Funct Integr Genomics
DOI 10.1007/s10142-009-0156-5

ORIGINAL PAPER

Comparative transcriptomics for mangrove species: an expanding resource

Maheshi Dassanayake · Jeff S. Haas · Hans J. Bohnert · John M. Cheeseman

Received: 6 October 2009 / Revised: 30 November 2009 / Accepted: 24 December 2009
© Springer-Verlag 2010

Abstract We present here the Mangrove Transcriptome Database (MTDB), an integrated, web-based platform providing transcript information from all 28 mangrove species for which information is available. Sequences are annotated, and when possible, GO clustered and assigned to KEGG pathways, making MTDB a valuable resource for approaching mangrove or other extremophile biology from the transcriptomic level. As one example outlining the potential of MTDB, we highlight the analysis of mangrove microRNA (miRNA) precursor sequences, miRNA target sites, and their conservation and divergence compared with model plants. MTDB is available at http://mangrove.illinois.edu.

Keywords Mangroves · Transcriptome · Database · Extremophile · miRNA target sequences · Small RNA

Introduction

The mangrove ecosystem is defined by a group of halophytes, predominantly trees that dominate tropical intertidal zones and estuaries. Being evolutionarily adapted to tolerate flooding, anoxia, high temperatures, wind, and high and extremely variable salt conditions in typically resource-poor environments (Cheeseman et al. 1991), mangroves are an untapped physiological and molecular resource for understanding and exploiting plant adaptations to extreme environments.

While mangroves have, evolutionarily, had a similar need to adapt to common environmental constraints, individual taxa—representing more than 15 unrelated plant families (Hogarth 2007)—have developed different physiological, life history, and morphological strategies. The genetic basis for these strategies is, however, virtually unknown: mangroves, like other extremophiles, are poorly represented in the plant molecular literature. Whether the goal is to understand the comparative biology of mangroves, or to exploit the group's unique genetic resources, much greater genome-level understanding is needed.

Here, we present, first, the design, implementation, and use of a database that can facilitate research addressing that need. The goal of the Mangrove Transcriptome Database (MTDB) is to provide a central resource, not only to the mangrove research community, but to the broader community that requires easy access to curated transcript sequences from extremophile plants. MTDB was developed largely because, for a taxonomically heterogeneous group such as mangroves, community-specific BLAST searches are not possible on NCBI.

Second, we present an example of the use of MTDB to approach a biologically interesting question. Using data extracted from MTDB, we explore post-transcriptional control based on small RNAs (sRNAs). Post-transcriptional regulation via microRNAs (miRNAs) and small interfering RNAs (siRNAs) has recently and explosively emerged as an additional ubiquitous and significant regulatory mode in plant development and stress responses (e.g., Griffiths-Jones et al. 2008; Rymarquis et al. 2008; Sunkar et al. 2007). While the use of EST databases in analyzing genetic structural diversity

M. Dassanayake · H. J. Bohnert · J. M. Cheeseman (✉)
Department of Plant Biology, University of Illinois,
505 S. Goodwin Ave,
Urbana, IL 61801, USA
e-mail: j-cheese@illinois.edu

J. S. Haas
Office of Networked Information Technologies (ONIT),
School of Integrative Biology, University of Illinois,
505 S. Goodwin Ave,
Urbana, IL 61801, USA

Published online: 27 January 2010

is well established, their use in investigating gene regulation has largely been limited to model systems.

MTDB is available via the World-Wide Web at http://mangrove.illinois.edu. Questions and comments related to MTDB should be addressed to mangrove@illinois.edu. We request that users of MTDB cite this article in publications related to its use.

Materials and methods

Data acquisition and annotation The MTDB was created to organize and make available the collection of mangrove sequences which resulted from a 454/Roche GSFLX pyrosequencing project and their subsequent assembly into unique contigs. Two normalized cDNA libraries were sequenced, one established with RNA from *Rhizophora mangle* L. (Rhizophoraceae) and the other from *Heritiera littoralis* Aiton. (Malvaceae) (Dassanayake et al. 2009). MTDB was further expanded to include 24,061 cDNA sequences from 26 additional mangrove species which were downloaded from NCBI nucleotide, protein, and EST databases. All sequences were formatted into searchable BLAST databases, enabling species and community specific sequence similarity searches based on sequences, queries, proteins, genes, and gene accessions. To enable BLASTn searches against the database, we have incorporated a standalone BLAST server using NCBI's wwwblast application. The website that overlays the database was designed to be simple and optimized for speed of sequence and BLAST searches. Additional details of the database structure and operation are included in the "About/Help" section of MTDB.

sRNA discovery For selected miRNA families, especially ones implicated in environmental responses, primary miRNA transcripts (pri-miRNAs) and precursor miRNAs (pre-miRNAs) (Kurihara et al. 2006) were identified using MTDB annotations coupled with the included BLAST tool. These were checked against the miRBase, microRNA registry (Griffiths-Jones et al. 2008) for further verification.

Indirect evidence for miRNA occurrence was also sought using conserved miRNA targets in mangrove mRNAs (Filipowicz et al. 2008). In practice, the problem is to accommodate the gaps and mismatches which occur between known, experimentally confirmed miRNA and target mRNA base pairing (Brennecke et al. 2005). Here, we limited candidate targets to near perfect complementary base-pairing between 2 and 12 nt from the 5′ end of the miRNA, allowing (1) only one mismatch, and not at the 10th or 11th position (thought to be the cleavage site), (2) a total of three, but not more than two consecutive, additional mismatches after the 12th nucleotide, and (3) calculated base-pairing free energies of at least 72% of the values that would result from perfect base pairing between the miRNA and the target (Schwab et al. 2005).

Results

Database sequence composition The 454/Roche sequencing project on which MTDB is based resulted in 97,155 *H. littoralis* and 67,524 *R. mangle* sequences, representing 13,598 and 13,049 gene models, respectively. We estimated these to represent ca. 50% of each transcriptome (Dassanayake et al. 2009). Additionally, more than 24,000 cDNA sequences from 26 additional mangrove species were downloaded from NCBI databases. Table 1 summarizes the number of annotations made for each species. As our intent was to make MTDB exploitable in functional genomics studies, we excluded mangrove entries from public databases containing promoter, intron and intergenic spacer regions that commonly harbor microsatellite repeats. For species with fewer than 1,000 ESTs, we manually excluded duplicate records. Because many EST sequences from GenBank were originally annotated with functionally uninformative names, e.g. clone numbers, or were assigned functions based on the project objective alone, e.g., "salt-tolerant protein," we re-annotated all sequences to find the most descriptive annotation possible before uploading to MTDB (Dassanayake et al. 2009). Whenever possible, Gene Ontology (GO) and KEGG pathway annotations were assigned based on appropriately annotated plant reference genomes in NCBI.

Approximately 88% of the successful annotations were based on *Arabidopsis thaliana*with the remainder commonly based on sequences from other extreme plants, e.g. *Salicornia bigelovii*, *Larrea tridentata*, *Euphorbia tirucalli*, *Thellungiella halophila*, and *Mesembryanthemum crystallinum*. Nearly fifty percent of all sequences, however, shared no significant sequence similarity, within our annotation criteria, with any other sequence in GenBank. In part, these might represent untranslated portions of mangrove transcripts (3′-UTRs) with low homology to the corresponding sequences in other species. However, a large number of these are likely to be novel mangrove coding region sequences, highlighting the large proportion of the unexplored sequence space in mangroves.

Data searching and retrieval There are two types of searches through the MTDB web interface: searches against the transcriptome database, and BLASTn searches against the mangroves sequence collection. Transcriptome searches can be based on nucleotide sequences, transcript names in MTDB (also denoted as query names), protein names,

Table 1 A summary of the numbers of mangrove sequences and annotations in MTDB, ordered by family and species

Species	No. of Sequences	Proteins[a]	Genes[b]
Myrsinaceae			
Aegiceras corniculatum	127	58	41
Avicenniaceae			
Avicennia alba	3	3	1
Avicennia bicolor	9	9	1
Avicennia germinans	56	56	4
Avicennia marina	1,896	1,659	1,470
Avicennia schaueriana	2	2	0
Rhizophoraceae			
Bruguiera cylindrica	126	56	40
Bruguiera gymnorhiza	20,664	11,204	11,063
Bruguiera sexangula	59	55	48
Ceriops australis	4	4	2
Ceriops decandra	4	4	4
Ceriops tagal	6	6	5
Kandelia candel	18	15	14
Rhizophora mangle	67,524	32,284	26,928
Combretaceae			
Conocarpus erectus	9	9	6
Laguncularia racemosa	6	6	4
Lumnitzera littorea	3	3	3
Lumnitzera racemosa	2	2	2
Euphorbiaceae			
Excoecaria agallocha	5	5	5
Excoecaria cochinchinensis	5	5	5
Malvaceae			
Heritiera littoralis	97,155	40,223	31,284
Lythraceae			
Sonneratia alba	446	443	403
Sonneratia apetala	154	112	77
Sonneratia caseolaris	297	294	253
Sonneratia griffithii	2	2	2
Sonneratia griffithii x alba	31	31	31
Sonneratia ovata	129	126	91
Sonneratia paracaseolaris	2	2	1

[a] Number of sequences that found a protein match through BLAST searches

[b] Number of sequences that had a match to genes in NCBI reference genomes

genes names, GO terms, KEGG IDs, and Entrez reference numbers. Searches are initiated by entering search terms into a form (see Fig. 1). Unless a search field is marked as "exact", case-insensitive substring searches are performed. "Exact" searches are used for species and ID-type fields. Terms can be entered in multiple fields to further restrict the results, one example being species specific searches initiated by selecting from a drop-down list in the species field. The main search page also has a "no hit sequences" link to download, in FASTA format, all sequences which do not share sequence homology within reasonable e-values to any known sequence in GenBank.

Search results are presented in tabular format, grouped by species (Fig. 2), and the matching sequences can be accessed individually or downloaded as FASTA files. Each annotation in the results table contains links to NCBI nucleotide, protein, gene, or EST databases, thereby enabling easy navigation to additional biological information.

The second MTDB search resource comprises the BLAST page. BLASTn searches are allowed against species specific databases within MTDB, or against the full mangrove community. As with the web-based BLAST services at NCBI, stand alone BLAST searchs can be optimized by altering parameters such as e-value, matrix, and filtering of low complexity regions. The results are displayed as a graphical overview and pairwise alignments (Altschul et al. 1990).

sRNA discovery using MTDB One interesting illustration of MTDB's potential is the search for small RNAs (sRNAs), including both microRNAs (miRNAs) and small-interfering RNAs (si-RNAs). This search is made possible by the fact that many plant miRNAs are evolutionary conserved (Axtell and Bartel 2005; Reinhart et al. 2002). Ambros et al. (2003) introduced a uniform system to identify miRNAs based on evidence for (1) expression and (2) biogenesis. These criteria are partially satisfied by miRNA searches against EST databases (as evidence of expression) and by identifying miRNAs through homology (as evidence for biogenesis) (Zhang et al. 2005). Thus, we searched for miRNAs in MTDB with particular attention to candidate miRNAs and their targets involved in developmental and environmental responses.

Because MTDB largely originated from the analysis of a size-selected poly-A enriched cDNA library designed to identify protein coding genes, the frequency with which *mature* miRNAs can be expected in MTDB is low. However, miRNAs originate as long primary transcripts (pri-miRNA) which are processed by DCL1 to form precursor miRNAs (pre-miRNA) (Kurihara et al. 2006). These form the stable stem-loop-hairpin structures by which they can be recognized (Bonnet et al. 2004). Pre-miRNAs are then further processed to mature miRNAs (miR) (Brodersen and Voinnet 2006). As pri-miRNAs arise from transcription by RNA polymerase II and they are subsequently polyadenylated (Bushati and Cohen 2007), putative mangrove pri-miRNA transcripts would be treated identically to mRNAs during construction of the libraries for transcriptome sequencing (Dassanayake et al. 2009). Additionally, because of the typically near perfect base-pairing observed between the target mRNA and the

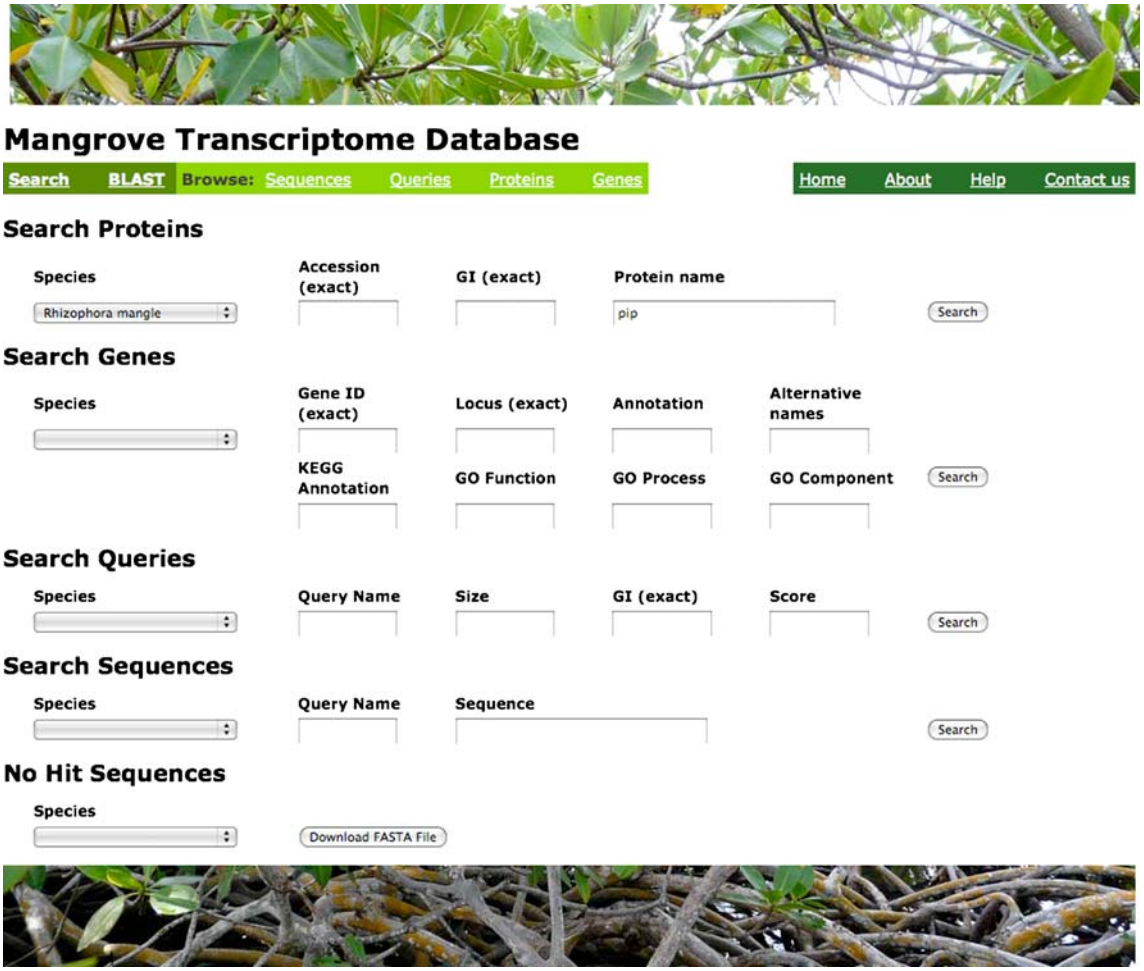

Fig. 1 The entry portal "Search" page from the Mangrove Transcriptome Databaseshowing different data searches/strategies (based on proteins, genes, queries, and sequences), and the "no hit sequences" download button. In this illustration, a protein search is to be conducted for PIP proteins, limited to *Rhizophora mangle*

regulatory miRNA in plants (Filipowicz et al. 2008), evidence for miRNAs in MTDB was also sought based on the sequences of their targets.

Table 2 summarizes 19 mangrove gene models representing 12 miRNA families identified using MTDB annotations coupled with the BLAST tool incorporated in MTDB. These were checked against the miRBase, microRNA registry (Griffiths-Jones et al. 2008) for further verification. Plant miRNAs are categorized into classes based on their level of conservation in the plant kingdom (Rajagopalan et al. 2006; Zhang et al. 2006). Based on analysis of putative targets, miR156, 166, 168, 172, and 396, in the 'highly conserved' class, are represented in MTDB. Mangrove gene models detected for the 'moderately conserved' miRNA class include miR162, 164, 169, and 397, while the gene model detected for miR408 is considered to be in the 'weakly-conserved' miRNA class.

miRNAs targeting siRNAs are also represented in MTDB. For example, *R. mangle* E5XRSP401AJ46V was annotated based on its high similarity with the *Physalis longifolia* trans-acting siRNA, TAS3 (gi225904417; BLASTn e-value 3e-13). TAS3 is thought to regulate expression of auxin (ARF2, ARF4) and ethylene (ETT) response factor genes. In *Arabidopsis*, TAS3 (At3g17185) is, in turn, regulated by miR390 (Allen et al. 2005). The base pairing complementarity of the *R. mangle* TAS3 target with the *Arabidopsis* miR390 is actually greater than that observed in *Arabidopsis*. *R. mangle* contig 4313, *H. littoralis contig* 11121, and *Bruguiera gymnorhiza* (gi 53821871), each also annotated as TAS3 in the MTDB, contain the miR390 target with the same number of mismatches at the same positions as those found in *Arabidopsis* TAS3. This suggests that two *R. mangle* TAS3 isoforms may be processed differently or with different efficiencies (Pillai et al. 2007).

While miRNA involvement in the regulation of development is well documented, their importance in plant responses to diverse stress conditions is also increasingly recognized. miR399, for example, which targets a

Mangrove Transcriptome Database

Search | BLAST | Browse: Sequences | Queries | Proteins | Genes | Home | About | Help | Contact us

Search Proteins

Species | Accession (exact) | GI (exact) | Protein name | Search

Listing Proteins

Proteins found: 21

GI	Protein Name							Sequences
	Query: Species	Query Name	Size	GI	Accession	Short ID	Score	E-value
104304208	Capsicum annuum PIP1 (PIP1) gene, complete cds							download
	Heritiera littoralis	E5VR0NL01DRNXE	193	104304208		Capsicum annuum PIP1 (PIP1) gene, co...	56	8e-006
15225913	PIP1B (plasma membrane intrinsic protein 1;2); water channel							download
	Bruguiera gymnorhiza	gi|53813582|gb|BP939081.1|BP939081 ...	514	15225913	NP_182120	PIP1B (plasma membrane intrinsic pr...	219	2e-057
47558817	CAPIP1 [Capsicum annuum]							download
	Rhizophora mangle	Rhizophora_allseq.fna.Contig1696	264	47558817		CAPIP1 [Capsicum annuum]	37	0.003
145339450	Arabidopsis thaliana PIP2A (PLASMA MEMBRANE INTRINSIC PROTEIN 2A); water channel (PIP2A) mRNA, complete cds							download
	Rhizophora mangle	E5XRSP401D65VJ	43	145339450	NM_115202	Arabidopsis thaliana PIP2A (PLASMA...	48	3e-005
145323704	PIP1C (PLASMA MEMBRANE INTRINSIC PROTEIN 1;3)							download
	Heritiera littoralis	Heritiera_allseq.fna.Contig26764	343	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	192	7e-050
	Rhizophora mangle	Rhizophora_allseq.fna.Contig9737	347	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	49	9e-007
	Rhizophora mangle	Rhizophora_allseq.fna.Contig17870	292	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	59	9e-010
	Heritiera littoralis	E5VR0NL01BEMUQ	77	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	52	1e-007
	Rhizophora mangle	E5XRSP401A0Y30	153	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	102	1e-022
	Bruguiera gymnorhiza	gi|53813326|gb|BP938825.1|BP938825 ...	500	145323704	NP_001077441	PIP1C (PLASMA MEMBRANE INTRINSI...	96	2e-020

Fig. 2 A sample of a protein search output, displaying the accession number, protein names, and GI number with related queries for each entry listed below. The target of the search was PIP proteins in all mangrove species. The query summary is displayed as species, query name, size of the sequence, short ID, best match GI number, short ID, score and e-value for the best match. In the absence of a good homolog in *A. thaliana* according to our annotation criteria, a homolog was sought within all green plants. The first entry lists a PIP protein found in *Capsicum annum* that has a best match to a *R. mangle* query sequence; the rest of the sequences find matches to proteins from *A. thaliana*

ubiquitin-conjugating enzyme (*AtPHO2,* At2g33770), is induced during low-phosphate stress (Bari et al. 2006); *PHO2* mRNA accumulation is decreased under stress, which leads to induction of the phosphate transporter gene, *AtPT1*, and attenuation of primary-root elongation (Fujii et al. 2005). Interestingly, sequences annotated as *PHO2* are present in MTDB entries for *R. mangle*, but the target sites for miR399 are not. Whether this reflects a case in which the 5'UTR (site of the target in *Arabidopsis*) is not represented in the database, or that there is a significant difference in the regulation of responses to P-deficiency in mangroves is an important question. *R. mangle*, for example, can survive for decades under extreme P limitation, and respond to a single fertilization with a dramatic increase in growth and restructuring of overall metabolism (Cheeseman and Lovelock 2004; Feller 1995; Lovelock et al. 2006).

The effects of miR399 are also modulated by an mRNA-like non-coding RNA (mincRNA), TPSI1 (Rymarquis et al. 2008). This RNA binds miRNA, but due to a base mismatch between bases 10 and 11, the critical region for cleavage, it is not cleaved, effectively immobilizing the miRNA. Rymarquis et al. (2008) noted that TPSI1 mincRNAs from different species share little sequence similarity except for the region complementary to miR399 where the similarity is very high. MTDB has sequences in both *H. littoralis* (E5VR0NL01BP386) and *R. mangle* (contig 1118) that share the mincRNA target sequence, including the mismatch at nucleotide 11 from the 5′ end.

Table 2 Precursor miRNA sequences in MTDB

Pri/pre-miRNA	MTDB reference ID	Mature sequence present in the MTDB sequence	Free energy of predicted pre-miRNA stem-loop structure (kcal/mol)	Highest % identity with another pre-miRNA
156	*R. mangle* contig 12525	Yes	−30.2	83% with *P. trichocarpa* 156f
156	*H. littoralis* E6PJTYN03DCHB4	Yes	−63.9	85% with *V. vivifera* 156f
162	*H. littoralis* E5VR0NL01DGZTG	Yes	−55.8	96% with *G. hirsutum* 162a
164	*R. mangle* E5XRSP401CKIDS	No	–	82% with *P. trichocarpa* 164e
166	*H. littoralis* contig 2702	Yes	−52.3	73% with *P. trichocarpa* 166j
166	*R. mangle* contig 9968	Yes	−32.4	87% with *V. vivifera* 166e
166	*R. mangle* E5VR0NL01BHKYB	Yes	−25.6	85% with *V. vivifera* 166e
168	*R. mangle* E5XRSP401AWKQ6	No	–	95% with *G. max* 168
169	*H. littoralis* contig 13803	Yes	−60.2	42% with *B. napus* 169l
172	*H. littoralis* E5VR0NL01BI7J1	No	–	59% with *G. max* 172b
396	*R. mangle* E5XRSP401CN5EB	No	–	97% with *G. max* 396a
396	*R. mangle* contig 21413	No	–	94% with *G. max* 396a
396	*B. gymnorhiza* gi53825750	Yes	−65.8	53% with *G. max* 396a
397	*H. littoralis* contig 237	Yes	–	88% with *P. trichocarpa* 397a
397	*H. littoralis* contig 12797	Yes	−36.8	96% with *S. lycopersicum* 397
397	*H. littoralis* contig 31097	Yes	−43.0	68% with *A. thaliana* 397a
408	*H. littoralis* contig 28953	Yes	−49.8	83% with *P. trichocarpa* 408
828	*H. littoralis* contig 24594	Yes	−53.5	48.7% with *A. thaliana* 828
1877	*R. mangle* contig 15138	Yes	−45.5	52.5% with *O. sativa* 1877

Many MTDB sequences are partial transcripts and therefore, free energies are not given for incomplete stem-loop structures based on sequence similarity predictions. Free energies were calculated using Mfold (http://www.bioinfo.rpi.edu/applications/mfold) (Zucker 2003)

However, because there are no significant sequence similarities on either side of that position, these are currently considered "unknown-unknowns" (Class R3, Dassanayake et al. 2009).

The distribution and positions of miRNA targets in mangroves suggest, in some cases, regulatory roles that differ in potentially significant ways from those in *Arabidopsis*. Our analysis has revealed examples of four different models, schematically represented in Fig. 3.

The simplest model (Fig. 3a) is that in which both the apparent target sequence and its position are conserved between species. miR164, for example, targets NAC domain-containing proteins in *Arabidopsis* (AT5G61430, ATNAC05) and *Populus* (gi224132939, NAC 21). MTDB contains putative targets conserved in syntenic positions in *H. littoralis* (contig 20266), *R. mangle* (E5XRSP401EYYTF), and *Avicennia marina* (gi17312651). Other transcripts conforming to this model include AFB2 (Auxin Signaling F-Box 2, At3g26810) which is a target for miR393, and mangrove homologs *H. littoralis* contig 20939 and *R. mangle* contig 20063. miR394 targets found in F-box family proteins (*R. mangle* contig 5783, *H. littoralis* E6PJTYN03C0P03, and *Arabidopsis* At1g27340) also conform to this model; even the mismatches are conserved between mangroves and *Arabidopsis* (Table 3).

In some cases, entire gene families conform to this model. miR160, for example, regulates a transcription factor gene family, the auxin response factors (ARFs) (Fukaki et al. 2007). *R. mangle* contig 23274 and its homolog *Arabidopsis* ARF10 (At2g28350), and *H. littoralis* contig 22754 and its homolog *Arabidopsis* ARF17 (At1g77850), all contain the miR160 target sequence (Table 3) conserved in the same position.

Strict conservation is not the rule, however. The second model is that in which a miRNA target is conserved between mangroves, but is not present in *Arabidopsis* (Fig. 3b). The *R. mangle* homolog of SPL5 (contig 18186), for example, carries the target site for miR156 in an extended 3′UTR lacking in its *Arabidopsis* homolog (At3g15270). Placement of a target in this region is uncommon in plants (Brodersen et al. 2008; Dezulian et al. 2006). As miR156 responds to salt stress in both glycophytic dicots and monocots (Dassanayake et al. 2009; Sunkar and Zhu 2004), this may reflect a modification associated with evolution of a mangrove lifestyle (Dassanayake et al. 2009). Alternately, the mangrove

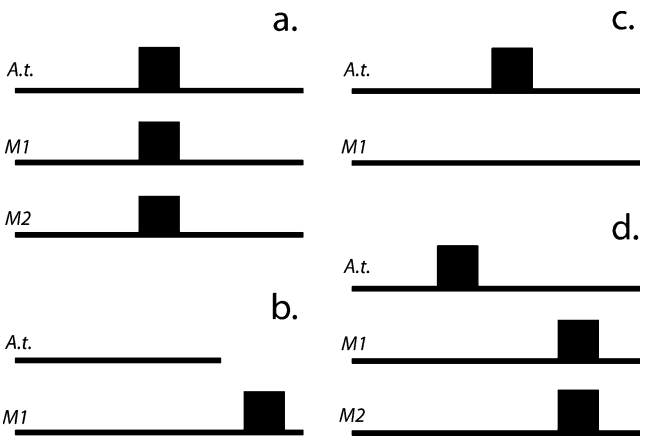

Fig. 3 Schematic comparison of miRNA targets in mangroves and *A. thaliana* (*A.t.*). *Lines* represent the target genes and the *boxes* represent miRNA target sequences. M1 and M2 symbolize homologous target genes of two mangrove species. **a** miRNA target is conserved across all species, or across multiple species and multiple genes in a family; **b** miRNA target is conserved in mangroves (either in the 3′-UTR or the coding region) but absent in *A. thaliana*; **c** miRNA target is absent in the mangrove homologs, but present in *A. thaliana*; **d** miRNA target is conserved in mangroves and *A. thaliana,* but the position is different. See text for specific examples of each scenario

targets may be conserved in the coding region, rather than in the UTR, but nonetheless be absent in the *Arabidopsis* homolog. *R. mangle* E5XRSP401DGYAA and *H. littoralis* contig 20644, for example, contain the target sequence for miR530 in the same position, while the target is absent in the homologous *Arabidopsis* sequences (At2g28670/ At2g28671).

In contrast, Fig. 3c illustrates a model in which a target is absent in the mangroves but present in the *Arabidopsis* homologs: the miR399 target found in the *Arabidopsis* vesicle associated membrane protein family (VAMP, At4g00170) is not present in the mangrove homologs, *H. littoralis* contigs 30188 and 21786, and *R. mangle* contig 24107.

Finally, miRNA target sequences may be conserved while their position within the transcript is not (Fig. 3d). This is the case when miR414 targets in *R. mangle* (E5XRSP401C8NVK) and *H. littoralis* (E5VR0NL01EHK95) are compared to *Arabidopsis* (At4g36860).

Discussion

The importance of genomics in understanding physiology, biodiversity, life history, and evolution is now indisputable (Bohnert and Sheveleva 1998; Cushman and Bohnert 2000; Sørensen and Loeschcke 2007). Beyond model systems research, integrative approaches, incorporating insight from physiological and molecular knowledge along with knowledge of ecological processes, have gained increasing recognition at all levels of conservation (Gutschick and BassiriRad 2003; Young et al. 2006) including mangrove conservation (Schwarzbach and Ricklefs 2001). As Karrenberg and Widmer (2008) noted, "... the only way to understand ecologically meaningful genetic variation ... is to bring the ecological and molecular perspectives together."

In this spirit, we have developed MTDB to serve the research community as a central sequence-anchored resource for studies ranging from conservation biology, to the physiology of plant adaptations to extreme environments, and to explication of convergent evolution in an ecologically and sociologically important plant community. As the breadth of genome coverage is improved both in more mangroves and in extremophile plants in general, we expect that insight from both data mining and interspecific comparisons will contribute increasingly to understanding plant adaptations to extreme environments. We plan to support this by updating MTDB continuously by adding newly published mangrove ESTs to our sequence assemblage and regularly re-annotating the "no hit sequences". We will also extend this database to include miRNAs and siRNAs as these become available.

In this report, using the first generation of MTDB, we have illustrated its use for one of many possible in silico explorations, the identification of mangrove miRNAs and the consideration of possible alternative regulatory models to those reported for *Arabidopsis*. The identification of plant miRNAs has received increasing attention in recent years, largely in a few model species such as *A, thaliana* and *Oryza sativa*. At the same time, post-transcriptional gene regulation by miRNAs has gained interest as a potential mechanism for crop improvement (Sunkar et al. 2007; Zhang et al. 2006). Many well-conserved miRNAs have been identified; however both family- and species-specific miRNAs are being discovered with the application of deep sequencing technologies (Morin et al. 2008; Moxon et al. 2008; Subramanian et al. 2008).

The example presented here illustrates the potential of MTDB for information retrieval beyond simple sequence similarity searches, highlighting a number of significant differences between mangrove targets and their *Arabidopsis* counterparts, and suggesting the additional potential value of miRNAs in studies of ecological genomics. We hypothesize that expression profiles of miRNAs and their predicted targets could be a useful tool in exploring the significance of their conservation patterns, particularly in responses of plants to abiotic stresses. Mangroves and other extremophiles, because of their well-developed abilities in this regard, are excellent models for this exploration.

Table 3 Selected miRNA targets found in mangroves and *Arabidopsis*. Mismatches in the alignment are underlined

miRNA	5′- 3′ miRNA sequence aligned to 3′- 5′target mRNA sequence	ID in MTDB/ *Arabidopsis* gene models (TAIR locus tag)	Annotation in MTDB	Free energy compared with a perfect match
miR156	TGACAGAAGAGAGTGAGCAC ACTGTCTTCTCTCTCTCGTA	*R. mangle* contig 18186	SPL5	82%
miR160	TGCCTGGCTCCCTGTATGCCA ACGGACCGAGGGACGTACGGT	*H. littoralis* contig 22754, AT1G77850	Auxin response factor 17	99%
miR160	TGCCTGGC-TCCCTGTATGCCA ACGGACCGGAGGGACATACGGA	*R. mangle* contig 23274	Auxin response factor 10	87%
miR160	TGCCTGGCTCCCTGTATGCCA ACGGACCGAGGGACATAAGGA	AT2G28350	Auxin response factor 10	83%
miR164	TGGAGAAGCAGGGCACGTGCA ACCTCTTCGTCCCGTGC–GT	*H. littoralis* contig 20266	ANAC100/ATNAC5 transcription factor	82%
miR164	TGGAGAAGCAGGGCACGTGCA ACCTCTTCGTCCAGTGCAC-T	*R. mangle* E5XRSP401EYYTF	unnamed protein product	73%
miR164	TGGAGAAGCAGGGCACGTGCA ACCTCTTCGTCACGTGCACCG	*A. marina* gi17312651	ANAC100/ATNAC5 transcription factor	74%
miR164	TGGAGAAGCAGGGCACGTGCA ACCTCTTCGTCCCGTGCATCT	AT5G61430	ANAC100/ATNAC5 transcription factor	86%
miR390	AAGCTCAGGAGGGATAGCGCC TTCGAGTCCTCTCTATAGCGG	*R. mangle* E5XRSP401AJ46V	Similar to *V. vinifera* contig VV78X098962.9; TAS3	80%
miR390	AAGCTCAGGAGGGATAGCGCC ATCGAGTCCTCCCTATCTGTT	AT3G17185	TAS3	71%
miR393	TCCAAAGGGATCGCATTGATCC AGGTTTCCCTAGCGTAACAAAG	*H. littoralis* contig 20939, *R. mangle* contig 20063, AT3G26810	AFB2	83%
miR394	UTGGCATTCTGTCCACCTCC AACCGTAAGACAGTTGGAGG	*R. mangle* contig 5783, *H. littoralis* E6PJTYN03C0P03, AT1G27340	F-box family protein	85%
miR399	TGCCAAAGGAGAGTTGCCCTG TCGGTATCCTCTCAACGGAAG	AT4G00170	VAMP (vesicle-associated membrane family protein)	85%
mir399	TGCCAAAGGAGAGTTGCCCTG ACGAGACCCTCCTGAGAGGAG	*H. littoralis* contig 30188	VAMP	37%
miR399	TGCCAAAGGAGAG-TTGCCCTG ACTAGACCCTCCTGCGGGGGAG	*H. littoralis* contig 21786	VAMP	51%
miR414	TCATCTTCATCATCATCGTCA AGGAGAAGTAGTAGTAGTAGT	*R. mangle* E5XRSP401C8NVK, *H. littoralis* E5VR0NL01EHK95	Unknown proteins	95%
miR414	TCATCTTCATCATCATCGTCA (non-syntenic valid target) AGTAGGAGTAGCAGTAGCAGT (syntenic invalid target) AGAAGTAGGAGTAGGAGTAGC	AT4G36860	zinc ion binding protein	88% and 59%
miR530	TGCATTTGCACCTGCACCTT ACGTAAACGTGGACGTGGAC	*R. mangle* E5XRSP401DGYAA	Unknown/similar to *V. vinifera* contig VV78X045012.8	99.8%
miR530	TGCATTTGCACCTGCACCTT ACGTAAACGTGGACGTGGCT	*H. littoralis* contig 20644	Unknown/similar to *V. vinifera* contig VV78X045012.8	95%
miR530	TGCATTTGCACCTGCACCTT CCATGGACGTGGACGTGGTC	AT2G28670/ AT2G28671	hypothetical/disease resistance-responsive family protein	64%
miR1310	GGCATCGGGGGCGTAACGCCCCU CCGTAGCCCCCGCGTTGCGGGAG	*A. marina* gi17312899	Hypothetical protein similar to At3g41950	90%
miR1310	GGCATCGGGGGCGTAACGCCCCU CCGTAACCCCCGCGTTGCGGGAG	AT3G41950	Hypothetical protein	80%

Free energies were calculated using nucleic acids sequence annealing available at http://www.mag.keio.ac.jp/~rsaito/Research/BasePAP/BasePAP.html (Osada et al. 1999)

Acknowledgements The authors are indebted to the Smithsonian Caribbean Coral Reef Ecosystem (CCRE) project, and especially to Klaus Rützler, Candy Feller, Mike Carpenter and all the Carrie Bow Cay station managers for continued support for the field work and access to Twin Cays; to the Vice Chancellor for Research at the University of Illinois at Urbana-Champaign for funding the sequencing; to Shahjahan Ali, Jyothi Thimmapuram and Deepika Vulaganthi in the Keck Center for Comparative and Functional Genomics at UIUC for sequencing and assembly; and to Robert Bocchino and Sahan Dissanayake for help with Perl scripts. This is contribution number 870 of the CCRE program, Smithsonian Institution, supported in part by the Hunterdon Oceanographic Research Endowment.

References

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121:207–221

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones SAM, Marshall M, Matzke M, Ruvkun G, Tuschl T (2003) A uniform system for microRNA annotation. RNA 9:277–279. doi:10.1261/rna.2183803

Axtell MJ, Bartel DP (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17:1658–1673. doi:10.1105/tpc.105. 032185

Bari R, Datt Pant B, Stitt M, Scheible W-R (2006) PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol 141:988–999. doi:10.1104/pp.106.079707

Bohnert HJ, Sheveleva E (1998) Plant stress adaptations - making metabolism move. Curr Opin Plant Biol 1:267–274

Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics 20:2911–2917. doi:10.1093/bioinformatics/bth374

Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. PloS Biol 3:e85

Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. Trends Genet 22:268–280

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science 320:1185–1190. doi:10.1126/science.1159151

Bushati N, Cohen SM (2007) microRNA functions. Annu Rev Cell Dev Biol 23:175–205. doi:10.1146/annurev.cellbio.23.090506.123406

Cheeseman JM, Lovelock CE (2004) Photosynthetic characteristics of dwarf and fringe *Rhizophora mangle* in a Belizean mangrove. Plant Cell Environ 27:769–780

Cheeseman JM, Clough BF, Carter DR, Lovelock CE, Eong OJ, Sim RG (1991) The analysis of photosynthetic performance in leaves under field conditions: a case study using *Bruguiera* mangroves. Photosyn Res 29:11–22

Cushman JC, Bohnert HJ (2000) Genomic approaches to plant stress tolerance. Curr Opin Plant Biol 3:117–124

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM (2009) Shedding light on an extremophile lifestyle through transcriptomics. New Phytol 183:764–775

Dezulian T, Remmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. Bioinformatics 22:359–360. doi:10.1093/bioinformatics/bti802

Feller IC (1995) Effects of nutrient enrichment on growth and herbivory of dwarf red mangrove (*Rhizophora mangle*). Ecol Monogr 65:477–505

Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Rev Genet 9:102–114

Fujii H, Chiou T-J, Lin S-I, Aung K, Zhu J-K (2005) A miRNA involved in phosphate-starvation response in *Arabidopsis*. Curr Biol 15:2038–2043

Fukaki H, Okushima Y, Tasaka M, Kwang WJ (2007) Auxin-mediated lateral root formation in higher plants. Int Rev Cytol 256:111–137

Griffiths-Jones S, Saini HK, Sv D, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158. doi:10.1093/nar/gkm952

Gutschick VP, BassiriRad H (2003) Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences. New Phytol 160:21–42

Hogarth P (2007) The biology of mangroves and seagrasses. Oxford University Press, New York

Karrenberg S, Widmer A (2008) Ecologically relevant genetic variation from a non-*Arabidopsis* perspective. Curr Opin Plant Biol 11:156–162. doi:10.1016/j.pbi.2008.01.004

Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. RNA 12:206–212. doi:10.1261/rna.2146906

Lovelock CE, Ball MC, Choat B, Engelbrecht BMJ, Holbrook NM, Feller IC (2006) Linking physiological processes with mangrove forest structure: phosphorus deficiency limits canopy development, hydraulic conductivity and photosynthetic carbon gain in dwarf *Rhizophora mangle*. Plant Cell Environ 29:793–802

Morin RD, Aksay G, Dolgosheina E, Ebhardt HA, Magrini V, Mardis ER, Sahinalp SC, Unrau PJ (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. Genome Res 18:571–584. doi:10.1101/gr.6897308

Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res 18:1602–1609. doi:10.1101/gr.080127.108

Osada Y, Saito R, Tomita M (1999) Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes. Bioinformatics 15:578–581

Pillai RS, Bhattacharyya SN, Filipowicz W (2007) Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biol 17:118–126

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–3425. doi:10.1101/gad.1476406

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16:1616–1626. doi:10.1101/gad.1004402

Rymarquis LA, Kastenmayer JP, Hüttenhofer AG, Green PJ (2008) Diamonds in the rough: mRNA-like non-coding RNAs. Trends Plant Sci 13:329–334

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D (2005) Specific effects of microRNAs on the plant transcriptome. Dev Cell 8:517–527

Schwarzbach AE, Ricklefs RE (2001) The use of molecular data in mangrove plant research. Wetl Ecol Manag 9:205–211

Sørensen J, Loeschcke V (2007) Studying stress responses in the post-genomic era: its ecological and evolutionary role. J Bioscience 32:447–456

Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu J-K, Yu O (2008) Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics 9:160

Sunkar R, Chinnusamy V, Zhu J, Zhu J-K (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309

Sunkar R, Zhu J-K (2004) Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell 16:2001–2019. doi:10.1105/tpc.104.022830

Young JL, Bornik ZB, Marcotte ML, Charlie KN, Wagner GN, Hinch SG, Cooke SJ (2006) Integrating physiology and life history to improve fisheries management and conservation. Fish Fish 7:262–283

Zhang BH, Pan XP, Wang QL, Cobb GP, AT A (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15:336–360

Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. Plant J 46:243–259

Zucker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucl Acids Res 31:3406–3415. doi:10.1093/nar/gkg595