The genomics of niche adaptation in the harmful alga *Aureococcus anophagefferens*

Sonya Dyhrman
Woods Hole Oceanographic Institution
Outline

• **Introduction**
  - Linking genomics to an environmental niche
  - Harmful algal blooms
  - *A. anophageffens* as a model system
  - Key questions

• Genome - *A. anophageffens* a work in progress

• Transcriptome - Serial Analysis of Gene Expression
  - *Introduction to SAGE*
  - *N and P stress in A. anophageffens*

• Biome - Physiological ecology of *A. anophageffens* in Long Island Sound
  - *Introduction to the approach*
  - *Preliminary results*

• Conclusions
From genomes to biomes

Predicting microbial responses

Genome  Transcriptome  Physiology and biogeochemistry

N fixation  C fixation  OM degradation
In the ocean, life was carnivorous and simple, a pyramid founded on the very simplest of forms the phytoplankton, which teemed in great colored tides wherever sunlight met raw materials.

*Earth* - David Brin
Two modes of harmful impacts

- Toxic effects, from toxin production and accumulation
- Non-toxic effects, typically from high biomass

High biomass bloom

Low biomass toxic event
Different toxic HAB impacts

- Neurotoxic Shellfish Poisoning (NSP)
- Amnesic Shellfish Poisoning (ASP)
- Diarrhetic Shellfish Poisoning (DSP)
- Azaspiracid Shellfish Poisoning (AZP)
- *Pfiesteria* complex
- Ciguatera Fish Poisoning (CFP)
- Paralytic Shellfish Poisoning (PSP)

- Freshwater toxins

Karenia  Pseudonitzschia  Pfiesteria  Alexandrium

NSP  ASP  *Pfiesteria* complex  PSP

A nutritious meal of mussels can cause illness and even death when algal toxins are present.
Non-toxic effects

- Non-toxic effects, typically from high biomass
  - Reduced light
  - Oxygen depletion (hypoxia or anoxia)
  - Economic impacts
- Example - Brown tide
Brown tide

• **Aureococcus anophagefferens**
  – Small cell from the Pelagophyte group of algae
  – Non-motile
  – No known toxin production
• Commonly blooms in the mid-Atlantic states
  – permanently destroyed important shellfisheries in Long Island Sound.
• 2007 one of the worst blooms to date in NY
• **Aureococcus** also blooms in Africa
• Blooms at very high cell density relative to other HAB

1 http://ccmp.bigelow.org
2 http://siddall.info/fifteen/bt.html
3 http://www.scottsbt.com/misc/enviro/browntide.html

Modified from: http://deathstar.rutgers.edu/projects/btide/
Brown tide
Aureococcus bloom characteristics

- Top down
  - Tolerant to low light
  - Not heavily grazed
- Bottom up
  - N: Found in low nitrate high DON environments
  - P: Able to grown on DOP
  - Vitamins
  - Trace elements
- Prediction is difficult
- What is the Aureococcus niche?

Figure 1. Dynamics of DON (open circles), nitrate (open squares), Aureococcus cells (closed squared) in Flanders Bay, NY, 1995.

Courtesy Chris Gobler
HAB genomics

- Growing sequence databases for toxic dinoflagellates
  - *Karenia*
  - *Alexandrium*
- Genome in progress for a toxic diatom
- *Aureococcus* is the first completed whole genome sequence for a HAB species!

1. Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP)
Niche adaptation in *Aureococcus*

- What does the *Aureococcus* genome tell us about its niche?
  - Molecular understanding of N and P metabolism
- How does *Aureococcus* respond to N and P stress?
- Can the genome be linked to bloom dynamics?
Outline

What does the *Aureococcus* genome tell us about its niche?

- **Introduction**
  - Linking genomics to an environmental niche
  - Harmful algal blooms
  - *A. anophageffens* as a model system
  - Key questions

- **Genome - *A. anophageffens* a work in progress**

- **Transcriptome - Serial Analysis of Gene Expression**
  - Introduction to SAGE
  - *N and P stress in A. anophageffens*

- **Biome - Physiological ecology of *A. anophageffens* in Long Island Sound**
  - Introduction to the approach
  - Preliminary results

- Conclusions
Aureococcus anophagefferens strain 1984 genome

- Isolated in 1984
- Long Island Sound
- DOE/JGI has sequenced 100,000 ESTs and a draft of the genome
- The genome size is ~56 MB
- The assembly resulted in 1,185 scaffolds at 7X coverage
- JGI annotation pipeline predicts 11,501 genes

Graphics: CCMP, JGI
<table>
<thead>
<tr>
<th>Organism Type</th>
<th>Species/Genomes</th>
<th>Authors</th>
<th>Date</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diatom</td>
<td><em>Thalassiosira</em> (genome, EST)</td>
<td>Armbrust et al.</td>
<td>2004</td>
<td>34MB</td>
</tr>
<tr>
<td>Prasinophyte</td>
<td><em>Ostreococcus</em> (genomes)</td>
<td>Palenik et al.</td>
<td>2007</td>
<td>12MB</td>
</tr>
<tr>
<td>Dinoflagellate</td>
<td><em>Karenia</em> (EST)</td>
<td></td>
<td></td>
<td>3,000 - 215,000MB</td>
</tr>
<tr>
<td>Dinoflagellate</td>
<td><em>Alexandrium</em> (EST)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pymnesiophyte</td>
<td><em>Emiliania</em> (genome, EST)</td>
<td>Read and Wahlund</td>
<td></td>
<td>165MB</td>
</tr>
<tr>
<td>Pelagophyte</td>
<td><em>Aureococcus</em> (genome, EST)</td>
<td>Gobler and Wilhelm</td>
<td></td>
<td>56MB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited genomic infrastructure</td>
<td>relative to marine cyanobacteria</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What makes *Aureococcus* a good competitor?

- **Light?**
  - Look for evidence of
    - Light harvesting complexes

- **Grazing deterrents?**
  - Look for evidence of
    - Natural product biosynthesis

- **Nitrogen?**
  - Look for evidence of
    - Refractory and labile DON hydrolysis
    - DIN uptake

- **Phosphorus?**
  - Look for evidence of
    - DOP hydrolysis
    - DIP uptake
    - P storage

---

Mirza and Armbrust 2007 Oceanography
Closest whole genome matches from *Phytophthora* and *Ostreococcus*
<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pectate lyase</td>
<td>70968</td>
<td>14504</td>
</tr>
<tr>
<td>Nitrilase hydratase</td>
<td>59241</td>
<td>14355</td>
</tr>
<tr>
<td>Ferrodoxin dependant nitrite reductase</td>
<td>10808</td>
<td>12289</td>
</tr>
<tr>
<td>Nitrite reductase</td>
<td>71105</td>
<td>3423</td>
</tr>
<tr>
<td>Nitrite reductase</td>
<td>77792</td>
<td>2331</td>
</tr>
<tr>
<td>NADPH dependant nitrite reductase</td>
<td>37238</td>
<td>4684</td>
</tr>
<tr>
<td>Nitrate reductases</td>
<td>53391</td>
<td>11818</td>
</tr>
<tr>
<td>Acetamidase/Formidase</td>
<td>60068</td>
<td>31049</td>
</tr>
<tr>
<td>Acetamidase/Formidase</td>
<td>37987</td>
<td>3830</td>
</tr>
<tr>
<td>Argininosuccinate synthase</td>
<td>26092</td>
<td>5262</td>
</tr>
<tr>
<td>Asparaginase (asparaginase/glutaminase)</td>
<td>25391</td>
<td>15684</td>
</tr>
<tr>
<td>Similar to Asparaginase/glutaminase</td>
<td>66990</td>
<td>160</td>
</tr>
<tr>
<td>Proline dehydrogenase</td>
<td>25925</td>
<td>3845</td>
</tr>
<tr>
<td>Proline dehydrogenase</td>
<td>51906</td>
<td>1945</td>
</tr>
<tr>
<td>L protein (dihyrdolipoamide dehydrogenase)</td>
<td>25795</td>
<td>2491</td>
</tr>
<tr>
<td>L protein (dihyrdolipoamide dehydrogenase)</td>
<td>26536</td>
<td>3873</td>
</tr>
<tr>
<td>Glycine cleavage system P-protein</td>
<td>59391</td>
<td>23168</td>
</tr>
<tr>
<td>Cystathionine Beta Lyase</td>
<td>58595</td>
<td>9296</td>
</tr>
<tr>
<td>Histdidine Ammonia Lyase</td>
<td>22572</td>
<td>17852</td>
</tr>
<tr>
<td>Amidase</td>
<td>28333</td>
<td>98</td>
</tr>
<tr>
<td>Asparaginase</td>
<td>11569</td>
<td>15684</td>
</tr>
<tr>
<td>Aliphatic amidase</td>
<td>28333</td>
<td>98</td>
</tr>
<tr>
<td>Cathepsin A; CATA</td>
<td>52916</td>
<td>26824</td>
</tr>
<tr>
<td>Nitrilase-like protein 3</td>
<td>59241</td>
<td>14355</td>
</tr>
<tr>
<td>Formamidase, AMIF</td>
<td>60068</td>
<td>31049</td>
</tr>
<tr>
<td>Cyanase</td>
<td>60366</td>
<td>2141</td>
</tr>
<tr>
<td>Peptidase M20</td>
<td>58971</td>
<td>19184</td>
</tr>
</tbody>
</table>
Armbrust et al. 2004 *T. pseudonana*
What makes *Aureococcus* a good competitor?

- **Light?**
  - Look for evidence of
    - Light harvesting complexes

- **Grazing deterrents?**
  - Look for evidence of
    - Natural product biosynthesis

- **Nitrogen?**
  - Look for evidence of
    - Refractory and labile DON hydrolysis
    - DIN uptake

- **Phosphorus?**
  - Look for evidence of
    - DOP hydrolysis
    - DIP uptake
    - P storage

Some unusual redundancy:
- Chitobiose, Formamide, Pectin, Cyanate, Urea, etc.
- Nitrate
- Ammonium
# Comparative genomics

<table>
<thead>
<tr>
<th>KEGG Metabolic Pathways</th>
<th>models in <em>Auripagatus</em> anophaga/keraea Filters/Models3 (var 1)</th>
<th>models in <em>Ostreococcus luorii</em> v2.0 Filters/Models (var 1)</th>
<th>models in <em>Ostreococcus lucimarinus</em> v2.0 Filters/Models (var 1)</th>
<th>models in <em>T. pseudonana</em> v3.0 Filters/Models (var 1)</th>
<th>models in <em>P. tricornutum</em> v2.0 Filters/Models2 (var 1)</th>
<th>models in all selected model sets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amino Acid Metabolism</strong></td>
<td>723</td>
<td>292</td>
<td>264</td>
<td>595</td>
<td>568</td>
<td>2448</td>
</tr>
<tr>
<td>Alanine and aspartate metabolism</td>
<td>33</td>
<td>25</td>
<td>23</td>
<td>34</td>
<td>36</td>
<td>151</td>
</tr>
<tr>
<td>Arginine and proline metabolism</td>
<td>52</td>
<td>16</td>
<td>18</td>
<td>53</td>
<td>48</td>
<td>187</td>
</tr>
<tr>
<td>Cysteine metabolism</td>
<td>13</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>Glutamate metabolism</td>
<td>42</td>
<td>23</td>
<td>18</td>
<td>41</td>
<td>32</td>
<td>156</td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td>61</td>
<td>31</td>
<td>26</td>
<td>52</td>
<td>50</td>
<td>220</td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>54</td>
<td>24</td>
<td>20</td>
<td>44</td>
<td>44</td>
<td>186</td>
</tr>
<tr>
<td>Lysine biosynthesis</td>
<td>34</td>
<td>16</td>
<td>13</td>
<td>27</td>
<td>26</td>
<td>116</td>
</tr>
<tr>
<td>Lysine degradation</td>
<td>66</td>
<td>18</td>
<td>18</td>
<td>56</td>
<td>54</td>
<td>212</td>
</tr>
<tr>
<td>Methionine metabolism</td>
<td>13</td>
<td>9</td>
<td>9</td>
<td>15</td>
<td>13</td>
<td>59</td>
</tr>
<tr>
<td>Phenylalanine metabolism</td>
<td>23</td>
<td>13</td>
<td>16</td>
<td>25</td>
<td>28</td>
<td>106</td>
</tr>
<tr>
<td>Phenylalanine, tyrosine and tryptophan biosynthesis</td>
<td>22</td>
<td>24</td>
<td>20</td>
<td>25</td>
<td>29</td>
<td>120</td>
</tr>
<tr>
<td>Tryptophan metabolism</td>
<td>168</td>
<td>23</td>
<td>18</td>
<td>72</td>
<td>57</td>
<td>338</td>
</tr>
<tr>
<td>Tyrosine metabolism</td>
<td>61</td>
<td>29</td>
<td>27</td>
<td>47</td>
<td>56</td>
<td>209</td>
</tr>
<tr>
<td>Urea cycle and metabolism of amino groups</td>
<td>27</td>
<td>13</td>
<td>13</td>
<td>24</td>
<td>23</td>
<td>100</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine biosynthesis</td>
<td>23</td>
<td>8</td>
<td>7</td>
<td>24</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>Valine, leucine and isoleucine degradation</td>
<td>47</td>
<td>12</td>
<td>10</td>
<td>45</td>
<td>43</td>
<td>157</td>
</tr>
</tbody>
</table>
Outline

How does *Aureococcus* respond to N and P stress?

- **Genome**
  - Introduction
    - Linking genomics to an environmental niche
    - Harmful algal blooms
    - *A. anophageffens* as a model system
    - Key questions
  - Genome - *A. anophageffens* a work in progress

- **Transcriptome - Serial Analysis of Gene Expression**
  - Introduction to SAGE
  - *N and P stress in A. anophageffens*

- **Biome - Physiological ecology of *A. anophageffens***
  - Introduction to the approach
  - Preliminary results

- Conclusions
Global gene expression analyses

- Expressed sequence tags - EST
- Microarrays
- Serial analysis of gene expression (SAGE)
  - Velculescu et al. 1995 Science (14bp)
  - Long-SAGE Saha et al. 2002 Nature Biotech. (21bp)
- No prior knowledge of genome required
- Provides a relative measure of gene expression
- Relatively high throughput
- SAGE-based expression studies now increasing (Dinoflagellate: Coyne et al. 2004, Coccolithophore: Dyhrman et al. 2006).
- Digital gene expression illumina (21bp) (Diatoms: pending in Dyhrman lab).
Long-SAGE

Extract RNA → Isolate Long-SAGE tags → Concatenate and Sequence → Expression Frequency → Annotation
• Compare *Aurecoccus anophagefferens* from Control, N-starved and P-starved conditions.
Long-SAGE: Tag isolation and sequencing

Extract RNA → Isolate Long-SAGE tags → Concatenate and Sequence → Expression Frequency → Annotation

CATG

CATGAACGACCTCATCTCCGA
Long-SAGE: tag isolation

cDNA synthesis

*Nla* III anchoring enzyme digestion (AE)

Divide sample

- Linker ligation
- *BsmF* I tagging enzyme digestion (TE)

Ditag synthesis

PCR amplify ditags &

*Nla* III ditag release

Primer A

Primer B
Long-SAGE: Tag isolation and sequencing

Extract RNA → Isolate Long-SAGE tags → Concatenate and Sequence → Expression Frequency → Annotation

CATG

CATGAACGACCTCATCTCCGA
Long-SAGE library characteristics

- **Singletons**: only one copy in one library
- Only the most highly expressed tags have been sampled

<table>
<thead>
<tr>
<th>Library</th>
<th>Total tags sampled</th>
<th>Total tags used in analyses</th>
<th>Total unique tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>+P</td>
<td>46,321</td>
<td>35,577</td>
<td>7,151</td>
</tr>
<tr>
<td>-P</td>
<td>31,356</td>
<td>24,440</td>
<td>6,353</td>
</tr>
<tr>
<td>-N</td>
<td>34,810</td>
<td>26,205</td>
<td>5,914</td>
</tr>
</tbody>
</table>
Long-SAGE: expression frequency

- Extract RNA
- Isolate Long-SAGE tags
- Concatenate and Sequence
- Expression Frequency
- Annotation
Long-SAGE: cluster analysis

A: Tags that are up-regulated under -P conditions

B: Tags that are up-regulated under -N conditions

C: Tags that are up-regulated under +P conditions

- Green: Up-regulated
- Red: Down-regulated
Long-SAGE in other algae: comparing N v. P stress

**Emiliana**

- N  Control  -P

55,000 tags /160Mb

**Aureococcus**

- P  Control  -N

113,000 tags/56Mb
Long-SAGE: tag annotation

Extract RNA → Isolate Long-SAGE tags → Concatenate and Sequence → Expression Frequency → Annotation
Long-SAGE: tag to genome mapping
• Total Unique Tags (including singletons): 31,862
  • Direct (100%) tag to gene maps
• Map to genome: 11,847 (37.2%)
  • Overlap with gene models: 4,709 (39.74%)
• Map to ESTs: 12,359 (38.8%)
• Gene modeling in the eukaryotic algae is very challenging
  • Tags provide expression support for gene models
  • Help to identify regions of the genome where we missed a gene
Long-SAGE: tag to predicted gene function

- **-P**
  - Control
  - -N

**High affinity P transporter**
- 5’-nucleotidase
- Esterase/lipase/thioesterase
- Conserved hypotheticals

**DIP and DOP**

**Ammonium transporter**
- Urea transporter
- Peptidase
- Formidase
- Conserved hypotheticals

**DIN and DON**

**Amino acid and nucleotide biosynthesis**
- Oxidoreductase
- Protein folding genes
Gene expression patterns

5'-nucleotidase

Asparaginase/glutaminase

Acetamidase/Formidase
Expression of an ammonia transporter (amt1) is detectable in N starved cultures (-N)
Introduction
  - Linking genomics to an environmental niche
  - Harmful algal blooms
  - A. anophageffens as a model system
  - Key questions

Genome - A. anophageffens a work in progress

Transcriptome - Serial Analysis of Gene Expression
  - Introduction to SAGE
  - N and P stress in A. anophageffens

Biome - Physiological ecology of A. anophageffens in Long Island Sound
  - Introduction to the approach
  - Preliminary results

Conclusions
Defining the niche - physiological ecology in the field

Figure 1. Dynamics of DON (open circles), nitrate (open squares), *Aureococcus* cells (closed squared) in Flanders Bay, NY, 1995.
Defining the niche - physiological ecology in the field

<table>
<thead>
<tr>
<th>Physiological role:</th>
<th>Gene target:</th>
<th>Support:</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIP metabolism</td>
<td>High affinity phosphate transporter</td>
<td>Long-SAGE/RT-PCR</td>
</tr>
<tr>
<td>DOP metabolism</td>
<td>5′ nucleotidase</td>
<td>Long-SAGE</td>
</tr>
<tr>
<td>DIN metabolism</td>
<td>Ammonium transporter, Nitrate transporter, Nitrite transporter</td>
<td>Long-SAGE/RT-PCR, None, None</td>
</tr>
<tr>
<td>DON metabolism</td>
<td>Urea transporter, Amino acid transporter</td>
<td>Long-SAGE/RT-PCR, None</td>
</tr>
<tr>
<td>House-keeping</td>
<td>Beta-tubulin, Actin</td>
<td>Long-SAGE/RT-PCR</td>
</tr>
</tbody>
</table>

![Graph showing N uptake over time](chart.png)
Preliminary “bloom phenotype” data from 2007

Water
→ Collect 0.5-0.75L onto GF/F
→ Store in RNA extraction buffer
→ Examine gene expression

Figure 1. Dynamics of DON (open circles), nitrate (open squares), *Aureococcus* cells (closed squared) in Flanders Bay, NY, 1995.

Courtesy Chris Gobler
Preliminary “bloom phenotype” data

Culture experiments show that expression of this urea transporter is correlated with N availability.

Quantuck Bay (7-9-07): provided by Chris Gobler
Outline

- Introduction
  - Linking genomics to an environmental niche
  - Harmful algal blooms
  - A. anophageffens as a model system
  - Key questions

- Genome - A. anophageffens a work in progress

- Transcriptome - Serial Analysis of Gene Expression
  - Introduction to SAGE
  - N and P stress in A. anophageffens

- Biome - Physiological ecology of A. anophageffens in Long Island Sound
  - Introduction to the approach
  - Preliminary results

- Conclusions
Conclusions

• The *Aureococcus* genome has given us new insights into its niche
  - *Top-down: low light, grazing deterrents*
  - *Bottom-up: DON, DOP, redundant inorganic transporters*

• Long-SAGE is a mechanism for examining transcriptional level responses in HAB models.
  - *Support gene models*
  - *Identify differentially regulated genes*
  - *Assign regulation patterns to genes with no database homology and unknown function*

• *Aurecococcus* has a strong transcriptional response to nutrient deficiency.
  - *Up-regulation of many genes involved in DIN, DON, DIP and DOP metabolism*
Conclusions

- Knowing the molecular underpinnings of N, and P metabolism will improve our understanding of the physiological ecology of this HAB group.

- Preliminary data from the Long Island bloom and shows promise with regards to the approach and our understanding of HAB physiological ecology.

How do we increase ocean literacy and teach the process of scientific inquiry?
Outreach

• How do you engage kids (8-12) with scientific content on a broad scale? - Whyville, a virtual world for kids www.whyville.net
Whyville is Seriously Fun!

Founded in 1999
Virtual world for kids 8-15
3.3 million registered users
Over 10 million user hours/month
Girls outnumber boys 2 to 1
Blooms and nutrients activity launches

WhOI Microscope

As a marine biologist at WhOI, your job is to figure out which type of phytoplankton is blooming in the ocean water near Whyville.

1. Drag the circular microscope field around to look at the organisms on your slide.

2. Consult the Culture Collection below to discover which type of phytoplankton is in bloom and what nutrient(s) it relies on to grow.

Once you have an idea, go back to the lab and choose a sensor to trace the nutrient to its source. Click here for more detailed instructions.
• How do you engage kids (8-12) with scientific content on a broad scale?
  – Whyville, a virtual world for kids
  – Whyville Oceanographic Institution (WhOI) is born

• To date the bloom activity has had over 200,000 different participants -
  making this a powerful mechanism for teaching about real world
  environmental issues and the process of scientific discovery.

• New microbial oceanography activity to launch Fall 2008 (CMORE)
  – Marine microbes are numerous!
  – Marine microbes are diverse!
  – Marine microbes are beneficial!
Acknowledgements

- Louie Wurch
- Sheean Haley
- Liz Orchard
- Chris Gobler (Stony brook)
- *Aureococcus* Genome Consortium
- Numedon Inc.

- DOE, JGI
- Earth Institute (Columbia)
- Ocean Life Institute (WHOI)
- CMORE
- NSF Environmental Genomics
- Woods Hole Center for Oceans and Human Health