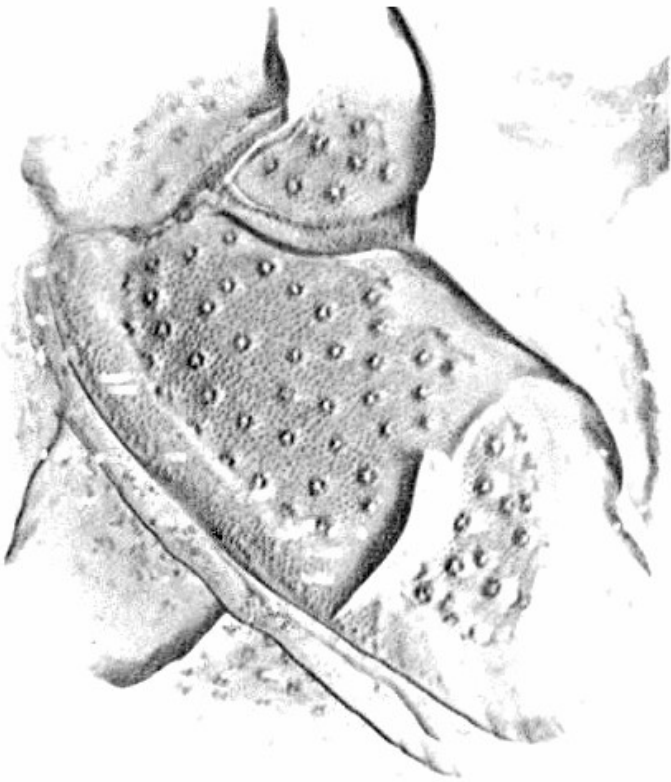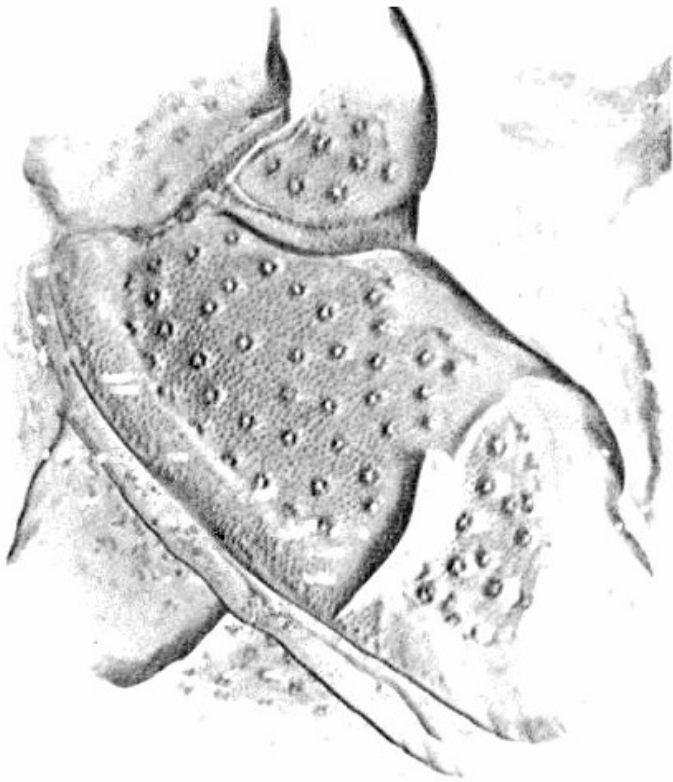# Genomic and Proteomic Studies on Toxic Algae for Enhanced Monitoring
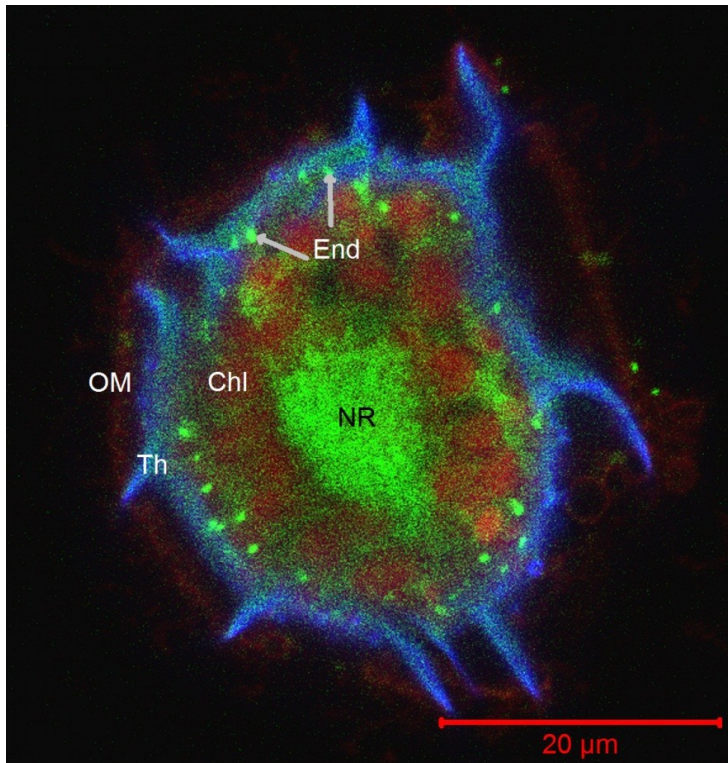
Arturo O. Lluisma
UP-Marine Science Institute

Why are we interested in dinoflaggelates

- many species produce toxins; pose threats to public health
- *Pyrodinium bahamense* var *compressum*
- species and genetic diversity in Philippne waters; implications on population structure / connectivity
- genomic peculiarities

Challenges

- not all species produce toxins
- levels of toxicity can vary
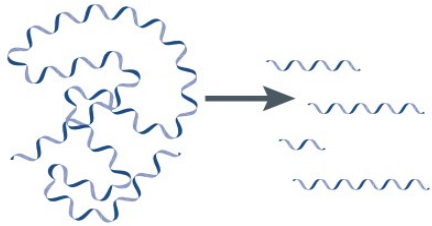- bacterial endosymbionts

Tools
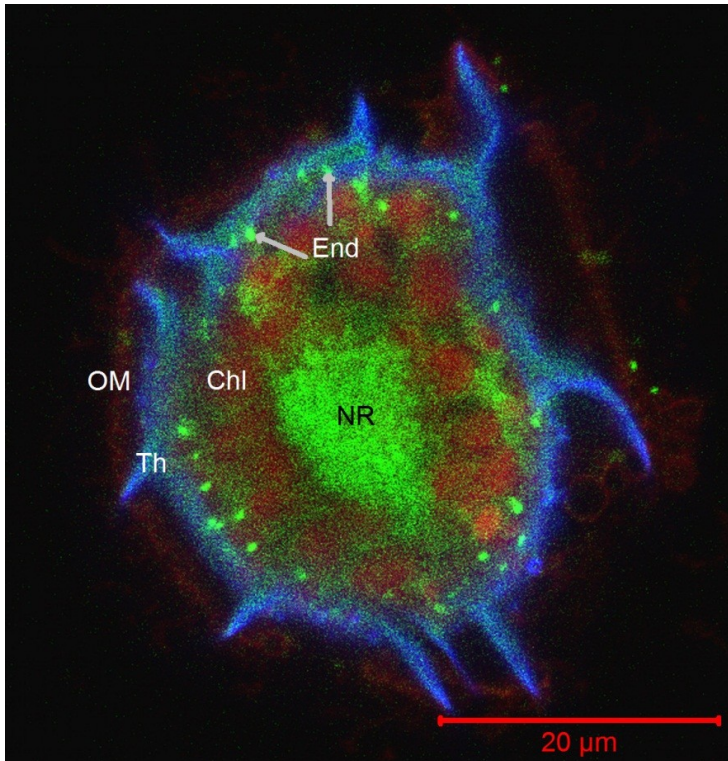
- genomics
- metagenomics
- proteomics
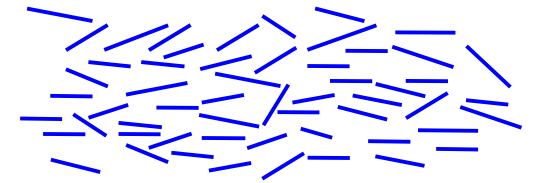- good old molecular tools

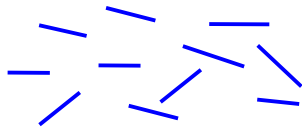# Next-generation sequencing (NGS)

gDNA

DNA extraction and library preparation

next-generation sequencing

raw reads (short sequences)

assembly of reads into contigs

unassembled reads

reads assembled into contigs

contig

contig

contig

search algorithms (BLAST, conserved-domain search, repeat-finding, etc.)

sequence / functional annotation, etc. (data-integration packages, e.g., MEGAN)

Dinoflagellate genomics: caveats

- very large genomes
- no reference genomes
- complex organization
- endosymbiotic bacteria: metagenome

Partial genome sequencing of
P. bahamense var *compressum*

	DNA	**mRNA**
Pbc state	Cystic	Vegetative
Platform	Illumina HiSeq 2000	Roche 454 GS Junior
No. of reads	239,161,531 x 2	161,854
Ave. read length (bp)	98	392
Total no. of bases (bp)	47,014,412,329	63,466,833

GC content histograms for DNA sequences and mRNA sequences
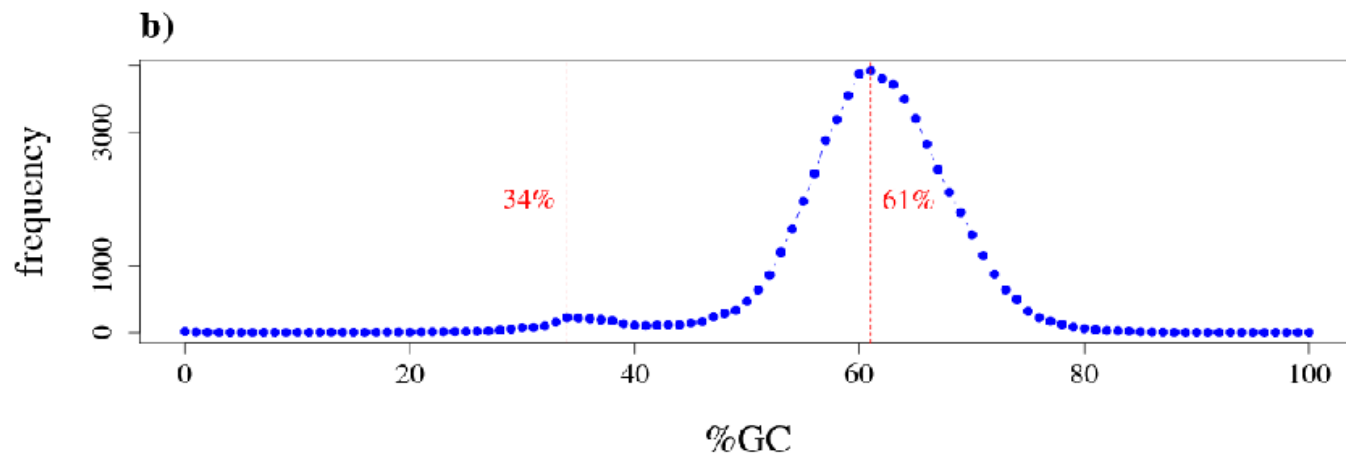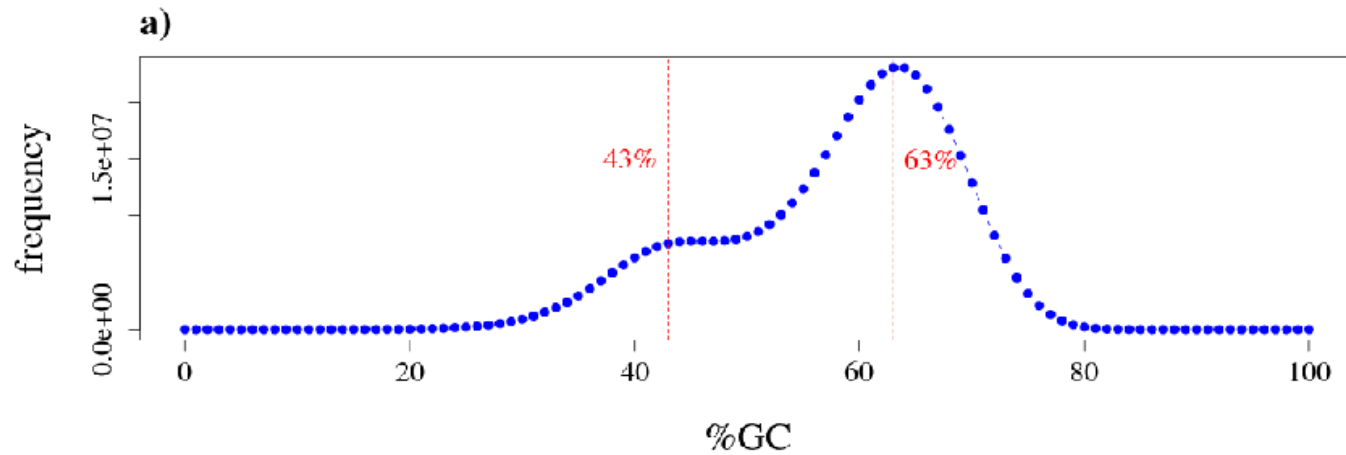
Saxitoxin Biosynthesis Pathway in Cylindrospermopsis raciborskii T3

Kellmann et al., 2008
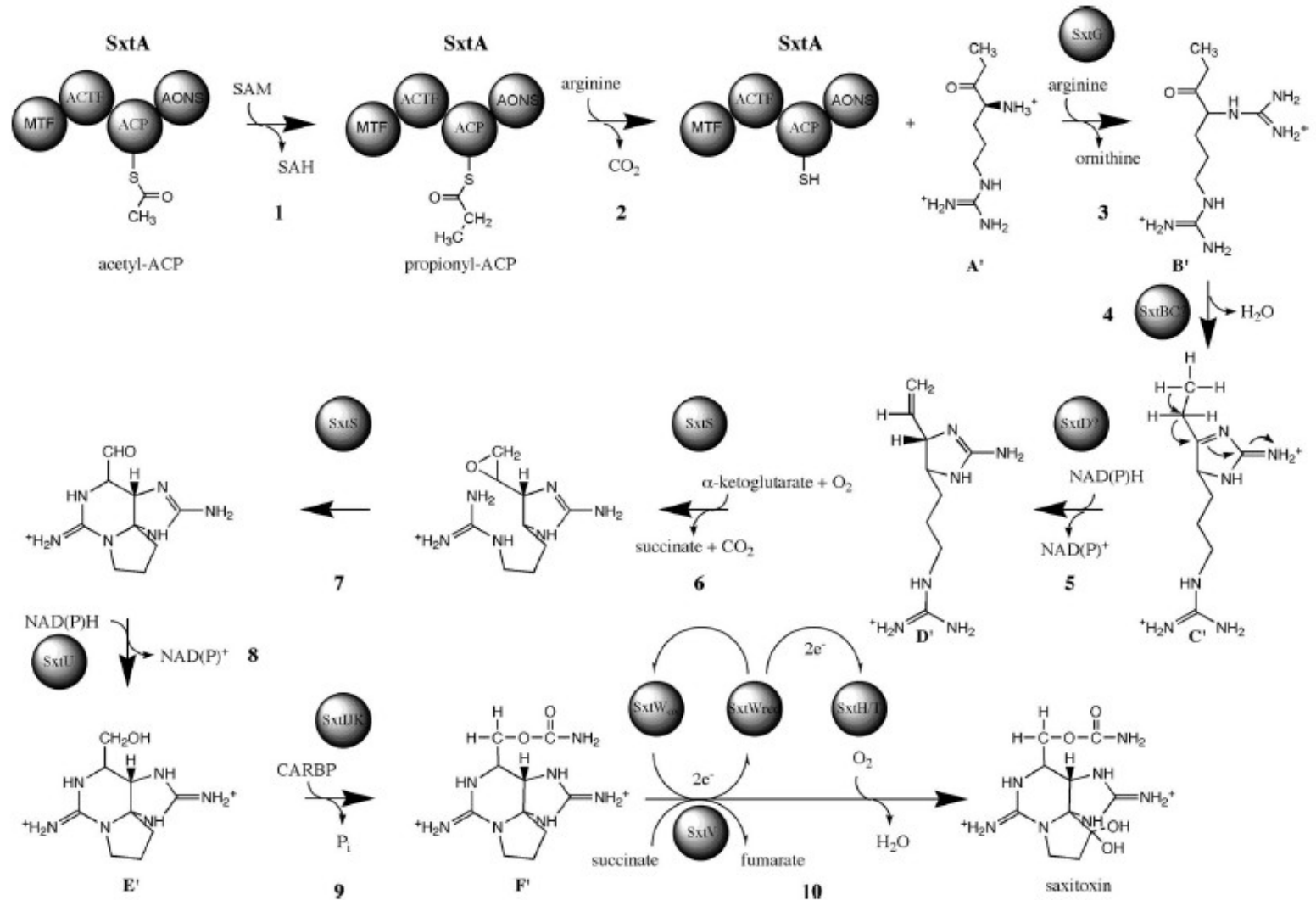
Putative homolog of *Alexandrium* sxtA

Reads aligned: 3

ADY62525 sxtA long isoform precursor (*A. fundyense*)

E-value: 0.0

Identity: 91-92%

Query coverage: ~80%

Hit region: residues 1165-1281
    Aspartate aminotransferase superfamily fold type I [AAT_I] (CDD199925)

```
   2 EGDLQRKRLLEATLGFCEGLKALECPHTYHGFPIINIYWTPVQVCAEVYRELMSARQGA  60
     EGDLQRKRLL ATL FCEGLKAL CPHTYHGFPI+NIYWTPV+VCAEVYRELMSARQGA
1165 EGDLQRKRLLAATLEFCEGLKALGCPHTYHGFPIVNIYWTPVEVCAEVYRELMSARQGA  1223

  61 FQRGVITTPMWYPIAPKGHEMLRFQFTSLHDEAAVRHILVILGDLIKRYPPSAEPPRI   118
     FQ+GV+TTPMW+PIAPKGHEMLRFQFTSLHDEAAVRHILVIL DLIKRYPPSA PPRI
1224 FQQGVVTTPMWHPIAPKGHEMLRFQFTSLHDEAAVRHILVILEDLIKRYPPSAVPPRI  1281
```

Conserved domain search for some of the

Domain	# read hits	# contig hits
GDSL-lipase	-	-
PKS/Acyl transferase	43	3
Alcohol dehydrogenase	148	10
Aminidinootransferase	5	-
Benzyl alcohol dehydrogenase	94	8
Cephalosporin hydroxylase	4	-
Cytidine deaminase	-	-
Ferredoxin	18	3
MATE	56	3
Methyltransferase	-	-
O-carbamoyltransferase	1	-
Phenylproionate dioxygenase	101	11
RTX toxin binding	3	1
Succinate dehydrogenase	60	5
Transposase	6	-

E-value cutoff: 0.01
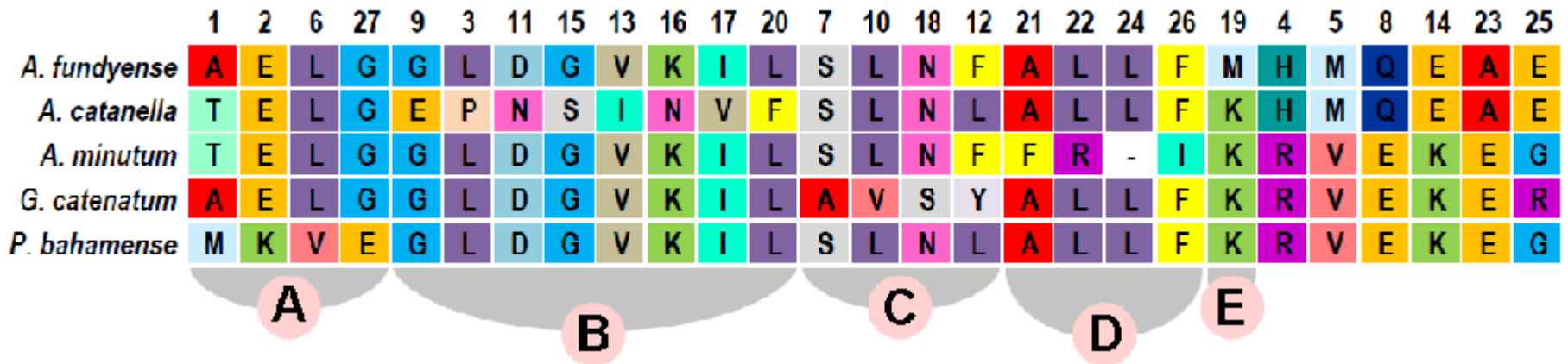
*STX-production-specific

Homologs in *P. bahamense*
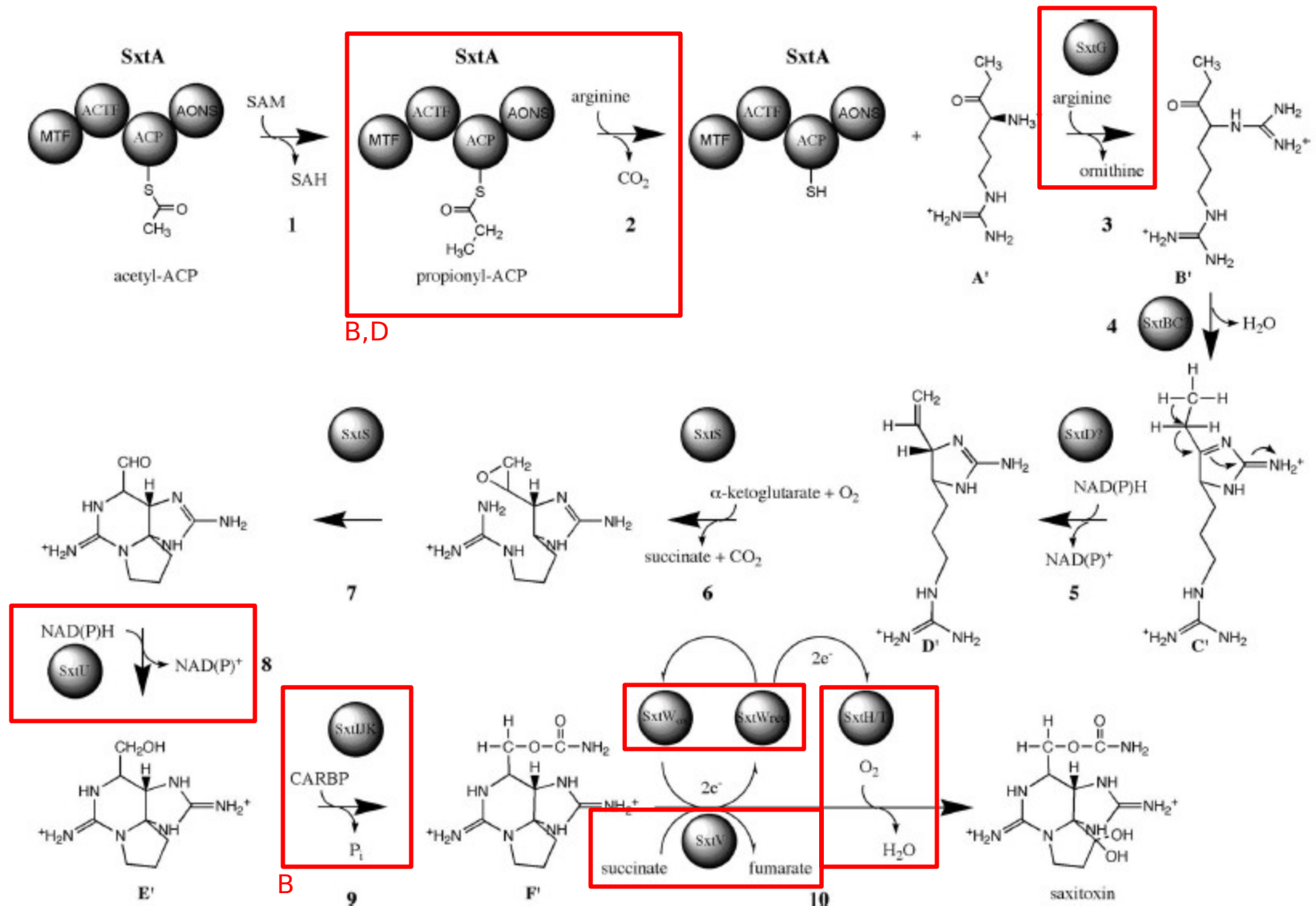
B – bacterial evidence; D – dinoflagellate evidence

Kellmann et al., 2008
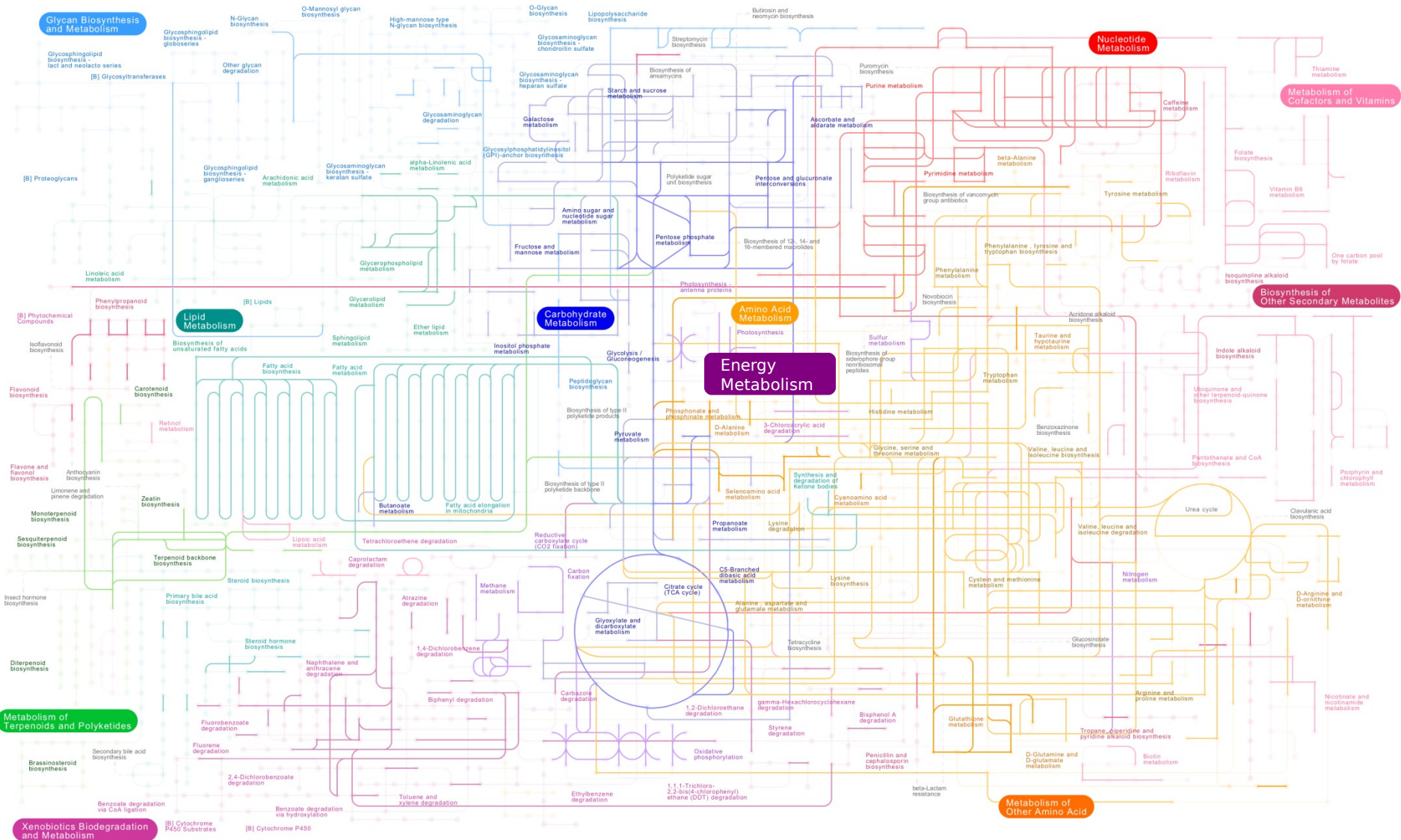
Variable sites in known *sxtA4*

Variable sites in known *sxtG*

Other significant matches to Alveolata sequences

GI Accession	Gene Name	Organism	# Reads
ABY78836	Luciferase-binding protein	*Alexandrium cantenella*	3
ACS13434	adenosylhomocysteinase	*Amphidinium carterae*	6
AAW80677	chloroplast PEP synthase	*Heterocapsa triquetra*	6
ABI14377	ubiquitin-conjugating enzyme E2 (catalytic domain)	*Pfiesteria piscicida*	1
ABI14376	small ubiquitin-like protein	*Pfiesteria piscicida*	1
ACU44795	senescence-associated like protein	*Pfiesteria piscicida*	3

Metabolic map based on KEGG for all reads

Diversity of bacterial sequences in the metagenome

Phyllobacteriaceae (0.9)
OMG group (0.5)
Rhizobiaceae (0.7)
Rhodothermaceae (1.3)
environmental samples <Bacteria> (0.5)
Pseudomonadaceae (0.6)
unclassified Flavobacteriales (1.2)
Cytophagaceae (0.6)
Hyphomonadaceae (2.2)
Rhodospirillaceae (1.7)
unclassified Verrucomicrobiales (2.2)
Flammeovirgaceae (1)
Piscirickettsiaceae (4)
Cyclobacteriaceae (1.9)
Erythrobacteraceae (2.1)
unclassified Alphaproteobacteria (2.1)
Planctomycetaceae (5.2)
Rhodobacteraceae (17.8)
Alteromonadaceae (4.7)
Flavobacteriaceae (2.9)

HAB Proteomics

- Bacterial proteins are expressed alongside dinoflagellate proteins

- May be used to identify bacterial endosymbionts

HAB Proteomics: insights into toxicity

- Toxicity studies on *Pyrodinium* with different light intensity and growth stage

- High toxin levels during exponential growth phase and moderate light intensity
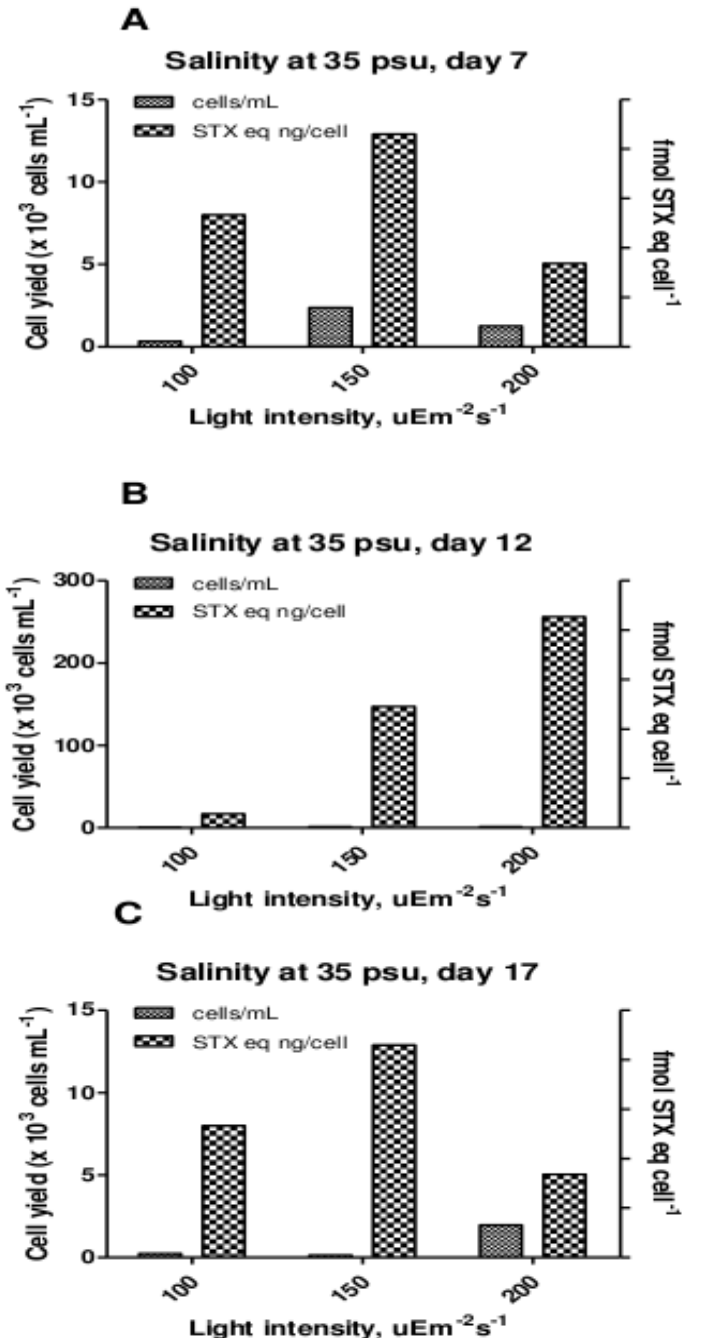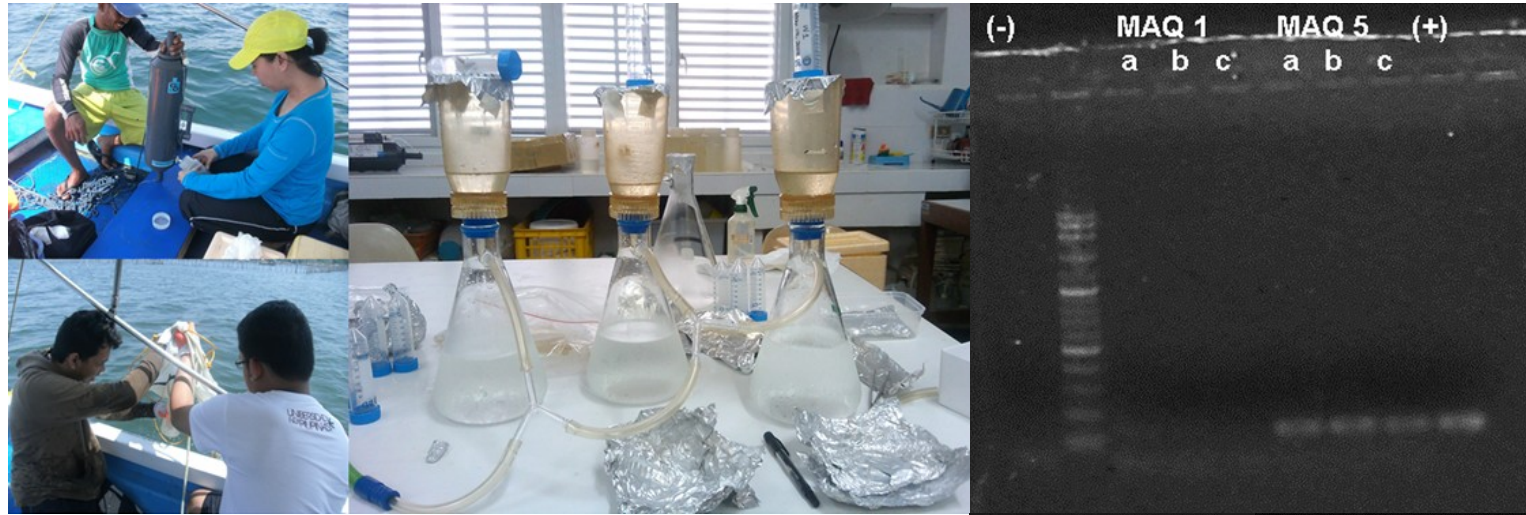
HAB Proteomics: insights into toxicity
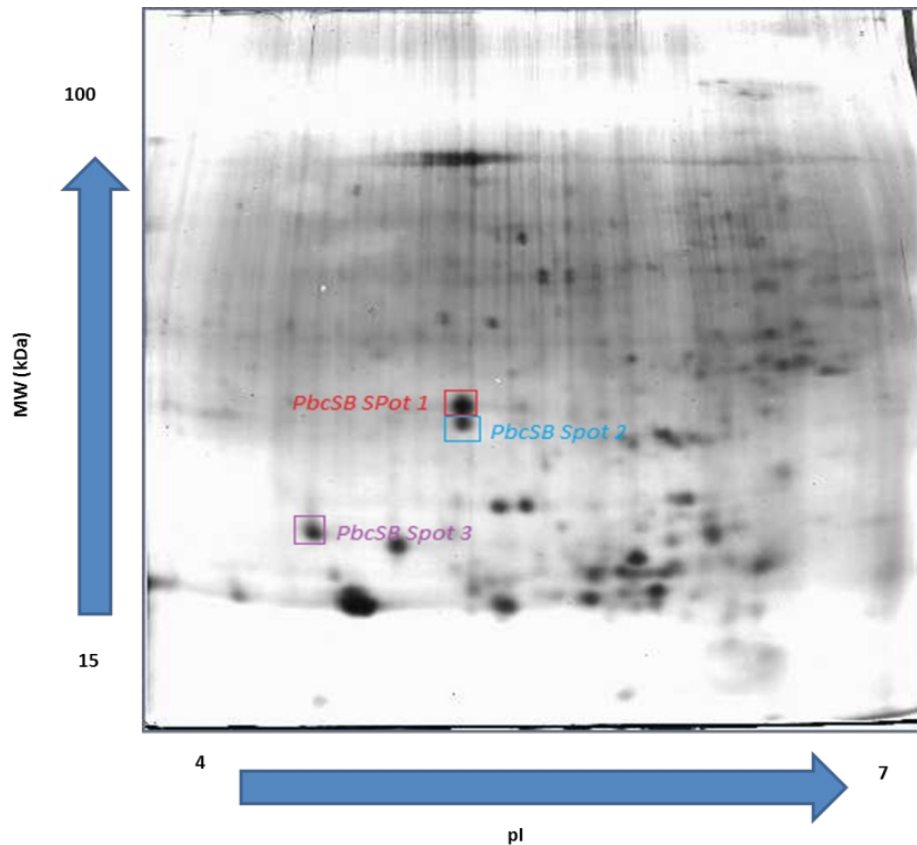
- Identification of proteins upregulated during high toxicity of *Pyrodinium*

- Identification of peptide transporter, transcription regulator, and metabolic enzyme

  - role in toxin production

On-field detection of *Pbc* using DNA amplification techniques

Developed two techniques for DNA-based detection of *Pbc:* PCR and LAMP

Testing of the PCR technique on actual field samples yielded positive results
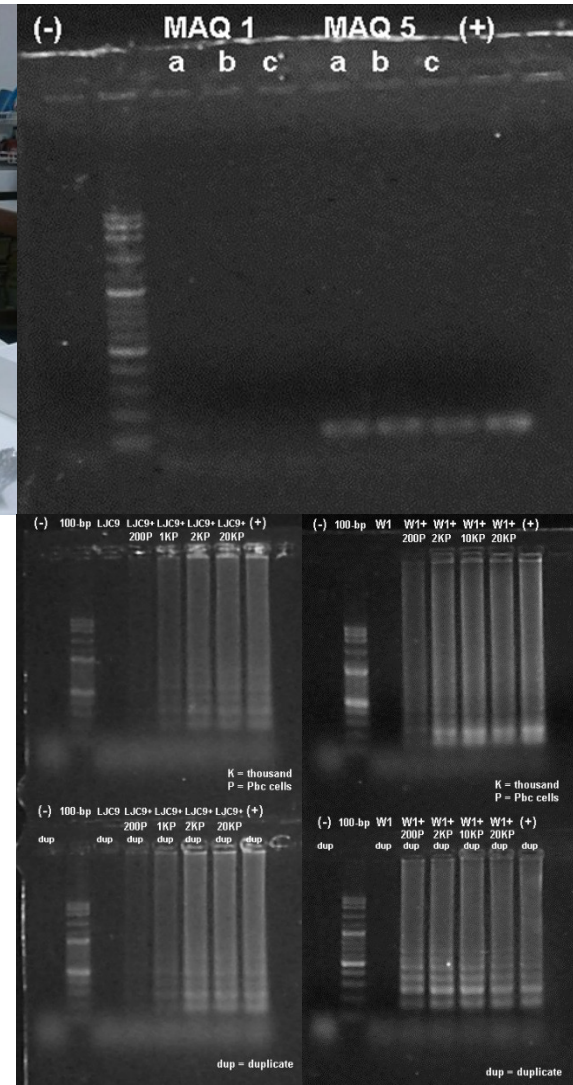
Summary

- partial genome sequence yields multiple types of data – bacterial symbionts, genes involved in various metabolic pathways, genes involved in saxitoxin biosynthesis

- hypothesis that the saxitoxin biosynthetic pathway in dinoflagellates is similar to that in cyanobacteria is partially supported

- similarity of putative *sxt* genes to bacterial homologs highlights the need to identify source of each gene (host vs. symbiont) and to evaluate role of endosymbionts in saxitoxin production

- genomic, proteomic, and molecular genetic approaches can provide us with tools to monitor harmful algae and their toxicity in the environment

Lab members who did the work

- David Gonzales

- Johanna Munar

- Kent Canlas

- Bryan Subong

- Deo Onda

Collaborators

- Dr. Rhodora Azanza, Program Leader, PhilHABs and HAB Genomics Programs

Acknowledgements

- DOST – PCAARRD