Phylogenetic characterization of transporter proteins in the cnidarian-dinoflagellate symbiosis


ARTICLE INFO

Keywords:
Glucose
Glycerol
Nitrogen
Coral
Membrane transport
Bioinformatics

ABSTRACT

Metabolic exchange between cnidarians and their symbiotic dinoflagellates is central to maintaining their mutualistic relationship. Sugars are translocated to the host, while ammonium and nitrate are utilized by the dinoflagellates (Symbiodinium spp.). We investigated membrane protein sequences of each partner to identify potential transporter proteins that move sugars into cnidarian cells and nitrogen products into Symbiodinium cells. We examined the facilitated glucose transporters (GLUT), sodium/glucose cotransporters (SGLT), and aquaporin (AQP) channels in the cnidarian host as mechanisms for sugar uptake, and the ammonium and high-affinity nitrate transporters (AMT and NRT2, respectively) in the algal symbiont as mechanisms for nitrogen uptake. Homologous protein sequences were used for phylogenetic analysis and tertiary structure deductions. In cnidarians, we identified putative glucose transporters of the GLUT family and glycerol transporting AQP proteins, as well as sodium monocarboxylate transporters and sodium myo-inositol cotransporters homologous to SGLT proteins. We hypothesize that cnidarians use GLUT proteins as the primary mechanism for glucose uptake, while glycerol moves into cells by passive diffusion. We also identified putative AMT proteins in several Symbiodinium clades and putative NRT2 proteins only in a single clade. We further observed an upregulation of expressed putative AMT proteins in Symbiodinium, which may have emerged as an adaptation to conditions experienced inside the host cell. This study is the first to identify transporter sequences from a diversity of cnidarian species and Symbiodinium clades, which will be useful for future experimental analyses of the host-symbiont proteome and the nutritional exchange of Symbiodinium cells in hospite.

1 Introduction

The cnidarian-dinoflagellate symbiosis is an endosymbiotic relationship involving intercellular exchange of many compounds. Dinoflagellates of the genus Symbiodinium reside within cnidarian endodermal cells and provide the host with the majority of its fixed carbon as photosynthetic products, with sugars being the major trafficked metabolites (Davy et al., 2012; Gordon and Leggat, 2010; Kopp et al., 2015). In exchange, the cnidarian provides nitrogenous compounds, generated as by-products from heterotrophic feeding, to the algal symbionts (Allemand et al., 1998; Davy et al., 2012; Houbrèque et al., 2004). Successful metabolite exchange between the host and symbiont is key to the health and survival of reef-building corals, as it allows the partners to obtain necessary nutrients that are often limiting in oligotrophic oceans (Muscatine et al., 1981; O’Neil and Capone, 2008).

Characterization of transported metabolites has long been a focus of cnidarian-dinoflagellate research. Glucose is now considered to be the primary carbon product translocated to the host from the symbiont (Burriscig et al., 2012; Hillery et al., 2016), although evidence from early studies pointed to glycerol (Lewis and Smith, 1971; Muscatine, 1967; Schmitz and Kremer, 1977). It is now thought that glycerol is primarily released by Symbiodinium cells in isolation (e.g. in culture), while glucose is the major photosynthetic product translocated from the symbiont to the host in hospite (Burriscig et al., 2012; Hillery et al., 2016; Ishikura et al., 1999; Whitehead and Douglas, 2003). High levels of glycerol have also been found to be released as an osmolyte in response to stress (Suescun-Bolivar et al., 2012). In the opposite direction, the major nitrogen metabolite transported to the symbiont by the host is thought to be ammonium, while nitrate is an important nutrient...
from the seawater or the host is used by the algal symbionts for photosynthesis. Carbon is obtained from seawater or the host waste products, then translocated to the symbiont through integral membrane proteins in the symbiosome membrane and symbiont cell membrane. The algae use the ammonium and nitrate for amino acid synthesis, with a portion of amino acids being translocated back to the host.

Although these carbon and nitrogen metabolites are an important source of energy and nutrition in the symbiosis, the transport mechanisms are still poorly understood. All molecules entering or leaving a cell must traverse the plasma membrane, most either by active transport or passive diffusion through transporter or channel proteins (Lodish et al., 2000). In the cnidarian-dinoflagellate symbiosis, Symbiodinium cells are housed within a host-derived vacuole called the symbiosome, which acts as the interface between host and symbiont (Fitt and Trench, 1983; Wakefield et al., 2000; Wakefield and Kempf, 2001). Therefore, the exchange of compounds must be facilitated by proteins in the cellular membranes of each of the partners as well as in the symbiosome interface (Fig. 1).

Insights on nutrient transport can be gained from studies of other systems. Glucose primarily enters eukaryotic cells through transporters in two families: facilitated glucose transporters (GLUT) and sodium-coupled glucose transporters (SGLT) (Scheepers et al., 2015; Wood and Trayhurn, 2003). Members of both families have been extensively characterized in humans and other animals (Doeege et al., 2001, 2000; Escher and Rasmussen-LeStander, 1999; Joost et al., 2002; Wright, 2013). While few invertebrate GLUT and SGLT homologs have been described (Escher and Rasmussen-LeStander, 1999; Peng et al., 2010) and transcriptomics (Lehnert et al., 2014) to identify proteins involved in the symbiosis; however, transporters specific to the symbiosome membrane have not been firmly established. Glycerol can be trafficked by aquaporin (AQP) channels, which are intrinsic water channels that transport small molecules (Hara-Chikuma and Verkman, 2006; Rojek et al., 2008). Since most eukaryotes already use AQP to facilitate the diffusion of water, we predict that cnidarians likely have AQP members capable of glycerol transport as well.

Currently, three genes encoding ammonium transporters (AMT proteins) and two encoding high-affinity nitrate transporters (NRT proteins) have been identified in Symbiodinium transcriptomes (Brown et al., 2011; Mayfield et al., 2010; Van Den Heuvel, 2012). The recent genome assembly of Symbiodinium kawagutii (Clade F) has also revealed genes encoding AMT and NRT family members (Lin et al., 2015). However, these studies have only focused on a single Symbiodinium type and did not consider the large genetic diversity within the genus. Currently, the Symbiodinium genus is classified into nine divergent clades (A–I), with each (except clade E) further divided into sub-clades or “types” (e.g., A3, B1, D1a) and some described species (Colfroth and Santos, 2005; Lajeunesse et al., 2012; Pochon and Gates, 2010). Genetic distances between these clades were found to be greater than the distances separating different species of free-living dinoflagellates (Krueger et al., 2015; Sampayo et al., 2009; Santos et al., 2002; Takabayashi et al., 2004). Therefore, analysis of these proteins from a diversity of Symbiodinium types may reveal differences in their conservation.

The goal of this study was to identify and characterize potential membrane transport mechanisms employed by both partners of the cnidarian-dinoflagellate symbiosis by performing informatics/
phylelogenetic investigations of well-described sugar and nitrogen transport proteins that are highly conserved among eukaryotes. We investigated five protein families: three families of secondary active transporters, GLUT, SGLT, and NRT2, and two families of channel homologs, AQP and AMT. We used a phylogenetic approach to identify cnidarian homologs to GLUT, SGLT, and AQP, and Symbiodinium homologs to AMT and NRT2. We then performed protein structural predictions, conservation analyses, and structural domain identifications to help establish whether or not these proteins function in metabolite exchange within the cnidarian-dinoflagellate symbiosis. Insights gained from these analyses will help inform further functional investigations of the symbiosome membrane and enhance our understanding of nutrient exchange in the cnidarian-dinoflagellate symbiosis.

2. Materials and methods

2.1. Identification of cnidarian and Symbiodinium transporter proteins

Separate analyses were completed to categorize the major transporters of the cnidarian host and Symbiodinium symbiont. Cnidarian and Symbiodinium transcriptomes and genomes were downloaded from publicly available sources (Table A1), and each was formatted into a searchable database. Reference sequences for each protein family were obtained from UniProt (Table A2); these included Homo sapiens GLUT, SGLT, and AQP, and Arabidopsis thaliana AMT and NRT2 (Table A2). Putative orthologues to the reference sequences were identified in the cnidarian and Symbiodinium databases by reciprocal BLAST (Telford, 2007; Voolstra et al., 2011) with a bit-score cutoff of 80 for any given alignment. To identify cnidarian sugar transporters, H. sapiens reference proteins (Table A2) were used as BLAST input queries to search cnidarian databases; A. thaliana reference proteins (Table A2) were similarly used as BLAST input queries for searching the Symbiodinium databases for nitrate and ammonium transporters. The best scoring matches were then converted to predicted proteins using custom Perl scripts (obtained from Eli Meyer, https://github.com/Eli-Meyer).

2.2. Retrieval of protein homologs

Sequences from the cnidarian and Symbiodinium databases with high sequence similarity to the reference proteins were searched and used for phylelogenetic analyses. Those protein sequences were also used to search with the NCBI non-redundant (nr) database using BLASTp, retrieving sequences with an E value cutoff of ≤ 1 × 10^-5 and excluding hypothetical proteins (Datasheets A1–A5). The top five annotated BLAST hits for each sequence were used as additional homologs in the phylelogenetic analyses (Datasheets A1–A5).

2.3. Sequence alignment and phylogenetic analysis

Protein family sequences were aligned using MAFFT v. 7.017 (default parameters) (Katoh and Standley, 2013). Alignments were manually edited in Aliview (Larsson, 2014) to eliminate false positive matches (Bucka-Lassen et al., 1999) by removing protein fragments (sequences of < 150 amino acids) and sequences that did not align with any of the defined motifs in Figs. S1–S5: the GLUT GR, GRR, and PETK motifs (Fig. S1), all of which are considered necessary for sugar transport function (Joost and Thorens, 2001); the SGLT Y, GG, and D motifs (Fig. S2), which are the most conserved (≥ 77%) between all pro- and eukaryotic taxa (Wright and Turk, 2004); the two NPA motifs and one D residue of AQP (Fig. S3), since NPA motifs form the aqueous pore and represent the AQP signature while the D residue is present in aquaglyceroporins and widens the pore opening allowing the transfer of larger solutes (Agre et al., 2002); the AMT D, AG, VH, and G motifs (Fig. S4), which are responsible for ammonium recognition and transport function (Marini et al., 2006); the NRT2 conserved F residue along with the proposed NNP signature A/G-G-W-G/A-N/D-M/L-G and the structurally important GG and GAGG glycine clusters (Fig. S5) (Forde, 2000; Trueeman et al., 1996).

Phylelogenetic analysis of each protein family was conducted using maximum-likelihood inference in RAxML v8.0.26 (Stamatakis, 2014). The PROTGAMMAUTO function was used to empirically (highest likelihood score of starting parsimony tree) determine the best substitution model for each dataset; all datasets in this study were best suited to using the AUTO model. The topology with the highest likelihood score for each protein tree was selected from 100 replicates, and then branch node support was calculated using 100 bootstraps. Protein homologs from the most ancestral species were used to root each tree. Phylelogenetic inference was then used to draw conclusions on the protein identities of novel cnidarian and Symbiodinium sequences (Eisen, 1998; Eisen et al., 1997; Sjölander, 2004; Zmasek and Eddy, 2001).

2.4. Structural predictions and characterization of proteins

Additional sequence analyses were conducted to further characterize the identified cnidarian and Symbiodinium proteins. The number of transmembrane (TM) helices and subcellular localization for each protein were predicted using the MemPye webserver, which outperforms other localization methods with an accuracy rate of 70% (Pierleoni et al., 2011), and protein domains were determined using the InterProScan (Quevillon et al., 2005) in Geneious v. 7.1.9. One protein sequence per gene family was also selected for protein structural analysis. Nemastostella vectensis sequences with highest percent similarity to the reference sequences were used for analysis of the cnidarian proteins because they were closest in amino acid length and had the highest similarity to the reference proteins. Furthermore, the N. vectensis genome had the most complete annotations. A sequence from Symbiodinium C was used for AMT structural analysis to further examine the high AMT protein copies found in this symbiont type and because at least one AMT sequence from this type was grouped with each eukaryotic clade (Fig. 6, clades 2–5), while a sequence from Symbiodinium clade A was chosen for NRT2 structural analysis since it was the only clade with sequences retrieved for this protein family. The three-dimensional structure and protein function were predicted using I-TASSER (Roy et al., 2010; Zhang, 2008). I-TASSER aligns predictions with the most similar protein structures from the Protein Data Bank (http://www.rcsb.org/pdb) and evaluates the structural similarity using a TM-score ranging from 0 to 1, where 1 indicates a perfect match. Ligands and ligand-binding sites are also predicted as part of I-TASSER using the COACH algorithm (Yang et al., 2013), which evaluates the confidence of the prediction using a C-score ranging from 0 to 1, where 1 indicates 100% confidence in the prediction. To determine the degree of structural conservation within each cnidarian and Symbiodinium protein family, all newly identified sequences were aligned within their family using MAFFT v. 7.017 and the alignment was submitted to the ConSurf server (Ashkenazy et al., 2010) along with the corresponding predicted 3D structure as a template to display residue conservation. Conservation scores for each residue were calculated using the Bayesian method with the best evolutionary substitution model determined empirically as WAG (Mayrose et al., 2004), and scores were subsequently depicted onto the protein structures using Polyview 3D (Porollo and Meller, 2007).

3. Results and discussion

3.1. Cnidarian GLUT proteins

A total of 23 cnidarian sequences were identified as GLUT homologs, representing putative glucose transporters (Datasheet A1). Sequence lengths were highly variable, with some sequences containing fewer than the 12 TM domains that are typical of proteins within the major facilitator superfamily (MFS), which includes GLUT proteins.
Fig. 2. Phylogenetic tree of cnidarian GLUT sequences and closest homologs from other organisms. The tree was rooted by the yeast *S. cerevisiae* homolog. The number of amino acid substitutions is indicated by the scale bar, and bootstrap support for nodes > 50% is represented by shaded circles. Cnidarian sequences are shown in bold text. Four main clades emerged: (1) OCT; (2) Class 1 and 2 GLUT proteins; (3) GLUT6 and GLUT8; and (4) GLUT10, GLUT12, and HMIT. Clades containing only vertebrate sequences are collapsed to enhance tree clarity.

The sequence used for structural analysis is marked by a yellow star (*N. vectensis* _4 GLUT8).
These sequences may represent incomplete proteins that were not excluded from our analysis because they contain the conserved protein motifs (Datasheet A1). Only cnidarian sequences from GLUT classes 1 and 2 were identified. All sequences include domains for MFS sugar transporters, while some also contain domains associated with specific GLUT members.

Our phylogenetic analysis of the GLUT proteins resolved the sequences into four main clades of solute transporters (Fig. 2): (1) organic cation transporters (OCT); (2) class 1 and 2 GLUT proteins; (3) GLUT6 and GLUT8; and (4) GLUT10, GLUT12, and HMIT. One divergent protein from Porites australiensis (Fig. 2) branches from near the root of the tree rather than clustering with any of the four main clades. This protein shares the highest sequence similarity with the yeast glucose transporter RGT2 (Datasheet A1). Since the P. australiensis transcriptome was generated from RNA derived from the entire holobiont rather than just host material, it is likely that this deduced protein represents RNA contamination from the algal symbiont or bacteria (Shinzato et al., 2014). Robust support is only observed for the OCT clade, with the other major clades being less well resolved. OCT is another MFS protein family that is closely related to the sugar porter family (Saier et al. 1999) (Fig. 2, clade 1). These proteins facilitate diffusion of organic cations such as neurotransmitters, amino and fatty acids, and nucleotides in vertebrates, but they have not been well studied in invertebrate systems (Clarimbol, 2008; Eraly et al., 2004).

The clustering of the GLUT sequences resulted in a branching order that follows previously constructed metazoan GLUT phylogenies (Wilson-O’Brien et al., 2010). Class 3 GLUT proteins are ancestral to the other GLUTs, and they have diverged into three well supported sub-clades (GLUT6 and GLUT8, GLUT10 and GLUT12, and HMIT), while class 1 and 2 proteins form a well-supported clade as sister groups that evolved more recently (Scheepers et al., 2015; Wilson-O’Brien et al., 2010). Overall, we identified cnidarian homologs to all class 1 and class 3 GLUT members, except for GLUT6 (Datasheet A1).

Ten cnidarian putative GLUT8 proteins were identified (Fig. 2). GLUT8 is of particular interest in the cnidarian-dinoflagellate symbiosis because it has a high affinity for glucose (Uldry and Thorens, 2004) and exhibits increased mRNA levels in symbiotic A. pallida relative to those in aposymbiotic anemones of the same species (Lehnert et al., 2014). As it is localized to endosomal compartments in mice, GLUT8 is predicted to function as a hexose transporter in intracellular membranes (Augustin et al., 2005). Four of the identified cnidarian GLUT8 proteins in this study were predicted to be localized to internal membranes, while six were predicted to exist in the plasma membrane (Datasheet A1). Internal GLUT proteins may reside in the symbiosome membrane, which is an internal vacuolar membrane of host origin that is derived from an early endosome (Fitt and Trench, 1983; Rands et al., 1993; Wakefield and Kempf, 2001). Three copies of GLUT8 were found in N. vectensis and two in P. australiensis. All copies in the non-symbiotic cnidarian N. vectensis were predicted to be localized to the plasma membrane. The two copies of GLUT8 in the symbiotic cnidarian, P. australiensis, had different predicted localizations based on the Mem-Loci algorithm; one is likely in the plasma membrane while the other appears to reside in internal membranes. These analyses raise the possibility that symbiotic cnidarians have a distinct GLUT8 that is associated with the symbiosome membrane where it facilitates movement of glucose from the symbiont to the host. The protein predicted to be internally located does form a highly supported group with one of the N. vectensis cell membrane proteins. Perhaps the internally localized P. australiensis protein is one of the most ancestral forms of the symbio-some-associated transporter as it still maintains high similarity to those in the cell membrane of apsymbiotic animals, or perhaps this particular isoform can be expressed in either the cell or internal membranes since prediction probabilities were over 50% for both (Datasheet A1).

The deduced 3D structure of the cnidarian GLUT8 protein closely resembles the crystal structure of human GLUT1 (TM score = 0.89, Fig. 3). All ten structural hits from the Protein Data Bank were GLUT proteins from other species ranging from bacteria to vertebrates (Datasheet A1, Tab C). Among the cnidarian GLUT sequences, conservation was highest for residues on the inner surface of the translocation pore, where the active binding sites are presumably located (Fig. 3). The predicted ligand was maltose (Fig. 3), a disaccharide composed of two glucose molecules (Berg et al., 2002). However, confidence of this prediction was low (C-score = 0.11), so we cannot make definitive conclusions about the solutes transported by the cnidarian GLUT protein. Overall, the results of the structural analysis support the hypothesis that the identified cnidarian proteins are highly conserved homologs of human GLUT proteins.

### 3.2. Cnidarian SGLT proteins

A total of 23 cnidarian sequences were discovered as putative homologs to SGLT proteins. While most sequences contain the 12 TM helices indicative of the amino acid-polyamine- organocation superfamily (APC), to which the SGLTs belong (Saier, 2000), three sequences appear to be protein fragments (Datasheet A2). Only one cnidarian sequence was identified as likely to encode for a protein that primarily
Fig. 3. Protein structural analysis of selected sequences from each gene family. Predicted 3D structures (left), conservation analysis (middle), and predicted ligands (right) of selected sequences from cnidarian GLUT (A), SMIT (B), and AQP (C) proteins and Symbiodinium AMT (D) and NRT2 (E) proteins. Predicted structures of each protein are colored by structure (alpha helix = pink, coil = white, with yellow arrows showing sequence direction) and aligned with the top matching structure from the Protein Data Bank, depicted by a cyan mesh ribbon. Conservation scores are rendered onto the protein models using a color gradient, from blue (low conservation, score = 1) to red (high conservation, score = 9), with yellow regions indicating insufficient data for analysis. Ligand predictions show the ligand at the predicted binding sites with transparent protein models colored by secondary structure (helices = red, coils = blue). (A, left) The cnidarian GLUT protein, as deduced from the amino acid sequence of N. vectensis, 4 GLUT8 aligned with the human GLUT1 protein (PDB ID = 4ppp). (A, middle) The N. vectensis GLUT8 predicted structure colored by degree of residue conservation from all newly identified cnidarian GLUT sequences. (A, right) The N. vectensis GLUT8 protein showing predicted ligand maltose (C-score = 0.11). (B, left) The SGLT protein as deduced from the N. vectensis 5 SGLT sequence aligned with the K294A mutant of VSGLT (PDB ID = 2aq2). (B, middle) The N. vectensis SMIT protein colored by degree of residue conservation from all newly identified cnidarian SGLT sequences. (B, right) The N. vectensis SMIT protein showing predicted ligands sodium (top, C-score = 0.21) and galactose (bottom, C-score = 0.19). (C, left) The cnidarian AQP protein as deduced from the N. vectensis 4 AQP3 sequence aligned with the E. coli glycerol facilitator GLPF (PDB ID = 1e6b). (C, middle) The N. vectensis AQP3 protein colored by degree of residue conservation from all newly identified cnidarian AQP sequences. (C, right) The N. vectensis AQP protein showing predicted ligand glycerol (C-score = 0.46). (D, left) The Symbiodinium AMT protein as deduced from the Symbiodinium C1_25 AMT sequence aligned with the S. cerevisiae methylammonium/ammonium permease Mep2 (PDB ID = SarxA). (D, middle) The Symbiodinium AMT protein colored by degree of residue conservation from all newly identified Symbiodinium AMT sequences. (D, right) The Symbiodinium AMT protein showing predicted ligands methylamine (top, C-score = 0.1) and ammonia (bottom, C-score = 0.09). (E, left) The Symbiodinium NRT protein as deduced from the Symbiodiniumam_24 NRT sequence aligned with the E. coli glycerol-3-phosphate transporter GlpT (PDB ID = 1pw4). (E, middle) The Symbiodinium NRT2 protein colored by degree of residue conservation from all newly identified Symbiodinium NRT2 sequences. (E, right) The Symbiodinium NRT2 protein showing predicted ligand monoacylglycerol (C-score = 0.06). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transports glucose (N. vectensis SLGT4) rather than myo-inositol. This sequence clustered with two other cnidarian SGLT sequences that were previously identified in A. pallida, and this group was nested inside a clade of cnidarian SMIT proteins. Another group of SMIT proteins resolved into a sister clade of the vertebrate SMIT clade. Additionally, one cnidarian sequence was identified as a choline transporter (CHT) with high similarity to the A. pallida CHT protein. The majority of cnidarian sequences were identified as putative sodium myo-inositol transporters (SMIT) or sodium-coupled monocarboxylate transporters (SMCT) (Fig. 4). All of the identified SGLT homologs were predicted to be localized to either the plasma membrane or internal membranes and contained the sodium solute symporter domain (Datasheet A2).

All proteins in the three main clades (CHT, SMCT, and SGLT) are part of the SLC5 gene family and function as sodium ion symporters. Recovered cnidarian proteins did not share high sequence similarity to the SGLT1-5 reference sequences, rather they were closely related to SMCT or SMIT proteins (Fig. 4). The SMCT1 protein (encoded by SLC5A8) couples sodium transport with that of short-chain fatty acids, and is associated with tumor suppression in humans (Gupta et al., 2006; Miyaschi et al., 2004). Cnidarian SMIT proteins were primarily predicted to be localized to internal membranes, with fewer in the plasma membrane. Additionally, multiple SMCT sequences were found in two of the symbiotic coral species, Fungia scutaria and P. australiensis, as well as in the aposymbiotic anemone N. vectensis (Datasheet A2). Fatty acids are a major component of the metabolites exchanged between partners in the cnidarian-dinoflagellate symbiosis (Dunn et al., 2012; Hillyer et al., 2016; Imbs et al., 2014). Therefore, the role of SMCT in symbiotic cnidarians may be to transport host-derived fatty acids to the symbionts, if isosorps of the protein can be expressed in the symbiosome.

The clade of SGLT family proteins has high bootstrap support, and the tree shows SGLT and SMIT sequences clustering together within vertebrate and invertebrate sub-clades, where one invertebrate clade is sister to the vertebrate SMITs and another consists entirely of cnidarian proteins (Fig. 4). The SMIT proteins resolving into two different clades may be due to differently functioning SMIT isoforms, where those branching from the vertebrate clade may act more similarly to the vertebrate homologs while proteins in the cnidarian SMIT/SGLT group may have more derived functions and different ligand transporting capabilities. Nine cnidarian proteins in total were identified as SMITs, which are closely related to SGLTs of the SLC5 gene family. These proteins function similarly to proton myo-inositol transporters (HMIT), primarily transporting myo-inositol (Bourgeois et al., 2005). SMIT1 can also transport a variety of sugars, including glucose, although it does so with low affinity (Wright, 2013), while SMIT2 can transport D-chiro-inositol (Coady et al., 2002). Myo-inositol is important for the growth of mammalian cells and is a key component in phospholipid biosynthesis and eukaryotic signaling pathways (Holub, 1986). The identification of both HMIT and SMIT transporters in cnidarians suggests an uncharacterized role for myo-inositol in cnidarian metabolism or cell structure.

A SMIT protein sequence from N. vectensis (N. vectensis 5 SMIT) was used for 3D structural analysis (Fig. 3). The predicted tertiary structure of the protein most closely resembles that of the K294A mutant of the VSGLT protein from Escherichia coli (TM score = 0.78). Overall residue conservation was low, with the most conserved residues lining the protein pore (Fig. 3). The ligands were predicted to be sodium (C-score = 0.21) and galactose (C-score = 0.19), which only weakly supports the idea that this protein functions as a sodium-coupled sugar transporter (Fig. 3). Proteins in the SGLT family are known to transport galactose, some with a high affinity (Zhao and Keating, 2007); however, experimental analysis is needed to confirm this function in cnidarians.

3.3. Metazoan AQP proteins

A total of 16 cnidarian sequences were identified as AQP homologs, all with at least five TM domains typical of AQPs (Datasheet A3). In cnidarians, the majority of the AQP homologs were predicted to be localized to internal membranes (Datasheet A3). All sequences contained domains present in members of the major intrinsic protein family and in AQP transporters, while some also had domains associated with specific AQPs (Datasheet A3). Our phylogenetic analysis resulted in the emergence of two sister clades: (1) aquaglyceroporins; and (2) classical aquaporins (Fig. 5). All sequences identified as aquaglyceroporins contained the conserved aspartate (D) residue that signifies a larger protein pore (Agre et al., 2002). Aquaglyceroporin proteins all share a common ancestor, but these proteins were shown to have diverged into vertebrate and invertebrate lineages before the emergence of the different protein members. The cnidarian aquaglyceroporins shared high sequence similarity with AQP3 or AQP9 and were labelled as such. However, since divergence of vertebrate and invertebrate lineages occurred before diversification of AQP isoforms within each taxonomic clade, functions of the AQP protein members could be different between invertebrates and vertebrates. These sub-clades of exclusively cnidarian sequences are highly supported and also include two previously identified AQP3-like sequences from A. pallida. In mammals, aquaglyceroporins can transport a diversity of solutes other than glycerol, including urea and water (Rojek et al., 2008). AQP3 primarily transports water and glycerol, while AQP9 can also transport larger molecules, such as lactate, purines, and pyrimidines (Ishibashi et al., 2011). Experimental evidence is needed to confirm which of these small solutes the cnidarian AQP proteins transport.

The sub-clade of metazoan classical aquaporins contains cnidarian sequences sharing a common ancestor with vertebrate and invertebrate AQP1, as well as vertebrate AQP4 and AQP0 (Fig. 5). While AQP1 and AQP4 are water-only channels, AQP0 has a variety of functions in other organisms (Agre et al., 2002). AQP0 is the lens major intrinsic protein,
Fig. 4. Phylogenetic tree of cnidarian SGLT sequences and closest homologs from other organisms. The tree was rooted by the bacterial sodium solute symporter (SSS) homolog from Acanthamoeba castellanii. The number of amino acid substitutions is indicated by the scale bar, and bootstrap support for nodes > 50% is represented by shaded circles. Cnidarian sequences are shown in bold text. Three main clades were resolved: (1) choline transporters (CHT), (2) sodium-coupled monocarboxylate transporters (SMCT), and (3) sodium/glucose cotransporters (SGLT and SMIT). The cluster of vertebrate SMIT proteins is collapsed to enhance tree clarity. The sequence used for structural analysis is marked by a yellow star (N. vectensis 5 SMIT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
originally found in lens fiber cells of vertebrate eyes and later discovered to function in increasing the water permeability of animal cells (Gorin et al., 1984; Mulders et al., 1995). Its suspected structural role in cell adhesion make AQP0 unique in the aquaporin family (Fotiadis et al., 2000). Cell adhesion proteins are important in the cnidarian-dinoflagellate symbiosis for cell-cell interactions, such as signaling between partners (Reynolds et al., 2000); cnidarian AQP0 may function in this capacity.

One AQP9 sequence from *N. vectensis* (*N. vectensis* AQP3) was used for structural analysis of the cnidarian AQP protein (Fig. 3). The deduced structure was very similar to that of the *E. coli* glycerol facilitator, GLPF (Fig. 3, TM score = 0.90). Additional hits from the Protein Data Bank were protein structures of aquaglyceroporins from other species (Datasheet A3, Tab C). Furthermore, the predicted ligand (Fig. 3) was glycerol (C-score = 0.46), and the most highly conserved residues among all cnidarian AQPs were those lining the protein pore (Fig. 3). Overall, the structural analysis supports the prediction that these proteins function as glycerol channels.

### 3.4. *Symbiodinium* AMT proteins

A total of 40 *Symbiodinium* sequences were putatively identified as AMT proteins. Most of these sequences contained 10–12 TM domains, which is typical of AMT proteins (Thomas and Mullins, 2000; von Wieren et al., 2000), while five sequences were shorter, containing only seven to nine TM helices. The sequences were primarily identified as general AMT proteins rather than specific members or isoforms and were predicted to be localized to either the cell membrane or internal membranes (Datasheet A4, Tab D). All sequences contained the domain for ammonium transport (Datasheet A4, Tab D).

Phylogenetic analysis showed the divergence of five main AMT clades: (1) bacterial cyclases; (2) metazoan-like AMT proteins; (3) diaphoretickes AMT proteins; (4) alveolate AMT proteins; and (5) bacterial-like AMT proteins (Fig. 6). The clade of bacterial cyclases is not well supported and includes the enzymes adenylate cyclase and guanylate cyclase exclusively from bacterial species.

Several clades of AMT proteins were identified in the phylogenetic analysis. The first clade shows AMT proteins diverging into three sub-clades, although not all are well resolved (Fig. 6). Clade 2 has robust support and shows metazoan AMT proteins emerging alongside a cluster of exclusively *Symbiodinium* proteins. This clade includes four *A. pallida* sequences in the metazoan cluster and five of the *Symbiodinium* sequences that contain the domain characteristic of AMT3. This finding indicates that horizontal gene transfer may have occurred between
Fig. 6. Phylogenetic tree of Symbiodinium AMT sequences and closest homologs from other organisms. The tree was rooted by the bacterial homolog from a Spirochaeta sp. The number of amino acid substitutions is indicated by the scale bar, and bootstrap support for nodes > 50% is represented by shaded circles. Symbiodinium sequences are shown in bold text. The proteins resolved into five main clades: (1) bacterial cyclases, (2) metazoan-like ammonium transporters, (3) plant-like ammonium transporters, (4) chromalveolate ammonium transporters, and (5) bacterial-like ammonium transporters. The sequence used for structural analysis is marked by a yellow star (Symbiodinium C1_25 AMT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Symbiodinium and the last common ancestor of this metazoan group. Evidence of horizontal gene transfer between symbiotic partners has also been discovered between Aiptasia and Symbiodinium by researchers who assembled the Aiptasia genome (Baumgarten et al., 2015). Twenty-nine Aiptasia gene products were found to best align with Symbiodinium proteins, rather than metazoan proteins, and clustered together in a phylogenetic analysis (Baumgarten et al., 2015).

Clade 3 is only weakly supported overall and includes proteins from organisms within the expansive diaphoretickes super-group (Adl et al., 2012). Nevertheless, a well-supported sub-clade shows a Symbiodinium AMT sequence (SymbiodiniumC1,12) grouping with vascular plant and green algal proteins. This Symbiodinium sequence also had high sequence similarity to the AMT1-3, which in plants has a high affinity for the uptake of both methylammonium and ammonium (von Wieren et al., 2000). While clade 4 is also lacking adequate bootstrap support, it contains a highly supported sub-clade that includes a large amount of expressed Symbiodinium C1 AMT proteins in relation to other alveolates (Fig. 6). While expression of multiple AMT proteins is observed for almost all alveolate species in the analysis, Symbiodinium C1 was found to express a total of 28 proteins, while other alveolate species have only one to four expressed AMT proteins. Symbiodinium C1 is known as a generalist symbiont that occurs worldwide, forming associations with a variety of invertebrate hosts (LaJeunesse, 2005; Stat et al., 2009).

Consequently, the upregulation of AMT proteins in this Symbiodinium type may have evolved as an adaptation to survive the varying nitrogen conditions experienced in different host cell environments.

The second lineage including Symbiodinium AMT proteins is a well-supported clade containing a mixture of unresolved sequences from bacteria, chlorophyta, and alveolata (Fig. 6, clade 5). Overall, our analysis of the Symbiodinium AMT proteins and their homologs has revealed a large diversity of these transportsers within this dinoflagellate species, which may have arisen through multiple evolutionary events.

For structural analysis of the Symbiodinium AMT proteins, one sequence from Symbiodinium clade C (SymbiodiniumC1,25) that falls within the alveolate cluster was used. The predicted 3D structure most closely resembles the structure of the Saccharomyces cerevisiae methylammonium permease, Mep2 (TM score = 0.84, Fig. 3). Residues lining the inside of the protein channel are highly conserved (Fig. 3), suggesting their importance in transport function. Ligand predictions (Fig. 3) had very low confidence scores but favored methylamine (C-score = 0.1) and ammonia (C-score = 0.09). The information gained from this structural analysis provides preliminary evidence that these highly conserved proteins serve as ammonium channels.

3.5. Symbiodinium NRT2 proteins

Only four sequences in total were identified as Symbiodinium NRT2 homologs, all of which were from Symbiodinium clade A, although a sequence from clade C and an unspecific Symbiodinium sequence within the NCBI database were also found to be related to NRT2. The four sequences had high similarity with one another (74–79%) and ranged in length from 549 to 613 amino acids with 11–12 TM helices, typical of MFS transporters (Datasheet A5). Half of these NRT2 homologs were predicted to be localized to the plasma membrane and the other half to internal membranes, while all sequences contained the domains for MFS and were related to high affinity nitrate transporter 2.3 (Datasheet A5, Tab D).

Phylogenetic analysis of the NRT2 proteins showed divergence of: (1) eukaryotic; and (2) bacterial nitrate transporters (Fig. 7). Surprisingly, very few NRT homologs were recovered from the Symbiodinium databases, and no other alveolate NRT2 homologs were identified (Fig. 7). NRT2 genes are highly regulated by environmental nitrate concentrations. As part of the high-affinity uptake system, the genes are upregulated during nitrogen starvation (< 0.5 mM) and repressed when nitrate levels are high (> 0.5 mM) (Lezhneva et al., 2014; Okamoto et al., 2003). If the Symbiodinium NRT2 protein level is regulated in the same way, then the gene would be repressed when the cells are grown in the commonly used 1/2 culture medium, since it contains 0.882 mM nitrate (Guillard, 1975). However, as copies of the NRT2 genes were not discovered in the Symbiodinium genomes surveyed, identification of the protein may indicate that this gene has either been lost in some derived species of this genus but retained by the more ancestral Symbiodinium clade A, or that the proteins discovered were from contaminants in the clade A transcriptome.

The presence of nitrate transporters in only Symbiodinium clade A could be due to the free-living nature of some dinoflagellate species within this clade (Hirose et al., 2008; Reimer et al., 2010), allowing the cells to more readily utilize nitrate from seawater. Ammonium is excreted by the host, while nitrate is not, and ammonium has been predicted to be the primary nitrogen source for symbiotic Symbiodinium cells (Kopp et al., 2013; Pernice et al., 2014). Many Symbiodinium types that have evolved to be strictly symbiotic may have lost the need for high-affinity nitrate transport, while those with a partially or wholly free-living lifestyle may have retained it.

While we had initially hypothesized that the identified Symbiodinium NRT2 homologs are nitrate transporters, the structural analysis indicates otherwise. The predicted 3D structure of the putative Symbiodinium NRT2 sequence (Symbiodinium A-24 NRT) was most similar to that of the glycero1-3-phosphate transporter from E. coli (TM score = 0.73, Fig. 3), rather than any nitrate transporters. Overall conservation was low, and one of the TM helices was only found in a single sequence (insufficient data, Fig. 3). The ligand was predicted as monoacylglycerol (Fig. 3) but with low confidence (C-score = 0.06). While the identified potential NRT2 homologs resemble integral membrane transport proteins, they are very likely MFS transporters with a different function.

4. Conclusions

This study identified sugar transport proteins from cnidarians and nitrogen transport proteins from their dinoflagellate symbionts. Homology searches identified previously undescribed cnidian GLUT, SGLT, and AQP homologs and Symbiodinium AMT and NRT2 homologs. Phylogenetic and structural analyses demonstrated high conservation of most of the investigated protein families, suggesting that the characterized transport functions of the reference sequences are likely retained in the identified cnidian and Symbiodinium sequences.

Considerably more cnidian sequences were identified in the GLUT family than in the SGLT family, signifying that GLUT proteins could be involved in the transport of glucose between the symbiont and cnidian host. Furthermore, potentially related transporters of the SLC5 gene family known to transport other compounds, such as monocarboxylates and myo-inositol, were identified in the cnidarians. Protein sequences for cnidian aquaglyceroporins were also discovered, suggesting a potential mechanism for glycerol translocation into cnidian cells.

Many of the sugar transporters discussed here have predicted localizations to internal membranes, which are defined in the MemPype pipeline as “the endoplasmic reticulum, the nuclear membranes, the Golgi apparatus, the vesicles, the vacuoles, the lysosomes, the peroxisome, the microsomes, and the endosome” (Pierleoni et al., 2011). Since the symbiosome is derived from an early endosome (Fitt and Trench, 1983; Wakefield and Kempf, 2001), we presume that symbiosome-localized proteins would fall into the “internal membrane” category. Transporters predicted as localized to only internal membranes were OCT and AQP9 (Datasets A2–A3). Investigations of these proteins might identify them as being associated exclusively with the symbiosome, which would be helpful in the establishment of specific marker proteins for identification of symbiosome membranes.

Additionally, more Symbiodinium AMT sequences were identified than NRT sequences, which is likely due to ammonium being the primary nitrogen source for symbiotic dinoflagellates (D’Elia et al., 1983;
Based on our findings, we suggest the use of AMT transporters to investigate the nutritional conditions experienced by symbionts in hospite. Since expression of these proteins is known to be regulated by environmental ammonium concentrations, AMT transporter expression patterns by Symbiodinium cells in hospite could reveal information about the nutrient conditions experienced within symbiotic host cells. It has previously been hypothesized that the host can regulate intracellular symbiont populations by limiting their growth and nutrient supply (Falkowski et al., 1993; Jones and Yellowlees, 1997; Muscatine and Pool, 1979; Wooldridge, 2010). In plants and various algal species, AMT isoforms are transcriptionally regulated by environmental conditions, in a manner similar to that for genes encoding NRT2 proteins, where high-affinity isoforms are expressed during nitrogen starvation and low-affinity isoforms are expressed under ample nutrient availability (Gazzarrini et al., 1999; González-Ballester et al., 2004; Lezhneva et al., 2014). Therefore, verification of regulated expression in Symbiodinium and analysis of AMT proteins during colonization experiments may help to establish whether or not the host imposes nitrogen limitations on the symbiont population.

The resources used to compile cnidarian and Symbiodinium databases in this study included mostly transcriptomic data. Molecular phylogenies conducted on proteins rather than DNA are more accurate at establishing relationships between lineages, since proteins are the molecules on which natural selection acts (Salemi and Vandamme, 2003)). While transcriptomes contain highly useful information about the proteins being transcribed during the times of sampling, they represent only a snapshot of an organism’s genome and not a comprehensive view of its total coding capacity. The results of this study can be interpreted as putative identifications of symbiont and host transport proteins, while the absence of any proteins in our datasets does not necessarily indicate their absence from the organism. All new protein identifications described here are inferred from homology and require verification through experimental analyses. While the results cannot be used to deduce the full range of cnidarian and Symbiodinium proteins from each of these five transporter families, they do give us a previously unseen glimpse of a subset of transporter proteins that are likely used in the cnidarian-dinoflagellate symbiosis. Using currently available genetic resources, we have taken the first steps towards uncovering mechanisms used in the exchange of carbon and nitrogen between the partner organisms in the cnidarian-Symbiodinium symbiosis. With the emergence of new genomic resources and proteomic-based experiments, we are likely to develop a more complete picture of transporter functions and mechanisms in symbiotic associations.

Acknowledgements

We thank Eli Meyer for the use of his Perl scripts and acknowledge the Center for Genome Research and Biocomputing at OSU for providing computational resources for this project. Financial support for this research was provided by the Marsden Fund of the Royal Society of New Zealand, grant number 1202, to S.K.D., V.M.W., and A.R.G. The funding source had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. We thank the library of science and medical illustrations (http://www.somersault1824.com/science-illustrations/), creative commons license CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/), for objects used to create Fig. 1. We also thank two anonymous reviewers for their


Muscatine, L., 1967. Glycerol excretion by symbiotic algae from corals and and


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muscietta, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muscietta, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muscietta, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami


Muschlitz, L., Mecloeskey, L.R., Hoeh-Guldberg, O., McAuley, P.J., 1994. Effect of ammonium enrichment on animal and algal biomass of the coral Porites dami