Dinoflagellate microevolution from species to populations

L. K. MEDLIN
Marien Biological Association of the UK, the Citadel, Plymouth PL1 2PB, UK
(e-mail: lkm@mba.ac.uk)

Abstract: Dinoflagellates belong to one of the most strongly supported macrolineages among the protists, the superphylum/kingdom Alveolata. This superphylum contains three phyla, the Dinoflagellata, the Apicomplexa and the Ciliata, all of which have cortical alveoli and micropores. As more taxa have been added to the molecular analyses and/or more genes have been used, our understanding of phylogenetic relationships within the dinoflagellates has changed and this is also true at the species level. Notwithstanding the fact that most analyses have focused at the genus level, cryptic species have been uncovered in some cases at the species level. Below species level however, microsatellites have shown immense spatial fragmentation in dinoflagellate populations. Expressed sequence tags (EST) libraries have been constructed for several species to try to understand toxin gene expression in relationship to stress conditions. Initial annotation results indicated that a low percentage of the genes could be annotated. Recent advances have pushed this level to nearly 29% of ESTs. Some common responses to stress conditions have been noted. Toxin genes been only been identified in EST libraries that have been screened by 454 sequencing Illumina. The complete genome of *Heterocapsa circularisquama* is near completion.

For the past 20 years, understanding and preserving biodiversity has been one of the most important global challenges and will continue to be an important scientific issue in the coming decades. The global environment is experiencing rapid and accelerating changes, largely originating from human activity. Whether these changes in the global environment come from local influences or from the more dispersed effects of global climate change, biodiversity is strongly modified by them.

It is estimated that less than 10% of the known biodiversity in the marine protistan community is known, but among the pico-fraction even less is known with new groups being discovered regularly (Kim *et al.* 2011). Medlin & Kooistra (2010) reviewed existing evidence to suggest that, in the marine environment, the marine protistan community is more genetically diverse below species level than the freshwater community. Many cosmopolitan species, which we think we can easily recognize, are found in both communities and are now being shown to be species complexes with little or no morphological characters to separate them. Spatial and temporal variation in the abundance and distribution of these species complexes are also unknown. With new molecular and analytical techniques, our knowledge of marine biodiversity at and below the species level begins to unfold enabling us to understand how marine biodiversity supports ecosystem structure, dynamics and resilience. With these techniques, our understanding of biodiversity and ecosystem dynamics in all areas of the planktonic community, not just the photosynthetic ones, can be augmented.

Some of the findings for hidden and cryptic diversity in the dinoflagellates are addressed in this paper and it is shown how genetic diversity studies have revealed their populations to be highly fragmented. A summary of what is known from gene expression studies to show how dinoflagellates respond to stress conditions is provided.

Evolution at the species level or cryptic diversity

Molecular studies in general have wreaked havoc with the species level taxonomy throughout the eukaryotic world, and dinoflagellates have not been an exception. Many cryptic species have been discovered, particularly among non-thecate taxa, because relationships are difficult to ascertain morphologically. In the Suessiales, eight clades of symbiotic dinoflagellates have been recovered (Pouchon *et al.* 2006) in contrast to the single species *Symbiodinium microadriaticum*, originally described from corals (Freudenthal 1962). Analyses of the *Alexandrium tamarense/fundyense/catenella* species complex have shown that these taxa are related by their geographic origin and not by their morphological features (Scholin *et al.* 1994; John *et al.* 2003). Each of these geographic clades is sufficiently distinct genetically to be considered a separate species, but the dinoflagellate community has been reluctant to formally separate the clades into species because of the absence of clear morphological characters to separate the clades.
thus making identification with light microscopy problematic. With further sequencing effort from more global isolates, it would appear that some of the geographic clades, such as the toxic North American clade, have a more world-wide distribution (Lilly et al. 2007). A scenario for the historical biogeographic history of this complex from a single global ancestor into the various geographic clades, and hence speciation, has been proposed by John et al. (2003) using a molecular clock. The closing of the Tethys Sea is hypothesized to have caused the divergence of the non-toxic Western European clade from the toxic Temperate Asian and this would explain the presence of relicts of this clade in the Mediterranean Sea. The closing of the Isthmus of Panama separated the non-toxic Mediterranean clade from the toxic North American clade. The Mediterranean clade became extinct in western tropical Atlantic waters, but it still thrives in the Mediterranean where it is endemic. Major geological events can therefore be related to the divergence of the geographic clades/groups into distinct species (Fig. 1).

Both true cryptic species and pseudogenes are present in dinoflagellates. In Scrippsiella trochoidea, several unique haplotypes have been found and identical morphotypes were recovered in different clades (Montresor et al. 2003). In contrast, Gribble and Anderson (2007) suggested that the variation at the intraspecific level in Protoperidinium and Diplopsalis species were primarily caused by the presence of pseudogenes. At the species level, cryptic diversity is often reconfirmed with data from other genes after first being recovered with sequence data from ribosomal genes. The rate of evolution in protein coding genes v. 18S ribosomal genes using whole genomes has been analysed by Piganeau et al. (2011). They found that the rate of evolution in these genes between unicellular and multicellular organisms is different and the protein genes evolve about five times faster than the ribosomal genes in unicellular organisms than in multicellular organisms. Piganeau et al. (2011) considered that this discrepancy in the rate of evolution between the ribosomal and protein genes was related to differences in effective population sizes, with larger population sizes (often proposed as metapopulations) of unicellular organisms having a relatively faster rate of evolution in protein genes. If this is true then we could find the same differences between dinoflagellate lineages that have two different population sizes, thus behaving effectively as unicellular or multicellular organisms, and these differences might be related to their effective population sizes. There are two different endosymbiotic lineages among the Suessiales, each with a different rate of evolution judging from the branch lengths in the trees given by Shaked & de Vargas (2006), and each lineage has a very different effective population size (Fig. 2). The dinoflagellate endosymbionts found in planktonic foraminifera have an effectively large population size. They therefore evolve much more slowly than the endosymbionts within corals, which have a much smaller effective population size, with some lineages being endemic to a single coral head. Piganeau et al. (2011) considered that because of the different rates of evolution, biodiversity in unicellular organisms is underestimated and that a single base difference in the 18S rRNA gene denotes a different species.

![Fig. 1. Map of the current distribution of the major clades within the Alexandrium tamarense/catenella/fundyense species complex. Large arrows point to the geological events (rising of the Isthmus of Panama and the closing of the Tethys Sea), which can be correlated to the divergence marked with a smaller arrow in the tree. Redrawn from John et al. (2003).](image-url)
Evolution below the species level genetic diversity

Among the fingerprinting methods available for measuring genetic diversity below the species level, microsatellites (short repeated sequences that are 1–6 nucleotides long, e.g. GTGTGTGT) are the most computationally intensive. They are codominantly inherited markers that provide very high levels of heterozygosity and allow gene flow from one population to another to be measured. Table 1 shows the dinoflagellate species for which microsatellites have been established. Each species has a distinct population structure and reduced gene flow between close populations. Some of the more interesting findings among the dinoflagellate populations with these markers are discussed below.

In Symbiodinium, Thornhill et al. (2009) have shown that population structure of the endosymbiont is dependent on how the endosymbiont is inherited as the coral reproduces. In horizontal inheritance of the endosymbiont, the coral larvae take their endosymbionts from the plankton. In vertical inheritance of endosymbionts, the coral larvae inherit their endosymbionts from the parent coral. In Symbiodinium sp. clade B with horizontal inheritance, there is host within host and even reef endemism; endosymbionts have a strong and temporally stable population structure. The endemism in Clade B could also be the result of the host coral expelling other genotypes to maintain a specific population. In contrast, in Symbiodinium sp. clade C with vertical inheritance, endosymbionts have no population structure.

Fig. 2. Phylogenetic tree showing the relationships and the rates of evolution (branch lengths) in the benthic and planktonic clades of foraminifera. Benthic clades with smaller population sizes are evolving faster and planktonic clades with larger population sizes are evolving slower. Redrawn from Shaked & de Vargas (2006).
As mentioned above, the *Alexandrium tamarense/fundyense/catenella* species complex is composed of genetically distinct clades which were originally delineated by geographic origin. The origin of the Mediterranean isolates of the Temperate Asian (TA) of the *Alexandrium tamarense/fundyense/catenella* species complex was first inferred with sequence data but later clarified with microsatellites. Using large subunit (LSU) rRNA sequence data to compare Mediterranean and Asian TA strains, Lilly et al. (2002) concluded that the presence of TA strains in the Mediterranean was a clear case of ballast water introduction. Using internal transcribed spacer (ITS) sequence data, Penna et al. (2005) also concluded that it was probably a case of ballast water introduction. Using microsatellites, Masseret et al. (2009) revealed that the Mediterranean populations were clearly distinct and distant from the Asian strains. However, they could not offer any reasonable explanation for their occurrence in the Mediterranean. Using the historical biogeography of the species complex proposed by John et al. (2003) (see above), the closing of the Tethys Sea, which is the vicariant event separating the West European strains from the Temperate Asian strains, possibly entrapped TA genotypes in the Mediterranean; they would therefore represent a relict population from when the two groups were once joined. In this scenario TA genotypes have probably always been present in French Lagoons; however, it is hypothesized here that a change in environmental conditions has caused them to bloom and be noticed.

Alpermann (2009) conducted the first hierarchical study of the genetic diversity of populations of *Alexandrium tamarense* NA clade in the Orkney Islands, UK. On a global scale, the Orkney Islands populations of the NA clade are not the most closely related to populations on the east coast of North America. Medlin et al. (1998) suggested that these populations entered from the Pacific and moved along coastal pathways of the eastern side of North America until they reached the Gulf Stream, which carried them across to the Orkney Islands. In contrast, they are most closely related to Pacific Isolates from Japan, which suggests that these populations were introduced to the region by cells directly coming across the Arctic Ocean. In a more detailed study of one location in the Orkney Islands, Alpermann (2009) found four populations that could interbreed and these populations were determined to be different year classes that had hatched from local cyst beds.

### Table 1. Summary of microsatellite studies in dinoflagellates

<table>
<thead>
<tr>
<th>Species</th>
<th>Establishment</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Alexandrium tamarense</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Alexandrium tamarense</em></td>
<td>Nagai et al. (2006a), Nishitani et al. (2007a)</td>
<td></td>
</tr>
<tr>
<td>fundyense/catennella TA</td>
<td>Nagai et al. (2006b)</td>
<td></td>
</tr>
<tr>
<td><em>Alexandrium minutum</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Coccolithus polykrikoides</em></td>
<td>Nishitani et al. (2007b)</td>
<td></td>
</tr>
<tr>
<td><em>Heterocapsa circularisquama</em></td>
<td>Nagai et al. (2007a)</td>
<td></td>
</tr>
<tr>
<td><em>Symbiodinium sp. Clades B and C</em></td>
<td>Santos &amp; Coffroth (2003), Bay et al. (2009)</td>
<td></td>
</tr>
<tr>
<td><em>Karenia brevis</em></td>
<td>Renshaw et al. (2006)</td>
<td></td>
</tr>
</tbody>
</table>

### Evolution of dinoflagellate genes and their transcription

The structure of the dinoflagellate nucleus has long been known to be unusual. The chromosomes are permanently condensed and the nucleus contains 2–200 pg/cell of DNA as compared to ≤0.5 pg/cell in other algae. There are many copies of each gene with long stretches of non-coding regions, but gene length may not be very different from that of other organisms (Triplett et al. 1993; Jaekisch et al. 2011). All copies of each gene were transcribed on a single mRNA. It is not known if the long stretches of non-coding regions are between clusters of multiple copies of a gene or between each copy of a gene cluster. If they are between each copy of a gene, then it will be more difficult for crossing over to take place during meiosis. It will also be more difficult to transcribe all of the multiple genes into a single mRNA transcript. Recycling of mRNAs seems to be a feature typical of dinoflagellates (Slamovits & Keeling 2008). Jaekisch et al. (2011) found for *Alexandrium ostenfeldii* that, in part, the genome has a relatively uniform sequence space interrupted by
coding sequences. The repetitive sequences in *A. ostenfeldii* were mainly constituted of large tandem arrays of only four repeat types. In contrast, the plastid and the mitochondrial genomes are dramatically reduced. When the first genome of a dinoflagellate is completed, we will have a better understanding of the strange features of the dinoflagellate nucleus. The immense size of the nucleus has held back many investigators from tackling dinoflagellates as a genome sequencing project. Some attempts at isolating a single chromosome or long DNA fragments for sequencing has resulted in so much non-coding data that contig overlaps were impossible (Jeackisch *et al.* 2012; U. John, 2012, pers. comm.). *Heterocapsa circularisquama* has been targeted for genome sequencing by the Bhattacharya lab at Rutgers University, USA, primarily because its virus has had its genome completely sequenced (Nagasaki *et al.* 2005).

It has been shown by a number of expressed sequence tags (EST) studies that all dinoflagellates possess a common unique splice leader sequence (Lidie & Van Dolah 2007; Zhang *et al.* 2007). It is therefore possible to isolate all dinoflagellate mRNAs from any environmental sample. Using this technique, the presence of genes not known to exist in dinoflagellates has been documented from environmental clone libraries (Lin *et al.* 2010). Genes, such as alveolins, nucleosome core histones and rhodopsin (the latter gene implicated in non-photosynthetic solar energy capture) are now known to occur in dinoflagellates (Lin *et al.* 2010; Toulza *et al.* 2010). The annotation of ESTs has improved as sequencing capacities have increased and more whole genomes have been edited. Initial coverages of 9% (John *et al.* 2005) have now risen up to nearly 44% (Table 2).

Studies of responses to stressed conditions using gene expression analysis has shown in nearly every case that a reduction in the expression of photosynthetic genes is followed by up-regulation of stress proteins to protect the cells (Leggat *et al.* 2007; Yang 2009; Moustafa *et al.* 2009) with cultures containing bacteria significantly enhancing the above stress response of the dinoflagellates (Moustafa *et al.* 2009). Their results suggest that about 73% of the *Alexandrium tamarense* transcriptome comprises a ‘core’ component and 27% comprises the regulated component. These proportions change under differing cellular or environmental conditions, which are enhanced in xenic over axenic cultures. Microarray-based comparisons of toxic and non-toxic strains of *A. minutum* indicated that a greater number of genes were more highly expressed in the toxic strains rather than the non-toxic strains (Yang *et al.* 2010); 145 genes were identified as being more highly expressed in the toxic strains. Eight of these not significantly expressed in the non-toxic strain were identified as putative helicase sequences and six hypothetical proteins for which no annotation could be found. Yang *et al.* (2010) assumed that these

<table>
<thead>
<tr>
<th>Study</th>
<th>Species</th>
<th>Percentage coverage and identification of EST</th>
<th>Notable genes found</th>
</tr>
</thead>
<tbody>
<tr>
<td>John <em>et al.</em> (2005)</td>
<td><em>Alexandrium ostenfeldii</em></td>
<td>10% of 3000 ESTs</td>
<td>Polyketide genes</td>
</tr>
<tr>
<td>Liodie and van Dolah (2007)</td>
<td><em>Karenia brevis</em></td>
<td>29% of 7000 ESTs</td>
<td>Microarray generated SNPs in multiple copy genes</td>
</tr>
<tr>
<td>Hackett <em>et al.</em> (2005)</td>
<td><em>Alexandrium tamarense</em> (NA)</td>
<td>20% of 6723 ESTs</td>
<td>Histone genes</td>
</tr>
<tr>
<td>Leggat <em>et al.</em> (2007)</td>
<td><em>Symbiodinium sp.</em></td>
<td>44% of 1456 ESTs</td>
<td>Glycerol-phosphate:phosphate antiporter</td>
</tr>
<tr>
<td>Yang (2009), Yang <em>et al.</em> (2010)</td>
<td><em>Alexandrium minutum</em></td>
<td>28% of 3000 ESTs</td>
<td>Four stress libraries, toxic and non-toxic strains express different genes</td>
</tr>
<tr>
<td>Moustafa <em>et al.</em> (2009)</td>
<td><em>Alexandrium tamarense</em> (NA)</td>
<td>19% of 11 171 ESTs</td>
<td>Bacteria influence gene expression under stress conditions</td>
</tr>
<tr>
<td>Toulza <em>et al.</em> (2010)</td>
<td><em>Alexandrium tamarense</em> (TA)</td>
<td>24% 21 236 ESTs</td>
<td>Alveolin protein genes</td>
</tr>
<tr>
<td>Lin <em>et al.</em> (2010)</td>
<td>3 splice leader environmental clone libraries</td>
<td>20–27% of 1000 ESTs</td>
<td>Nucleosome histone core, modification and assembly genes, Rhodopsin</td>
</tr>
<tr>
<td>Jaeckisch <em>et al.</em> (2011)</td>
<td><em>Alexandrium ostenfeldii</em></td>
<td>15% of 12 287 ESTs</td>
<td>58% of genome is complex repeats</td>
</tr>
</tbody>
</table>
eight genes were likely candidates for genes associated with the biosynthesis or regulation of paralytic shellfish poisoning (PSP) toxins, or for adaptive responses to intracellular PSP toxins.

A putative entire operon for the saxitoxin genes has recently been recovered using 454 sequencing of EST libraries (Hackett et al. 2010). Three of the eight genes in this operon are of cyanobacterial origin, one is of eukaryotic origin and four are of unknown origin. Hackett et al. (2010) suggested that there has not been a lateral gene transfer from the Cyanobacteria into the dinoflagellates (see Murray et al. 2010). Because these same genes are not routinely recovered in lower sequencing capacity EST clone libraries (Hackett et al. 2005), this would suggest that the dinoflagellates have some unique control over the transcription and production of toxins and that they have still more secrets waiting to be discovered. In other studies carried out to find the putative saxitoxin genes in dinoflagellates, putative candidates were found for sxtA, sxtB, sxtF/M, sxtH, sxtI, sxtU (Stücker et al. 2010), sxtW, sxtZ, sxtT/H and sxtO (S. Wohlrab & U. John, 2012, pers. comm.). The putative saxitoxin gene candidates found in another study (R. Diwan & U. John, 2012, pers. comm.) were for the genes sxtA, sxtG, sxtU, sxtX, sxtH, sxtI, sxtO, sxtN and sxtT. Another exciting example of gene evolution is that of polyketide synthases in dinoflagellates. Phylogenetic analyses have shown that the polyketides synthase (PKS) genes are derived from the modular Type I group (Snyder et al. 2003; Monroe & van Dolah 2008); however, detailed analyses have shown that the modular structure is broken apart and that the single modules become single genes as known for PKS Type II.

**Conclusions**

Species level and genetic diversity in the dinoflagellates follow similar patterns to other eukaryotic organisms. Cryptic species are present where cosmopolitan species have been studied and, below species level, extensive fragmentation of oceanic populations has been recovered using microsatellite markers. Gene expression studies have recovered several unusual genes including those such as histones, not believed to exist in dinoflagellates. The presence of a dinoflagellate specific splice leader in all their mRNAs provides a unique advantage to studying gene expression in the field; this is not afforded by other members of the eukaryotic world except in the Euglenids (see accompanying paper on dinoflagellate macrolevolution; Medlin & Fensome 2013). The difficulty in recovering mRNA transcripts for the toxin genes in routine clone libraries suggest that the dinoflagellates have different mechanisms for enhancing transcription and translation rather than just by increasing gene expression. It would seem that we still have many things to learn about the dinoflagellates, such as how they have speciated, how their populations are interconnected in the oceanic environment and how they respond to environment cues to express the genes needed to survive. Hopefully the first genome sequence will help to unravel and solve many of their secrets.

This work was funded in part from EU MIDTAL contract no. 201724.

**References**


Hackett, J. D., Scheetz, Z., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L. & Bhattacharya,


