Population genetic structure and connectivity of the harmful dinoflagellate *Alexandrium minutum* in the Mediterranean Sea

Silvia Casabianca$^{1, †}$, Antonella Penna$^{1, *, †}$, Elena Pecchioli$^2$, Antoni Jordi$^3$, Gotzon Basterretxea$^3$ and Cristiano Vernesi$^2$

$^1$Department of Biomolecular Sciences, University of Urbino, Pesaro 61121, Italy
$^2$Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Trento 38040, Italy
$^3$IMEDEA (UIB-CSIC), Esorpes 07190, Spain

The toxin-producing microbial species *Alexandrium minutum* has a wide distribution in the Mediterranean Sea and causes high biomass blooms with consequences on the environment, human health and coastal-related economic activities. Comprehension of algal genetic differences and associated connectivity is fundamental to understand the geographical scale of adaptation and dispersal pathways of harmful microalgal species. In the present study, we combine *A. minutum* population genetic analyses based on microsatellites with indirect connectivity ($G_i$) estimations derived from a general circulation model of the Mediterranean sea. Our results show that four major clusters of genetically homogeneous groups can be identified, loosely corresponding to four regional seas: Adriatic, Ionian, Tyrrhenian and Catalan. Each of the four clusters included a small fraction of mixed and allochthonous genotypes from other Mediterranean areas, but the assignment to one of the four clusters was sufficiently robust as proved by the high ancestry coefficient values displayed by most of the individuals (>84%). The population structure of *A. minutum* on this scale can be explained by microalgal dispersion following the main regional circulation patterns over successive generations. We hypothesize that limited connectivity among the *A. minutum* populations results in low gene flow but not in the erosion of variability within the population, as indicated by the high gene diversity values. This study represents a first and new integrated approach, combining both genetic and numerical methods, to characterize and interpret the population structure of a toxic microalgal species. This approach of characterizing genetic population structure and connectivity at a regional scale holds promise for the control and management of the harmful algal bloom events in the Mediterranean Sea.

Keywords: connectivity; population structure; HAB (harmful algal blooms); microsatellite; genetic distance; dinoflagellate

1. INTRODUCTION

Management efforts against eutrophic events and high biomass proliferation of toxic microbial species in marine environments can benefit from information about how environmental factors influence genetic structure and population connectivity. Marine planktonic microbial species are generally thought to be dispersed throughout the ocean worldwide. Their ubiquity is attributed to huge population sizes [1,2] and their vast capacity for dispersal: they are passively transported and distributed homogeneously by different system currents, and there is an apparent absence of dispersal barriers [3]. This extensive dispersal capability has led to the assumption that microplanktonic populations are highly connected and they should therefore exhibit extensive gene flow preventing genetic isolation, fixation of genetic differences among populations and hence allopatric speciation [4,5]. As a consequence, these microbial species have been viewed as consisting of less genetically structured populations with a lower level of speciation than terrestrial animals and plants. However, molecular data have recently revealed a high degree of genetic diversity within microbial species with evident cryptic biodiversity that is well structured in the open ocean and benthic environment [6–9]. Together with other permanent and temporary components of the planktonic community, such as meroplankton, larval stages of animals and propagules of seaweeds [10–14], micro-organisms show long-distance dispersal modulated by geographical and physical or ecological barriers [15,16]. The processes of gene flow within microbial species can result in different populations with different genetic structures that reflect the degree of connectivity of a geographical location with adjacent areas [17–19]. Gene flow may also be influenced by the duration of the vegetative planktonic form during the migration across the open ocean, as well as
by local adaptation mediated by resting stages [20,21]. These circumstances are responsible for the successful dispersal and colonization of new areas by microbial micro-organisms [22].

The genetic seascape provides a scenario of connectivity and dispersal of both planktonic and benthic organisms through the coupling of sets of population genetic data and environmental descriptors within a spatial genetic structure [23,24]. Traditionally, estimates of dispersal have notoriously been difficult to obtain in marine environments owing to the high dispersal potential of many organisms and a variety of processes involved at different scales. Increasing efforts are underway in order to link population genetic structure with large-scale circulation patterns [25]. Most of these studies rely on isolation-by-distance analysis, but they very often show no linear relationship [26,27]. Alternatively, spatially explicit physical models, including dispersal kernels of different complexity, have been implemented in order to explain the genetic distances among populations within a species [28]. The combination of numerically modelling dispersion with population genetic structure analysis has notably improved the scenario of several marine species population dynamics and can also open new frontiers on the comprehension of the mechanisms for HAB species’ dispersion and expansion.

This is the first reported study to explore population structure using microsatellite loci combined with a connectivity model of the toxic dinoflagellate Alexandrium minutum populations in the Mediterranean Sea. The study represents a novel advancement in the evaluation of unknown population genetic structure of a bloom-forming phytoplankton species by incorporating a model of indirect connectivity within the Mediterranean basin. Using a numerical simulation model of dispersion throughout different generations in the Mediterranean Sea, we constructed a hypothetical scenario of natural connectivity of this toxic microbial species that may explain some aspects of the observed genetic structure of A. minutum populations. Alexandrium minutum is responsible for harmful phenomena which negatively alter marine ecosystem functions and services with outbreaks of paralytic shellfish poisoning (PSP) [29,30]. In the Mediterranean Sea, the increase and expansion of A. minutum outbreaks have coincided with exploitation of the coastline that increasingly offers confined nutrient-enriched waters suitable for microalgal proliferation [31,32]. Despite the dinoflagellates’ preference for settling in confined environments near shore, A. minutum has an enormous natural potential of dispersal because of its capacity to grow and produce resting cysts under a wide range of environmental conditions [33]. Our approach allowed characterizing the protist A. minutum population structure and connectivity at spatial scales for the control and management of the HAB events in the Mediterranean Sea.

2. MATERIAL AND METHODS
(a) Origin and clonal cultures
A total of 116 clonal isolates of A. minutum were collected from different coastal areas of the Mediterranean Sea and the eastern Atlantic (electronic supplementary material, table S1). The Spanish Atlantic locality was sampled as it represented an external site to be compared with the Mediterranean sampling sites. The isolates were obtained from environmental samples which included both surface sea water and sediment, as described in Penna et al. [34].

(b) DNA extraction and microsatellite genotyping
Exponential growth phase of each isolate was collected by centrifugation for DNA extraction. Genomic DNA was purified using DNeasy Plant Mini Kit (Qiagen). The ribosomal gene 5.8S rDNA and internal transcribed spacer (ITS) regions of each microalgal isolate genomic DNA were amplified according to the protocol of Penna et al. [34] and the polymerase chain reaction (PCR)-amplified products were sequenced (electronic supplementary material, table S2).

After testing 12 microsatellite loci [35], seven loci were selected for genotyping the 116 isolates as the other five loci could not be confidently amplified in all the samples. PCR amplifications were carried out in a DNA Thermocycler 2720 (Applied Biosystems). PCR mixture contained 5 ng of template DNA, 200 μM of each dNTP, 0.5 μM of each primer, 2 mM MgCl2, 1 x reaction buffer and 0.6 U AmpliTaq Gold (Applied Biosystems) in a final volume of 25 μL. A concentration of 3 mM MgCl2 was used for the Aminu22 locus. For the Aminu11, Aminu29 and Aminu48 loci, only 0.8 μg μL⁻¹ of BSA was added to the reaction mixture. The PCR conditions for Aminu22, Aminu41, Aminu43 and Aminu44 loci were as follows: an initial denaturation step of 10 min at 94°C, 38 cycles of 30 s at 94°C, 30 s at 50°C, 1 min at 72°C and a final extension step of 5 min at 72°C. The PCR conditions for Aminu11, Aminu29 and Aminu48 loci were as follows: an initial denaturation step of 10 min at 94°C, five cycles of 30 s at 94°C, 30 s at 58°C, 1 min at 72°C; five cycles of 30 s at 94°C, 30 s at 55°C, 1 min at 72°C; 28 cycles of 30 s at 94°C, 30 s at 50°C, 1 min at 72°C and a final extension step of 5 min at 72°C. Sizing of microsatellite loci was carried out by GENEMAPPER (Applied Biosystems).

(e) Population genetic analysis
The data were analysed in the following steps.

— Genetic diversity. The linkage disequilibrium (LD) test was performed with FSTAT v. 2.9.3.2 (http://www2.unil.ch/cppgen/softwares/fstat.htm). The number of alleles per locus was calculated using ARLEQUIN v. 3.11 (http://cmpg.unibe.ch/software/arlequin3/), and gene diversity and allelic richness using FSTAT.

— Population differentiation. Global and pairwise FST fixation indices and an analysis of molecular variance (AMOVA) were calculated using ARLEQUIN v. 3.1. The significance of the FST values was assessed after 1000 permutations. A locus-by-locus AMOVA with FST calculation was carried out using ARLEQUIN v. 3.1. Unrooted dendrogram of populations was constructed using the neighbour-joining (NJ) method, as implemented in the PHYLIP v. 3.6 (http://evolution.gs.washington.edu/phylip.html) using chord distance (Dc) [36] calculated with MICROSAT (http://hpgl.stanford.edu/projects/microsat/) v. 1.4d as genetic distance. The robustness of the tree was tested by 10 000 bootstrap pseudoreplicates. An isolation-by-distance analysis was carried out to correlate genetic distances, expressed as FST/(1 − FST), with geographical distances through the Isolation by Distance v. 3.16 (http://www.bio.sdsu.edu/pub/andy/IBD.html). Geographical distances were measured as the linear distances between

pairs of locations. The significance of each test was examined by 10 000 permutations.

— Inference of population structure. A principal coordinates analysis (PCoA), based on the matrix of genetic distances among alleles, was performed using GENALEX v. 6.1. Population structure was investigated with a Bayesian clustering method implemented in STRUCTURE v. 2.2 [37]. The program implements a model-based clustering method to infer population structure and assign individuals to populations using multilocus genotype data (microsatellite data). Using the estimated allele frequencies for each individual also based on a modest number of loci, it is possible to compute the likelihood of origin of a given genotype from a population. After inferring the most likely number of different clusters, this method estimates the ancestry coefficient of each individual, which represents the likelihood of it belonging to one of the identified clusters. In this study, the model for the assignment of ancestry of individuals was the admixture model, where each individual draws some fraction of his genome from each of the \( K \) populations. The most likely number of \( K \) cluster was identified through three independent runs. Each run was based on 1 000 000 MCMC simulations with an initial ‘burn-in’ period of 250 000; 10 iterations were run for each \( K \) value ranging from 1 to 10. The second-order statistics \( \Delta K \) was calculated to better refine estimation of \( K \). The STRUCTURE analysis was also run hierarchically with the following scheme: a first run (same settings as above) was performed setting \( K = 2 \). Each of the two identified groups was then subjected to a canonical analysis in order to find the most likely number of inferred group. Finally, the estimated cluster membership coefficient \((Q)\) matrices of multiple runs generated by STRUCTURE were analysed using CLUster Matching and Permutation Program (CLUMPP v. 1.1.2; [39]). The program evaluates the similarity of outcomes between populations structure estimates [40] identifying the run that produces the highest average pairwise similarity \((H)\). The input of CLUMPP is the output files generated by STRUCTURE for each of the \( K \)-values tested. After running the program, determination of the best configuration (in terms of \( K \)) is accomplished by taking the highest \( H \)-value among those estimated for different \( K \).

(d) Population connectivity estimation

The dispersal of \( A. \) minutum in the Mediterranean Sea was investigated with a coupled biophysical model and an individual-based model (IBM). The ocean circulation model, described by Jordi and Wang [41], reproduced the individual-based model (IBM). The ocean circulation investigated with a coupled biophysical algorithm, which— Inference of population structure. A principal coordinates analysis (PCoA), based on the matrix of genetic distances among alleles, was performed using GENALEX v. 6.1. Population structure was investigated with a Bayesian clustering method implemented in STRUCTURE v. 2.2 [37]. The program implements a model-based clustering method to infer population structure and assign individuals to populations using multilocus genotype data (microsatellite data). Using the estimated allele frequencies for each individual also based on a modest number of loci, it is possible to compute the likelihood of origin of a given genotype from a population. After inferring the most likely number of different clusters, this method estimates the ancestry coefficient of each individual, which represents the likelihood of it belonging to one of the identified clusters. In this study, the model for the assignment of ancestry of individuals was the admixture model, where each individual draws some fraction of his genome from each of the \( K \) populations. The most likely number of \( K \) cluster was identified through three independent runs. Each run was based on 1 000 000 MCMC simulations with an initial ‘burn-in’ period of 250 000; 10 iterations were run for each \( K \) value ranging from 1 to 10. The second-order statistics \( \Delta K \) was calculated to better refine estimation of \( K \). The STRUCTURE analysis was also run hierarchically with the following scheme: a first run (same settings as above) was performed setting \( K = 2 \). Each of the two identified groups was then subjected to a canonical analysis in order to find the most likely number of inferred group. Finally, the estimated cluster membership coefficient \((Q)\) matrices of multiple runs generated by STRUCTURE were analysed using CLUster Matching and Permutation Program (CLUMPP v. 1.1.2; [39]). The program evaluates the similarity of outcomes between populations structure estimates [40] identifying the run that produces the highest average pairwise similarity \((H)\). The input of CLUMPP is the output files generated by STRUCTURE for each of the \( K \)-values tested. After running the program, determination of the best configuration (in terms of \( K \)) is accomplished by taking the highest \( H \)-value among those estimated for different \( K \).

The Adriatic emerged as the group with the highest genetic diversity, having the highest values of both gene diversity and allelic richness. The AMOVA based on allele particles were released at each model grid point adjacent to the coast every 10 days over the last 8 years of model currents. Each release was carried forward for 30 days (considered to represent the vegetative cell phase) without inclusion of cell mortality. The probability that a particle is transported from one coastal grid point to arrive at another in 30 days is the direct connectivity \((C_d)\) calculated as the average percentage of particles released from one model grid point that arrived to another grid point adjacent to the coast at the end of the 30 days. Indirect connectivity \((C_i)\) is the probability of a vegetative cell being transported from one coastal grid point to another along indirect routes of migration by means of island stepping stones and coastal-zone diffusion and provides information on dispersal over several life cycles (or cell phases). According to graph theory, \( C_i \) is defined as:

\[
C_i = \sum_{n=1}^{N} C^n_d
\]

where \( N \) is the number of life cycles, \( C_d \) identifies the direct routes and \( C^2_d \) includes all the routes requiring one step (or two direct routes), etc. [42]. \( C_i \) reached an approximate steady state within 200 life cycles. Therefore, the life cycle of \( A. \) minutum limits the geographical range over which vegetative cells can advect or mix, even taking possible indirect routes of migration into account. Despite this limit, \( C_i \) was calculated over 500 life cycles to ensure that the steady state was reached.

3. RESULTS

Most of the \( A. \) minutum isolates were sequenced for species-specific identification. The length of the 5.8S rDNA gene and ITS1–ITS2 regions of \( A. \) minutum isolates was 520 bp. Alignment analyses of sequences of different dinoflagellate species in GenBank showed complete identity (100%) with the Mediterranean \( A. \) minutum confirming the species assignment of our isolates. Of the 12 polymorphic microsatellites tested for the amplification of the \( A. \) minutum isolates, seven microsatellites gave robust and repeatable amplification products, namely Aminu10, Aminu11, Aminu22, Aminu29, Aminu41, Aminu43, Aminu44 and Aminu48. The remaining five microsatellites were discarded, as PCR products were obtained in less than 50 per cent of the \( A. \) minutum isolates. As expected, all seven loci showed a single allele in each isolate owing to the haploid phase of the vegetative cells of this species. Statistical analyses of the seven microsatellites gave no signs of linkage disequilibrium (data not shown). The seven loci were polymorphic in all sampled localities except for the Aminu48 locus in the Catalan population, where only one allele was found. An NJ tree was constructed based on the assumption that each \( A. \) minutum sampling site corresponded to a distinct genetic group. The NJ tree identified six distinct groups corresponding to the different regional seas: north-western Adriatic, Ionian, Tyrrhenian, Catalan, Balearic and Atlantic (figure 1). The topology of the NJ tree was strongly supported by high bootstrap values.

The main indices of genetic diversity for each group are shown in the electronic supplementary material, table S3. The Adriatic emerged as the group with the highest genetic diversity, having the highest values of both gene diversity and allelic richness. The AMOVA based on allele
Bayesian clustering method (implemented in *STRUCTURE*) was used on single genotypes to infer the most likely number of homogeneous genetic groups, *K*. Comparison of the ln-likelihood of the different runs for the different *K* tested, revealed no peak, although a slight but constant increase in ln-likelihood up to the highest number of *K* tested (detailed analyses of *STRUCTURE* were also shown in a paragraph of the electronic supplementary material). At the same time, the second-order statistic devised to infer the most likely number of *K*, namely Δ*K*, did not yield unambiguous results either, there being a major peak at *K* = 2 and another peak at *K* = 4. To distinguish between *K* = 2 and *K* = 4, we estimated the highest average pairwise similarity index (H-value) according to the method described in Nordborg *et al.* [40] and implemented in the software CLUMPP. We inferred *H* for *K* = 2, *K* = 3 and *K* = 4: the results clearly showed that the highest *H*-value is that associated with *K* = 4 (electronic supplementary material, table S6). Furthermore, adopting a hierarchical *STRUCTURE* analysis, we recovered the same four distinct genetically homogeneous groups identified by the standard *STRUCTURE* analysis and by CLUMPP. Not only did these four clusters correspond to those identified by PCoA, but with *K* = 4 we recorded the highest ancestry coefficient for each genotype. The overall genetic structure of the seven sampling sites can be therefore assigned to four distinct homogeneous clusters or populations (table 1). Taking *K* = 4, we found that the two sampled Adriatic sites could be assigned to cluster 1, the Ionian site to cluster 2, the Tyrrhenian site to cluster 3 and the Catalan site to cluster 4. The Balearic and Atlantic sampled sites could not be assigned to any of the four clusters as they consisted either of individuals with mixed genotypic components or of individuals with components from two or more clusters. It was also evident from the individual ancestry coefficient that each population contained allochthonous individuals as well as individuals with mixed genotypes. At any rate, it is also relevant that a very high percentage of individuals (84.4%) showed ancestry coefficients higher than 75 per cent highly indicative of assignment to one of the four inferred clusters (figure 3).

Hydrodynamic model simulations indicated that direct cell exchange (*G*<sub>d</sub>) on a regional scale, even between the geographically closest locations, is highly unlikely on the
origin do not appear to be geographically structured but rather dispersed in different clusters [43,44]. This lack of structure is not supported by our data that demonstrate the existence of regional scale genetic population structure of the toxin-producing phytoplankton species *A. minutum* in the Mediterranean Sea. Moreover, we suggest that the genetic population differentiation is related to the basin scale transport patterns through successive generations of the vegetative microalgal cells. By combining indirect connectivity and genetic models, we were able to add new insights into microbial species connectivity patterns relevant to oceanic scenarios and useful for understanding the distribution, dispersal and evolution for the so defined cosmopolitan marine micro-organism. The diversity detected with microsatellite markers significantly exceeds that of ribosomal DNA markers revealing enough allelic variation to investigate the genetic structure of *A. minutum* in the Mediterranean Sea. The *A. minutum* isolates were collected from seven different geographical areas. We started from the strong assumption that each sampling area corresponded to a distinct, genetically homogeneous *A. minutum* population. The NJ dendrogram confirmed that six of the seven sampling areas were well differentiated, while the two Adriatic sites merged into a single cluster. A PCoA of individual genotypes, with no *a priori* assumptions regarding their origin, identified four major groups of isolates. The geographical origin of most of the isolates in each group showed that these four groups correspond to four different regional seas: Adriatic, Ionian, Tyrrhenian and Catalan. Allochthonous individuals were present within each group. The STRUCTURE and subsequent CLUMPP analyses highlighted the best configuration was that with the setting *K* = 4. Most of the isolates in the Mediterranean Sea were assigned to one of the four inferred clusters with a very high probability. Each of the four clusters included a small fraction of mixed and allochthonous genotypes from other Mediterranean areas, but the assignment to one of the four clusters was sufficiently robust, as proved by the high ancestry coefficient values which most of the individuals (>84%) showed. The strong population structure was also confirmed by the global *F*<sub>ST</sub> value, while the pairwise *F*<sub>ST</sub> values among the four Mediterranean clusters were also highly significant. Within each cluster, high gene diversity
values were encountered. This unexpected within-population variability that was not related to any gene flow among sampling areas can be explained by the biological and genetic role of the seed bank in marine habitats. It has been proved that resting cysts settled in the bottom sediment can remain viable for many decades and accumulate over several years constituting a reservoir of potential genetic diversity and resulting in high values of gene diversity. It is therefore expected that *A. minutum* local cyst banks in each examined area in the Mediterranean Sea also potentially host high gene diversity. The role of resting cysts in promoting genetic diversity has been demonstrated in other geographical areas, such as the northern European Gullmar Fjord, where the diatom population of *Skeletonema marinoi* is genetically differentiated from the open sea population. Furthermore, the genotypes in this fjord represent a continuous gradient of genetic differentiation that permits the recruitment of different clones with different live-permanence [45]. Plankton and benthic environments are genetically linked with the marine ecosystem, as benthic stages of organisms are continuously resuspended from a large gene pool in the bottom to the surface where the propagules germinate and persist in the vegetative form. Populations or clones are probably selected by environmental conditions to grow and dominate the water column [46]. Like other phytoplankton species, the dinoflagellate *A. minutum* alternates between planktonic (vegetative cells) and benthic (resting cysts) stages as a response to endogenous and/or environmental factors [47,48].

Calculation of oceanographic distance in relation to the population genetic structure based on *F*<sub>ST</sub> was applied to tentatively explain the genetic pattern of *A. minutum* in the Mediterranean basin. The population structure of *A. minutum* was not correlated to Euclidean distance as expected under an isolation-by-distance model; in other terms, the probabilities of *A. minutum* gene flow among sites have little to do with the physical distance between them [49,50]. The decoupling between the genetic structure and Euclidean distances between the sampled sites is probably due to the complex of geography and hydrodynamics of the Mediterranean regions as already depicted for other organisms. Marine biotic communities exhibit patterns of genetic structure that are more in accordance with asymmetric dispersal processes than to geographical distances [51,52]. More recent genetic studies have demonstrated an asymmetric gene flow under the Antarctic Circumpolar Current (ACC) in fish population [53], as well as an asymmetrical dispersal of bull-kelp in southern New Zealand [54] or multiple events of introgression of allochthonous genotypes and rapid dispersal into the Mediterranean Sea that explain the population genetic variability of the seaweed *Asparagopsis taxiformis* [55].

Straight geographical distances are not good indicators for inferring dispersal patterns in oceanic systems, particularly in the Mediterranean Sea, which is characterized by closed circulation patterns produced by coastal topography. In our study it seems, therefore, that the four major distinct genetic clusters reflect a situation of rather well-established geographical isolation with very little, if any, gene flow between them. Although the situation regarding the Balearic Sea and the eastern Atlantic deserves further investigation, mainly by additional sampling, some preliminary hypotheses can be advanced with respect to the four major clusters identified.

We considered the indirect connectivity model based on the Mediterranean basin circulation at different
regime scales. Further, as this study demonstrates, direct connectivity at a regional level is highly unlikely. This may be attributed essentially to the relatively short vegetative stage that precludes dispersal over long distances. *Alexandrium minutum* vegetative cells are known to spend probably less than one month in the water column, although they are present throughout the year [56]. Moreover, long-distance propagation of *A. minutum* in open ocean conditions is unlikely because of the severe nutrient limitations [57]. Instead, a scenario of the indirect connectivity provides a more likely explanation of the main features of genetic differentiation among the four populations of *A. minutum* examined. The $C_i$ allowed two areas with relatively high internal connections to be identified: the western Adriatic and Mediterranean Sea. This is consistent with the previously described ecoregions [58] and with the general concept that the Mediterranean Sea consists of two relatively independent basins communicating through the Strait of Sicily, a situation which constrains basin-scale circulation and has potential consequences for the distribution of planktonic organisms [59]. Within each region, the $C_i$ was relatively high. In particular, there was considerable internal connectivity in the Adriatic Sea ($C_i = 0.92$) attributed to its closed circulation pattern. Stable cyclonic circulation in the Adriatic Sea not only favours alongshelf transportation, but also acts as a frontal structure reducing offshore transportation of coastal phytoplankton [60]. Furthermore, the phytoplanktonic cell survival in the northern Adriatic Sea can be sustained by outflow from the river Po, which markedly influences the hydrology of the western side of the northern Adriatic Basin [61,62]. This $C_i$ scenario in the Adriatic Sea is consistent with the genetic structure of the two sampling sites, Trieste and Marotta identified by the NJ and PCoA configurations, and STRUCTURE assignment to belong to the same cluster. In the western Mediterranean Sea, connectivity estimates revealed the close link between Palma and Arenys de Mar ($C_i = 0.50$), which are connected by the anti-clockwise-circulating northern current [63]. Less robust were the connections between Olbia, which appeared genetically well-differentiated. The presence of a semi-permanent sub-basin gyre in the northern Tyrrhenian Sea [64] may play a key role in promoting genetic differentiation between the Arenys de Mar and Olbia sites. Finally, the Ionian site, Siracusa, had the highest $F_{ST}$ pairwise values when compared with the other Mediterranean clusters, making this site highly differentiated and isolated, and apparently uninfluenced by any gene flow. In fact, there was little connectivity between Siracusa and the other areas probably owing to constraints on exchanges exerted by the narrow Strait of Sicily and more crucially by the Strait of Otranto. Although it is possible that *A. minutum* may eventually enter the Adriatic Sea from Siracusa, dispersal in this direction is low as the main flow in the area is towards the eastern Mediterranean and it only enters the Adriatic after cyclonically circulating through the entire eastern Mediterranean Sea [65].

This connectivity model allows us to advance some hypotheses regarding the presence in each of the four populations identified of isolates with genotypes assignable to a different geographical cluster or with mixed genotypic components. Here, it is important to recall that mixed genotypes can arise given that sexual reproduction, being one of the dinoflagellate *A. minutum*’s life stages, permits recombination between two different genotypes of diploid cells and therefore confers greater genetic variability than that displayed by haploid organisms during the late phase of bloom events.

In the northern Adriatic area, allochthonous genotypes with a main Tyrrhenian component (16%) were found (electronic supplementary material, table S7). Furthermore, genotypes with Catalan and Ionian components were retrieved in the Olbia site. While the connectivity model shows that some kind of contact between the Tyrrhenian and Catalan Seas cannot be ruled out, contact between the Tyrrhenian and Adriatic Seas and the Tyrrhenian and Ionian Seas appears very unlikely. Moreover, the northern Adriatic and Tyrrhenian areas are clearly separated by the large- and meso-scale circulation systems and the geographical barrier of the Italian peninsula.

Furthermore, the worth of indirect connectivity as a predictor of the genetic structure within the Mediterranean basin was confirmed by a Mantel test that showed a highly significant correlation between matrices of $F_{ST}$ values among the four populations and the correspondent values of $C_i$ (indirect connectivity, $p < 0.001$; data not shown).

Anyway, we cannot dismiss that alternative ways of planktonic cell-dispersal exist, and these could include human-mediated transportation. Ballast water discharges and shellfish stock translocation have recently been proposed as a human-aided vector of planktonic microbial species propagation [66], even though the relative importance of this propagation pathway is highly debated [67]. The northern Adriatic and Tyrrhenian sites are characterized by intense commercial shipping, as well as widespread aquaculture with potential mussel translocations; therefore human-mediated transport cannot be excluded.

Application of high-resolution molecular markers, such as microsatellite loci, allowed us to identify a robust population structure within the harmful phytoplankton species *A. minutum* based on connectivity between distinct areas of the Mediterranean basin. To the best of our knowledge, this study marks the first attempt to correlate genetic differentiation of *A. minutum* isolates with a model of Mediterranean water circulation. Our model of indirect connectivity was consistent with the results of *A. minutum* population genetic structure analyses on a regional scale. The low number of sampling sites is certainly a limit for a more comprehensive description of *A. minutum* population structure in the Mediterranean Sea considering the varied coastal systems and hydrographic regimes in the different Mediterranean regions. Despite the admittedly low number of sampling sites across the Mediterranean region, we obtained a strong signal of differentiation. It seems highly likely that other coastal sites, especially those confined or semi-confined, would host well-differentiated populations and therefore, our data can provide a basis for planning the control and management of harmful algal blooms in Mediterranean waters.

Finally, our results served as validation for estimating population connectivity using an oceanographic approach of $C_i$ for simulating phytoplankton dispersal in the marine environment.
REFERENCES


8 Hughes Martiny, J. B. et al. 2006 Microbial biogeography: putting microorganisms on the map. Nat. Rev. 4, 102–112. (doi:10.1038/nrmicro1341)


Alexandrium minutum population structure  S. Casabianca et al. 137


