THE SYNAPSE (PTHR10098:SF126)		regulation of mitotic cell cycle(GO:0007346)	centrosome(GO:0005813) ;midbody(GO: 0030496)
Triad10384	PHENAZINE BIOSYNTHESIS-LIKE DOMAIN-CONTAINING PROTEIN (PTHR13774:SF17)	catalytic activity(GO:0003824)	biosynthetic process(GO:0009058)	
Triad10385	PHENAZINE	catalytic activity(GO:0003824)	biosynthetic	

	BIOSYNTHESIS-LIKE		process(GO:0009058)
	DOMAIN-CONTAINING		
	PROTEIN		
	(PTHR13774:SF17)		
Triad11269	PHENAZINE	catalytic activity(GO:0003824)	biosynthetic
	BIOSYNTHESIS-LIKE		process(GO:0009058)
	DOMAIN-CONTAINING		
	PROTEIN		
	(PTHR13774:SF17)		
Triad8318	PHENAZINE	catalytic activity(GO:0003824)	biosynthetic
	BIOSYNTHESIS-LIKE		process(GO:0009058)
	DOMAIN-CONTAINING		
	PROTEIN		
	(PTHR13774:SF17)		
Triad5100	RETINOID-INDUCIBLE	serine-type carboxypeptidase	proteolysis(GO:0006508)
	SERINE	activity(GO:0004185)	

CARBOXYPEPTIDASE

(PTHR11802:SF3)

Triad6224 NO VEIN-LIKE PROTEIN-

RELATED (PTHR32387:SF0)

Triad1431 SUBFAMILY NOT NAMED

(PTHR16222:SF12)

| Triad8072 | FI16807P1 (PTHR11767:SF58) | inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244) | transport(GO:0006810);ion transport(GO:0006811);potassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane | membrane(GO:0016020);integral component of membrane(GO:0016021) |

transport(GO:0071805)

Triad8076	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810);ion transport(GO:0006811);potassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane transport(GO:0071805)	membrane(GO:0016020);integral component of membrane(GO:0016021)
Triad8077	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810);ion transport(GO:0006811);po	membrane(GO:0016020);integral

		tassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane transport(GO:0071805)	component of membrane(GO:0016021)
Triad6223	NO VEIN-LIKE PROTEIN-RELATED (PTHR32387:SF0)		
Triad6222	NO VEIN-LIKE PROTEIN-RELATED (PTHR32387:SF0)		
Triad7683	ABSCISIC ACID 8'-HYDROXYLASE 4 (PTHR24286:SF77)	iron ion binding(GO:0005506);oxidoreductase activity, acting on paired donors, with incorporation or reduction of	oxidation-reduction process(GO:0055114)

molecular oxygen(GO:0016705);heme binding(GO:0020037)

Triad5102	RETINOID-INDUCIBLE SERINE CARBOXYPEPTIDASE (PTHR11802:SF3)	serine-type carboxypeptidase activity(GO:0004185)	proteolysis(GO:0006508)
Triad5587	AMIDOHYDROLASE FAMILY PROTEIN (PTHR22642:SF2)	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds(GO:0016810)	metabolic process(GO:0008152)
Triad5585	AMIDOHYDROLASE FAMILY PROTEIN (PTHR22642:SF2)	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds(GO:0016810)	metabolic process(GO:0008152)
Triad1430	SUBFAMILY NOT NAMED (PTHR16222:SF12)		

| Triad476 | PEPTIDYL-GLYCINE ALPHA-AMIDATING MONOOXYGENASE (PTHR10680:SF14) | catalytic activity(GO:0003824);monooxygenase activity(GO:0004497);peptidylglycine monooxygenase activity(GO:0004504);copper ion binding(GO:0005507);oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced ascorbate as one donor, and incorporation of one atom of oxygen(GO:0016715) | peptide metabolic process(GO:0006518);oxidation-reduction process(GO:0055114) | membrane(GO:0016020) |
| Triad8202 | FERREDOXIN-FOLD ANTICODON-BINDING DOMAIN-CONTAINING | | | |

	PROTEIN 1 (PTHR11538:SF26)			
Triad8243	PROTEIN ROADKILL (PTHR24413:SF77)			
Triad7945	CATION CHANNEL SPERM-ASSOCIATED PROTEIN 3 (PTHR10037:SF196)	ion channel activity(GO:0005216);voltage-gated ion channel activity(GO:0005244);voltage-gated calcium channel activity(GO:0005245);calcium channel activity(GO:0005262)	transport(GO:0006810);ion transport(GO:0006811);calcium ion transport(GO:0006816);regulation of ion transmembrane transport(GO:0034765);transmembrane transport(GO:0055085);calcium ion transmembrane transport(GO:0070588)	voltage-gated calcium channel complex(GO:0005891);membrane(GO:0016020);integral component of membrane(GO:0016021)

Triad7684	ABSCISIC ACID 8'-HYDROXYLASE 4 (PTHR24286:SF77)	monooxygenase activity(GO:0004497);iron ion binding(GO:0005506);oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen(GO:0016705);heme binding(GO:0020037);metal ion binding(GO:0046872)	oxidation-reduction process(GO:0055114)	
Triad5101	RETINOID-INDUCIBLE SERINE CARBOXYPEPTIDASE (PTHR11802:SF3)	serine-type carboxypeptidase activity(GO:0004185)	proteolysis(GO:0006508)	
Triad8073	FI16807P1 (PTHR11767:SF58)	inward rectifier potassium channel activity(GO:0005242);voltage-gated ion channel activity(GO:0005244)	transport(GO:0006810);ion transport(GO:0006811);po	membrane(GO:0016020);integral

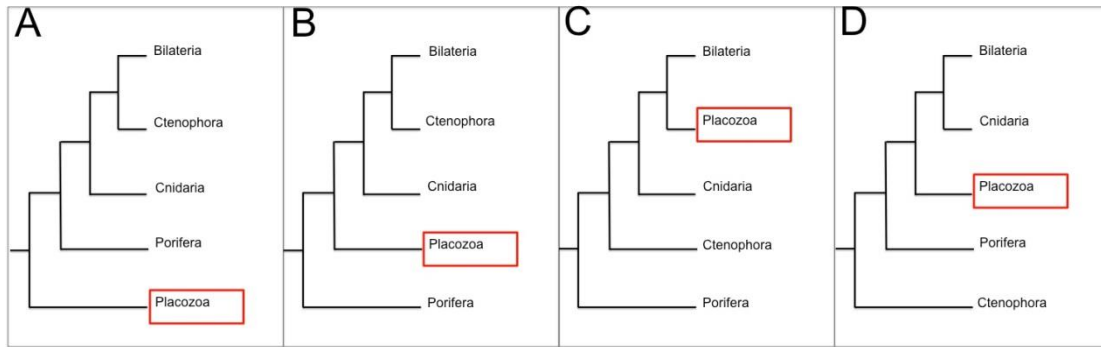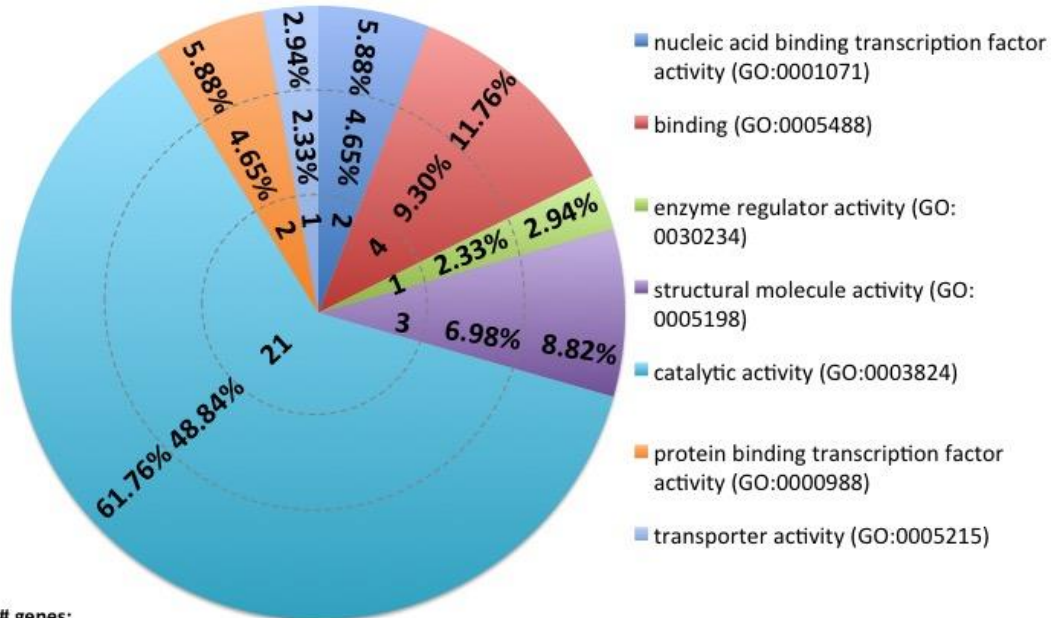| tassium ion transport(GO:0006813);regulation of ion transmembrane transport(GO:0034765);potassium ion transmembrane transport(GO:0071805) | component of membrane(GO:0016021) |

Fig. 6. Four hypotheses on the phylogenetic relationship of placozoans to other

metazoans.

A

GO Molecular Function

Total # Genes: 43; Total # function hits: 34

Legend:
- nucleic acid binding transcription factor activity (GO:0001071)
- binding (GO:0005488)
- enzyme regulator activity (GO:0030234)
- structural molecule activity (GO:0005198)
- catalytic activity (GO:0003824)
- protein binding transcription factor activity (GO:0000988)
- transporter activity (GO:0005215)

Pie chart values: 5.88%, 4.65%, 2 (nucleic acid binding transcription factor activity); 11.76%, 9.30%, 4 (binding); 2.94%, 2.33%, 1 (enzyme regulator activity); 8.82%, 6.98%, 3 (structural molecule activity); 61.76%, 48.84%, 21 (catalytic activity); 5.88%, 4.65%, 2 (protein binding transcription factor activity); 2.94%, 2.33%, 1 (transporter activity)

genes;
Percent of gene hit against total # genes;
Percent of gene hit against total # Function hits
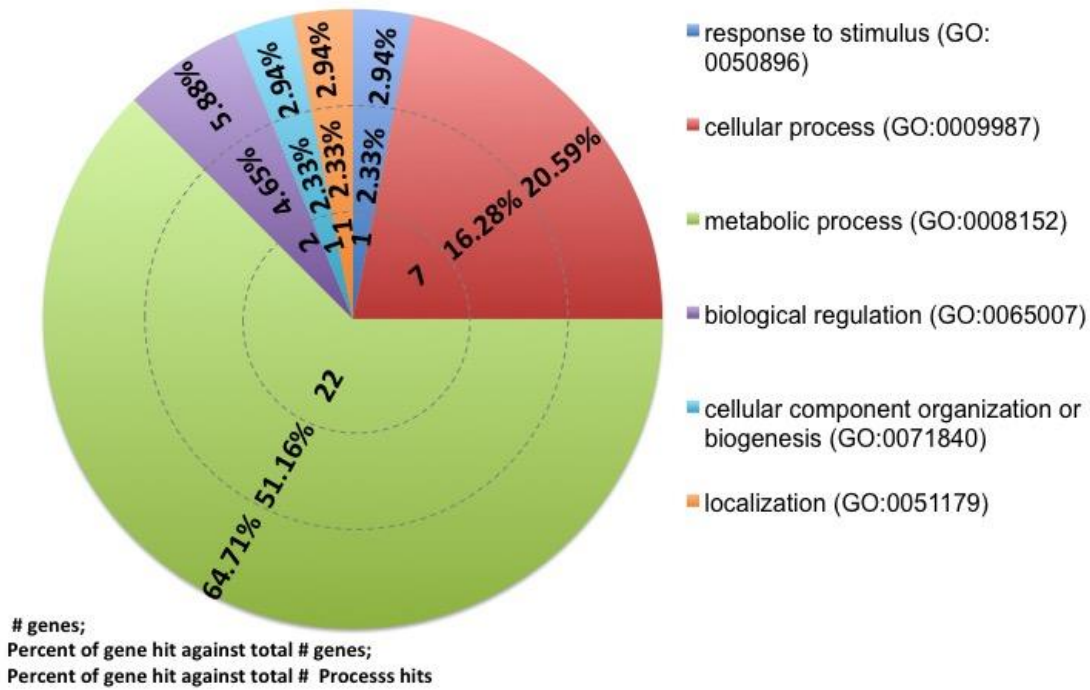
B

Fig. 7. For both pie charts, the inner ring represents number of genes in that category; middle ring represents percentage over total number of genes; outer ring represents percentage over total number of functional hits. GO analyses used PATHER database. In "Molecular Function" analysis, more than half of the genes are involved in "catalytic activity" (61.76%). In "Biological Process" analysis, the largest category is "metabolic process" (58.82%).
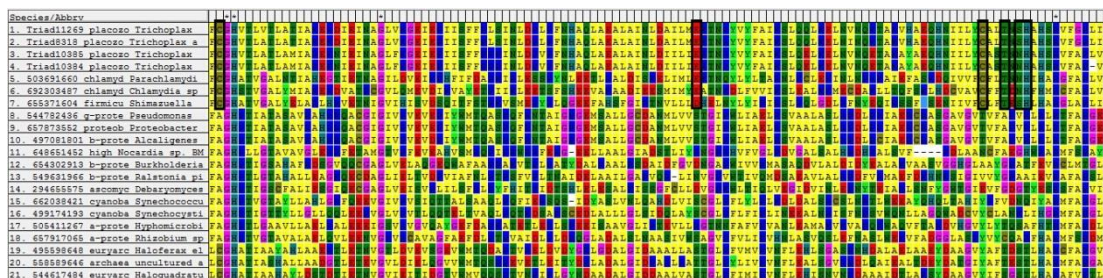
A
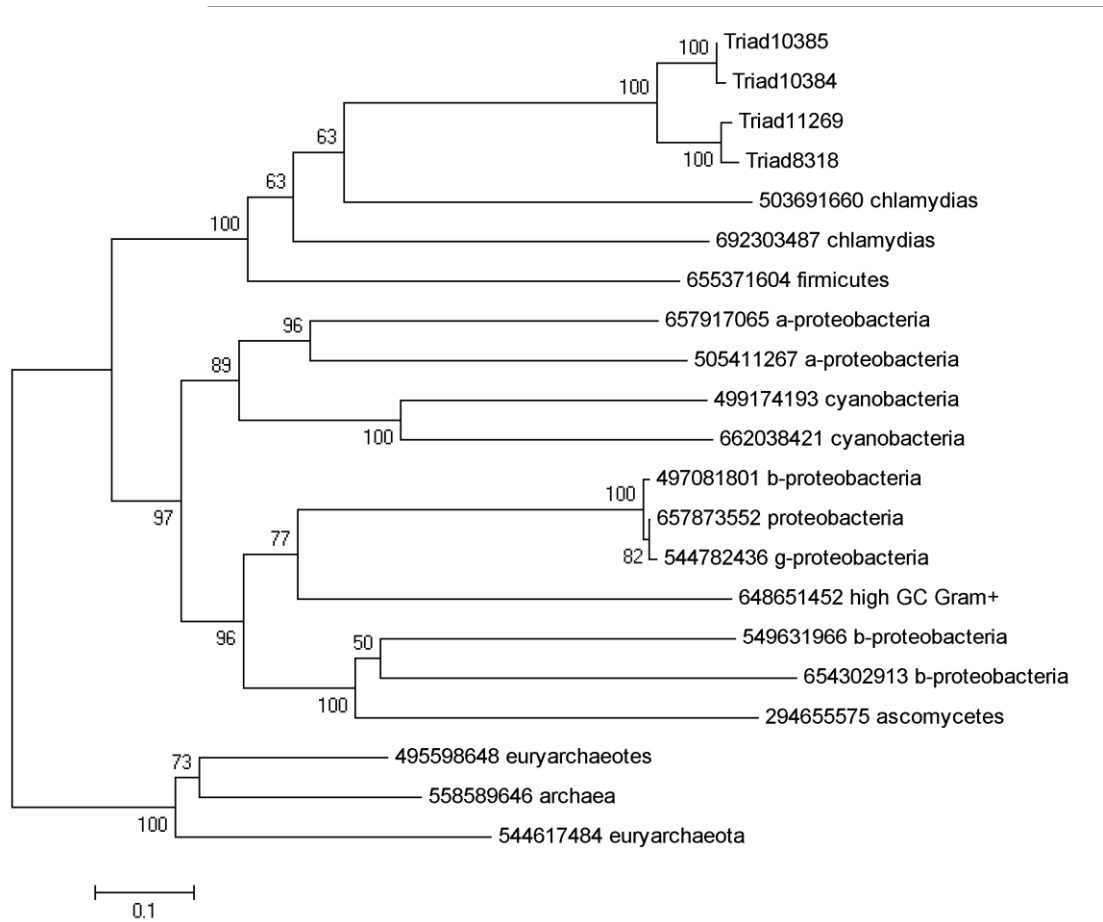
B

Fig. 8. Molecular phylogeny (A) and multiple sequence alignment (B) of phenazine biosynthesis-like domain-containing protein (PBLD). Boxed columns indicate the amino-acid residues shared by bacterial and *T. adhaerens* PBLD sequences.

Numbers above branches are bootstrap values from maximum likelihood method.

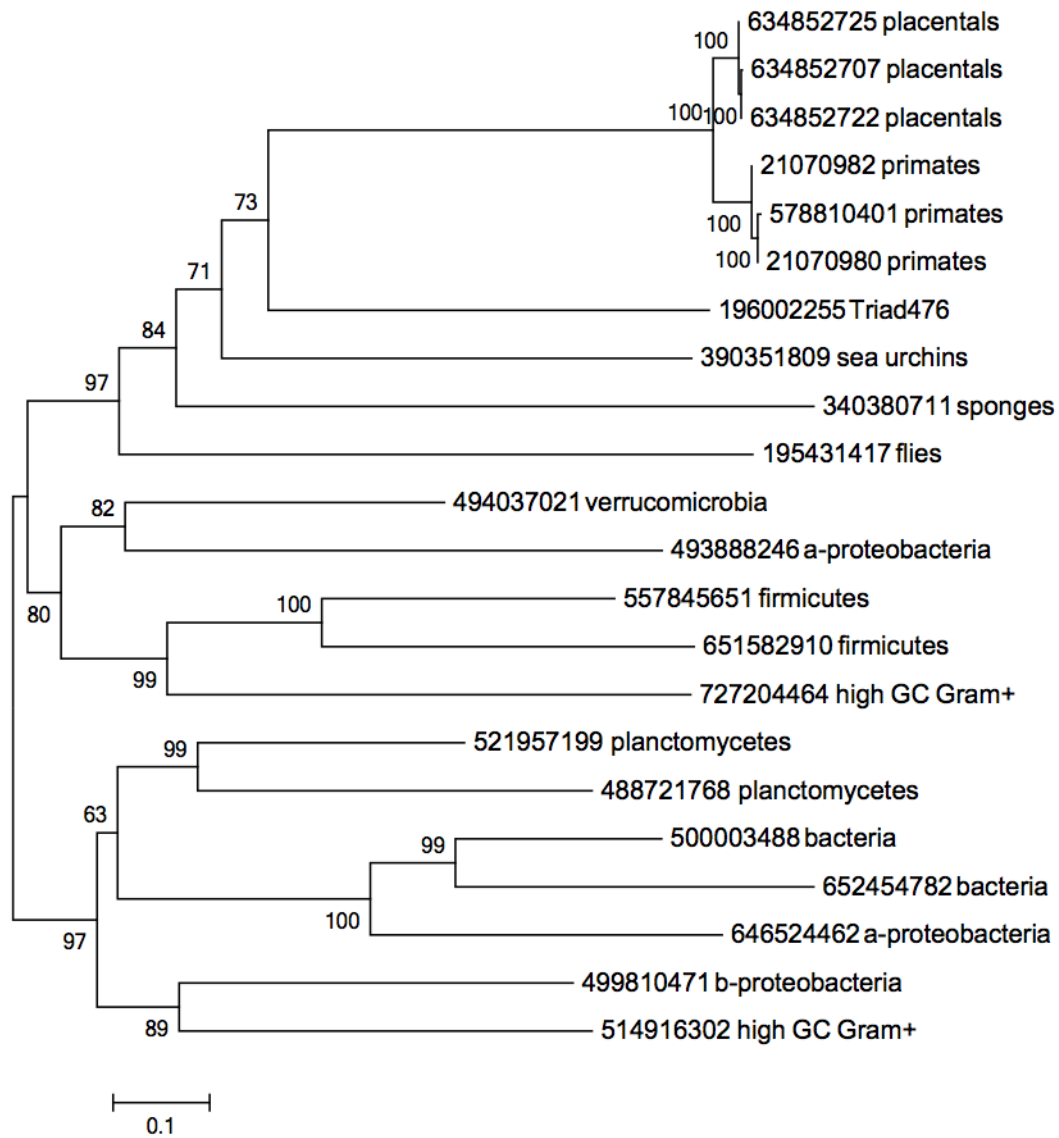Bootstrap values that were lower than 50 were deleted from the tree.

Fig. 9. Molecular phylogeny of peptidylglycine alpha-amidating monooxygenase (PAM). Numbers above branches are bootstrap values from maximum likelihood method. Bootstrap values that were less than 50 were deleted from the tree.

Chapter 5: Conclusions

This study is one of the first researches to explore the HGT scale in all metazoans, and can be divided into three parts: scales of HGT, the occurrence of HGT in *Ciona* and *Trichoplax*. Comparing with a recent study by Crisp A. et al. (Crisp, Boschetti et al. 2015), chapter two of my study includes fewer genomes/transcriptomes, but cover all the major linages of metazoans, rather than just model species such as nematodes or fruit flies. Sampled species include aquatic and terrestrial animals, early-branching and "higher" animals, animals of sexual and asexual reproduction, parasites and non-parasites, which make it possible to analyze factors that may affect HGT occurrence. Another advantage of chapter two is the filtration algorithm AlienG. It has been demonstrated to be computationally efficient, and can identify more HGT genes with fewer false positive candidates, compared with other programs, such as PhyloGenie and Darkhorse (Tian, Sun et al. 2011). However, the study of HGT scales in chapter two does have some pitfalls. HGT candidates, instead of verified HGT genes, were used to analyze the importance of different factors. Plenty of predicted candidates may be false positives, so further study on these HGT candidates is needed. Another potential pitfall of this study is the limited species sampling, especially parasitic species. Further study with more parasitic genomes will help uncover whether parasitism enhances or restrains HGT events.

Chapter three scrutinizes the HGT candidates of *C. intestinalis* obtained from AlienG filtration. Due to lacking of annotation of HGT genes, it is difficult to conclude on the roles of the acquired genes in evolution of metazoans or in the innovation of

tunicates. Gene functions and their evolutionary importance of validated HGT genes need to be explored. An interesting case is *CesA* genes found in tunicates, which cannot be detected in other metazoan species. Previous research suggested that tunicate *CesA* genes are derived from bacteria. However, my data clearly show that they are derived from green algae instead. With discoveries of evolution of *Ces/Csl* genes in green plants, chapter three helps to uncover the origin and evolution of *Ces/Csl* genes in nature.

In chapter four, two interesting genes, *PBLD* and *PAM*, were discussed specifically. *PBLD* encodes a key enzyme involved in the biosynthesis of phenazine, an antibiotic used by some bacteria. Whether this gene is functional in *Trichoplax* needs to be clarified. If it does express and has functions in placozoans, it represents an evolutionary innovation of this clade. *PAM* is a gene usually found in Bilateria and Cnidaria with nervous systems. However, it is also present in *Trichoplax*, which does not have a nervous system, and serves as an epitheliopeptide amidase. This discovery is consistent with previously proposed "two-step" model to explain the evolution of nervous system in metazoans(Galliot and Quiquand 2011), which states that neural components consist of genes recruited before and after the emergence of neurogenesis. Since *PAM* cannot be found in *Ctenophora*, it implies that *Ctenophora* may be the earliest-branching clade in metazoans, which is also supported by other recent studies (Dunn, Hejnol et al. 2008; Hejnol, Obst et al. 2009; Ryan, Pang et al. 2013; Moroz 2014; Whelan, Kocot et al. 2015).

This dissertation is definitely not a systematic study of HGT events in all metazoans, but its generated data pave the way for future research. Additional studies with more metazoan species are necessary, especially with more parasites selected. All HGT candidates need to be verified using rigorous phylogenetic analyses and other methods, so that relationships between HGT occurrence and different factors can be demonstrated with more confidence. Wet-lab experiments need to be done to investigate the functions of acquired genes, in order to better understand their role in the development and evolution of different animals.

References

Crisp, A., C. Boschetti, et al. (2015). "Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes." Genome Biol **16**: 50.

Dunn, C. W., A. Hejnol, et al. (2008). "Broad phylogenomic sampling improves resolution of the animal tree of life." Nature **452**(7188): 745-749.

Galliot, B. and M. Quiquand (2011). "A two-step process in the emergence of neurogenesis." Eur J Neurosci **34**(6): 847-862.

Hejnol, A., M. Obst, et al. (2009). "Assessing the root of bilaterian animals with scalable phylogenomic methods." Proc Biol Sci **276**(1677): 4261-4270.

Moroz, L. L. (2014). "The genealogy of genealogy of neurons." Commun Integr Biol **7**(6): e993269.

Ryan, J. F., K. Pang, et al. (2013). "The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution." Science **342**(6164): 1242592.

Tian, J., G. Sun, et al. (2011). AlienG: An Effective Computational Tool for Phylogenomic Identification of Horizontally Transferred Genes. BICoB.

Whelan, N. V., K. M. Kocot, et al. (2015). "Error, signal, and the placement of Ctenophora sister to all other animals." Proc Natl Acad Sci U S A **112**(18): 5773-5778.