Allele-specific PCR shows that genetic exchange occurs among genetically diverse *Nodularia* (Cyanobacteria) filaments in the Baltic Sea

Gary L. A. Barker,¹ Barbara A. Handley,¹ Panmuk Vacharapiyasophon,¹ Jamie R. Stevens² and Paul K. Hayes¹

Author for correspondence: Paul K. Hayes. Tel: +44 117 928 7483.
e-mail: Paul.Hayes@Bristol.ac.uk

¹ School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

² Hatherly Laboratories, School of Biological Sciences, University of Exeter, Prince of Wales Road, Exeter EX4 4PS, UK

Some cyanobacteria have been shown to exchange genetic information under laboratory conditions, but it has not been clear whether such genetic exchange occurs in the natural environment. To address this, a population genetic study was carried out on the filamentous diazotrophic cyanobacterium *Nodularia* in the Baltic Sea. *Nodularia* filaments were collected from 20 widely distributed sampling stations in the Baltic Sea during June and July 1998. Allele-specific PCR (AS-PCR) was used to characterize over 2000 filaments at three loci: a non-coding spacer between adjacent copies of the main structural gas vesicle gene *gvpA* (*gvpA*-IGS), the phycocyanin intergenic spacer (PC-IGS) and the rDNA internal transcribed spacer (rDNA-ITS). The three loci were all found to be polymorphic in the 1998 population: two alternative alleles were distinguished at the *gvpA*-IGS and PC-IGS loci, and three at the rDNA-ITS locus. All 12 possible combinations of alleles were found in the filaments studied, but some were much more common than others. The index of association (I_A) for all possible pairwise combinations of isolates was found to differ significantly from zero, which implies that there is some linkage disequilibrium between loci. The I_A values for 16 out of 20 individual sampling stations also differed significantly from zero: this shows that the observed linkage disequilibrium is not due to pooling data from genetically distinct subpopulations. Monte-Carlo simulations with random subsets of the data confirmed that some combinations showed significantly more linkage disequilibrium than expected by chance alone. It is concluded that genetic exchange occurs in the natural *Nodularia* population, but the frequency is not high enough for the loci to be in linkage equilibrium. The distribution of the 12 genotypes across the Baltic Sea was found to be non-random, but did not correlate with temperature, salinity or major nutrient concentrations. A significant relationship was found between the gene diversity among filaments at each station and the distance of the station from the centre of the sampling area: possible reasons for this trend are discussed.

Keywords: AS-PCR, index of association, population genetics, gene diversity

INTRODUCTION

Many agronomically and medically important bacteria have been shown to undergo extensive genetic recom-

bination (Maynard Smith *et al.*, 1993). Although there have been numerous population genetic studies of pathogens (e.g. Spratt *et al.*, 1995; Stevens & Tibayrenc, 1996; Whittam, 1995; Trott *et al.*, 1997; Feizabadi *et al.*, 1997) and soil micro-organisms (Eardly *et al.*, 1990; Hagen & Hamrick, 1996; Silva *et al.*, 1999), there are few such studies for aquatic micro-organisms (Wise

Abbreviations: AS-PCR, allele-specific polymerase chain reaction; IGS, intergenic spacer; ITS, internal transcribed spacer; PC, phycocyanin.

et al., 1996) and none for cyanobacteria, despite the fact that many members of the group form toxic blooms which threaten water quality. Laboratory studies of cyanobacterial cultures have shown that there is the potential for gene exchange in the environment. Conjugation has been demonstrated both within (Muro-pastor *et al.*, 1994) and between genera (Sode *et al.*, 1992; Chiang *et al.*, 1992; Thiel & Wolk, 1987). Many taxa are naturally competent (Singh *et al.*, 1987; Barten & Lill, 1995), and studies have shown that potentially transforming DNA is often present in seawater (Lorenz & Wackernagel, 1994). Transduction is also a potential mode of genetic exchange, as cyanophages are known to be present in high concentrations in most aquatic environments (Wilson & Mann, 1997).

The filamentous, diazotrophic cyanobacterium *Nodularia* is a convenient organism on which to perform population genetic analysis as it is reliably present during the summer months in the Baltic Sea, and sufficient template DNA for PCR can be obtained from single filaments isolated from the water column. Whilst the extraction of DNA from filaments precludes any further analysis of them, the advantages of directly analysing filaments picked from water samples, rather than clonal cultures established from them, are twofold: first, a great deal of time and effort is saved in isolating and maintaining thousands of clonal cultures; and secondly (and more importantly), culture collections may not be representative of the natural population if genetic differences between isolates have differential effects on fitness in the artificial conditions of laboratory culture.

Genetic diversity in natural populations of *Nodularia* has been demonstrated in the Baltic Sea (Hayes & Baker, 1997; Barker *et al.*, 1999) and elsewhere (Neilan *et al.*, 1995; Bolch *et al.*, 1996, 1999), but these studies have either focussed on a single locus (Hayes & Barker, 1997; Neilan *et al.*, 1995; Bolch *et al.*, 1999) or had a sample size too small for population analysis (Barker *et al.*, 1999). In this study we have characterized three loci from 2336 *Nodularia* filaments collected during research cruises in June and July 1998. The large sample size minimizes the sampling error of population genetic parameters such as gene diversity (Nei, 1987) and index of association (I_A) (Maynard Smith *et al.*, 1993). The three loci chosen were (1) part of the *cpc* operon that includes the phycocyanin intergenic spacer (PC-IGS; Neilan *et al.*, 1995), (2) the rDNA internal transcribed spacer (rDNA-ITS; Wilmotte, 1994) and (3) the intergenic spacer between two adjacent copies of *gvpA* (the gene encoding the main structural gas vesicle protein; Tandeau de Marsac *et al.*, 1985), *gvpA*-IGS. These loci were selected for study because each is known to be polymorphic in *Nodularia* (Barker *et al.*, 1999). For the PC-IGS locus four sequence variants were found among cultured isolates. Sequence analysis of cultured isolates similarly revealed the presence of three distinct rDNA-ITS, and two *gvpA*-IGS genotypes (Barker *et al.*, 1999). The four PC-IGS sequence types fall into two main groups that can be distinguished by allele-specific PCR

(AS-PCR) (Hayes & Barker, 1997), as can the three rDNA-ITS sequence types. The AS-PCR approach uses PCR to amplify a region with broad-specificity primers and then a second, nested PCR with one broad-specificity and one or more allele-specific primers: the presence/absence and/or size of the resulting PCR products indicate which allele(s) were present in the original template DNA.

In addition to questions regarding population structure, there is currently disagreement concerning the taxonomic status of *Nodularia* in the Baltic Sea, with some studies suggesting that several species co-exist in this region (Komárek *et al.*, 1993; Bolch *et al.*, 1999), whilst other data show that many phenotypic and genotypic markers are not congruent, implying that gene flow occurs between isolates (Barker *et al.*, 1999). The taxonomic implications of our results are discussed.

METHODS

Sampling. *Nodularia* filaments were collected from the mixed layer of the water column during June and July 1998 using a 100 µm mesh plankton net as described by Barker *et al.* (1999). Sampling locations are shown in Table 1. Individual filaments were picked from ~50 µl drops of population sample, washed, and isolated with the aid of a dissecting microscope (Walsby *et al.*, 1995). To avoid sample bias (for example systematically picking long filaments in preference to shorter ones), plankton catches were diluted with filtered sea water to give <5 filaments per 50 µl and then every filament in a drop of the sample was picked for analysis. Washed filaments were transferred to 0.5 ml Eppendorf tubes containing 25 µl Supertaq PCR buffer: 20 mM Tris/HCl (pH 9.0), 3.0 mM MgCl₂, 100 mM KCl, 0.2% (w/v) Triton X-100, 0.02% (w/v) gelatin (HT Biotechnology). Tubes were overlaid with light mineral oil (Sigma), heated by immersion in freshly boiled water for 5 min to lyse the cells, and stored at -20 °C: the filament lysate was used, without further treatment, as the source of template DNA in the reactions described below.

Multilocus genotype determination using PCR. An overview of the PCR reactions performed, and the primers used at each stage, is shown in Fig. 1. Amplification of the PC-IGS and rDNA-ITS loci was performed in two stages. In the first stage, the primer pairs PCβF/PCαR and 1358F/ITSREV3 (Table 2) were used to amplify both the PC-IGS and rDNA-ITS regions in a duplex reaction. In separate second-stage reactions the forward primers PCβF and 1358F were used in combination with allele-specific reverse primers (Fig. 1, Table 2) to reveal the genotype of each filament. Amplification of the *gvpA*-IGS was carried out in a single stage, and alternative alleles were detected as product length variants. The PC-IGS/rDNA-ITS stage 1 reaction was carried out in a volume of 32 µl; all other reactions were carried out in 25 µl volumes. All amplification reactions contained Supertaq buffer to 1× final concentration, 12.5 pmol of each primer, 100 µM deoxynucleoside triphosphates (Pharmacia), and 0.5 units Supertaq DNA polymerase (HT Biotechnology). Reactions were overlaid with light mineral oil (Sigma), and subjected to the thermocycling parameters shown in Table 3 using an Omnigene thermocycler (Hybaid). Negative control tubes, containing all reaction components except *Nodularia* filaments, and positive controls, containing single filaments of *Nodularia* isolate BC-Nod 9402, 9408 or 9427 (Barker *et al.*, 1999), were set up for each batch of test filaments, and subjected to the two-stage

Table 1. Sampling location and genotype data

Sampling station	Date*	Latitude	Longitude	Genotype (PC-IGS, rDNA-ITS, <i>gvpA</i> -IGS)												$I_A \pm \text{error}^\dagger$
				1 I/II-I-I	2 I/II-I-II	3 I/II-II-I	4 I/II-II-II	5 I/II-III-I	6 I/II-III-II	7 III-I-I	8 III-I-II	9 III-II-I	10 III-II-II	11 III-III-I	12 III-III-II	
<i>Litorina</i>																
L2	30-06-98	54-58° N	11-19° E	0	0	0	0	0	0	0	0	83	0	1	0	0-012 ± 0-686
L4	01-07-98	54-47° N	12-31° E	0	0	0	0	0	0	0	0	80	0	7	1	0-223 ± 0-163
L7	02-07-98	54-39° N	13-09° E	0	0	2	0	3	0	1	1	66	0	14	0	0-224 ± 0-006
L11	03-07-98	54-58° N	14-04° E	0	0	8	0	1	0	1	0	72	2	10	0	−0-047 ± 0-017
L14	04-07-98	55-13° N	15-28° E	0	0	7	0	1	1	0	0	63	1	19	1	0-118 ± 0-023
L16	05-07-98	55-36° N	15-01° E	0	0	30	0	0	0	2	0	65	1	6	1	−0-035 ± 0-024
L17	05-07-98	55-36° N	15-00° E	0	1	4	2	0	0	1	1	70	1	7	0	0-414 ± 0-010
L19	07-07-98	55-05° N	15-81° E	0	0	3	0	0	0	0	0	80	0	9	0	−0-039 ± 0-113
L20	08-07-98	55-05° N	14-62° E	1	3	28	2	0	0	1	1	57	1	2	1	0-339 ± 0-045
L24	09-07-98	55-13° N	13-28° E	0	0	3	1	0	0	0	0	69	0	4	0	0-176 ± 0-130
L26	10-07-98	54-65° N	12-02° E	1	0	14	0	0	0	0	0	93	0	9	0	−0-008 ± 0-035
<i>Alexander von Humboldt</i>																
AvH1	24-06-98	55-23° N	15-59° E	0	0	26	0	0	0	1	3	104	1	7	0	0-067 ± 0-000
AvH2	25-06-98	55-61° N	16-53° E	0	0	12	0	1	0	0	1	131	1	7	0	0-092 ± 0-060
AvH3	26-06-98	55-89° N	17-02° E	0	1	21	1	0	0	0	4	140	1	0	0	0-417 ± 0-031
AvH4	27-06-98	56-38° N	18-58° E	3	1	18	0	0	0	2	2	95	5	2	2	0-265 ± 0-017
AvH5	28-06-98	57-15° N	19-53° E	0	0	10	0	0	0	1	3	127	2	2	1	0-273 ± 0-057
AvH6	29-06-98	58-17° N	20-00° E	1	2	16	0	0	0	1	8	114	2	1	0	0-527 ± 0-008
AvH7	30-06-98	59-07° N	20-01° E	0	0	21	0	0	0	0	2	133	1	0	1	0-212 ± 0-052
AvH8	01-07-98	58-58° N	18-58° E	0	0	7	0	0	0	0	1	138	2	2	0	0-116 ± 0-130
AvH9	02-07-98	57-75° N	17-58° E	0	1	17	0	0	0	0	0	101	2	1	1	0-277 ± 0-041
			Total	6	9	247	6	6	1	11	27	1881	23	110	9	0-203 ± 0-001

* Day.month.year.

† The error for I_A assuming linkage equilibrium (see text for method of calculation).

Fig. 1. Flow diagram showing the amplification reactions used to determine the multilocus genotypes of single *Nodularia* filaments. The primers (shown in bold type) and template volumes used at each stage are indicated. The PC-IGS/rDNA-ITS stage 1 reaction was carried out in a volume of 32 µl; all other reactions were carried out in 25 µl volumes. The primer sequences are shown in Table 2; other reaction conditions are described in the text and in Table 3.

amplification procedures described above. All reaction products were resolved on agarose gels as described by Hayes & Barker (1997).

Detection of linkage disequilibrium. The index of association (Maynard Smith *et al.*, 1993) was used to determine whether the *Nodularia* population was in linkage equilibrium (panmixis) or disequilibrium (predominantly clonal growth). The index of association I_A is equal to $(V_{\text{obs}}/V_{\text{exp}}) - 1$, where V_{obs} is the variance of pairwise genetic distance values for all possible combinations of individuals, and V_{exp} is the variance expected if the population is in linkage equilibrium. Values of I_A which differ significantly from zero are indicative of some degree of linkage disequilibrium (Maynard Smith, 1995). For the Baltic Sea as a whole, a Monte-Carlo approach was used to test whether the observed I_A value was significantly different from that expected if the population was in linkage equilibrium. Five hundred new random data sets were produced with identical allele frequencies at each locus to the original data, and an I_A value calculated for each one. The standard error on all 500 I_A values was then calculated. For the calculation of I_A for individual sampling stations, the expected error was calculated as described by Maynard Smith (1995).

The association of alleles at different loci in *Nodularia* was also studied using the method of Stevens & Tibayrenc (1996), a modified Mantel test (Mantel, 1967), to detect linkage disequilibrium based on correlations between independent genetic markers (the PC-IGS, rDNA-ITS and *gvpA*-IGS loci); non-random associations can provide evidence that gene flow is restricted in the population under study, irrespective of the reasons for this. The technique is a generalized Monte-Carlo test for linkage, independent of mating system or ploidy. In brief, the loci are shared into two groups (for example PC-IGS in one group and both rDNA-ITS and *gvpA*-IGS in the other). Next, a genetic distance matrix is calculated for each of these

Table 2. PCR primers used in this study

Primer	Sequence (5'–3')	Position	Region	Organism	EMBL accession no.
PCβF	GGCTGCTTGTTTACGCGACA	668–687	<i>cpcB</i>	<i>Synechococcus</i> PCC 7942	AB008546
PCαR	CCAGTACCACCAGCAACTAA	1299–1318	<i>cpcA</i>	<i>Synechococcus</i> PCC 7942	AB008546
9402Ra	CTAAATTCTTAGGTATGGAG	15–34	PC-IGS	<i>Nodularia</i> BC-Nod 9402	AJ224916
9427R	ATCAATCAACTTTTGAGCG	144–162	<i>cpcA</i>	<i>Nodularia</i> BC-Nod 9427	AJ224915
1358F	CACACCATGGAAGCTGGTCAC	1352–1372	16s rDNA	<i>Nodularia</i> BC-Nod 9427	AJ224447
ITSREV3	CATTTACATTGGTGTTTCTGC	333–353	rDNA-ITS	<i>Nodularia</i> BC-Nod 9427	AJ224447
9408ITSR	GGTTCTGGCTGGATTATC	192–209	rDNA-ITS	<i>Nodularia</i> BC-Nod 9427	AJ224447
9427ITSR	GGTTCTGGCTGGATATTT	192–209	rDNA-ITS	<i>Nodularia</i> BC-Nod 9427	AJ224447
PVA1F2	TTGAAATATGCAGAAGCAGT	29–48	<i>gvpA1</i>	<i>Anabaena flos-aquae</i>	M32060
PVA2R2	ATCGATGACAATACCTTTGT	62–81	<i>gvpA2</i>	<i>Anabaena flos-aquae</i>	M32060

Table 3. Amplification conditions used in this study

Reaction*	Denaturation	Annealing	Extension	Cycles
PC-IGS/rDNA-ITS stage 1	94 °C, 1 min	50 °C, 1 min	72 °C, 1 min	15
PC-IGS stage 2	94 °C, 1 min	45 °C, 1 min	72 °C, 2 min	30
<i>gvpA</i> -IGS and rDNA-ITS stage 2	94 °C, 30 s	50 °C, 1 min	72 °C, 1 min	30

* All reactions were subjected to an initial denaturation step of 94 °C for 3 min, and a final extension step of 72 °C for 5 min.

groups, showing the distance between a given individual and all other members of its group. A correlation coefficient (r) is then calculated between the two groups of distances for this, the 'observed' arrangement of genotypes. The genotypes (the allele patterns for each locus for each *Nodularia* filament) are then randomly reassorted a given number of times (e.g. 100 times), each time creating a 'new' set of *Nodularia* 'specimens' of randomly mixed genetic composition. A new r -value is calculated for each random reassortment. Following the required number of genotype reassortments and r -value calculations, a frequency distribution of r -values can be constructed, allowing a probability value to be attached to the observed correlation coefficient. The process is then repeated and the loci are shared into two different groups. When the number of loci is six or less, as in this study, this is repeated until all combinations (three in this study) of loci have been evaluated. For each analysis, 100 random reassortments of each of the three combinations were used to construct each distribution of r -values. Significance levels of $P < 0.01$ and $P < 0.05$ were used for each combination. Computational constraints necessitated that linkage analyses were performed on 20 subsets of 100 randomly selected filaments, rather than the full 2336.

Other statistical analyses. Gene diversity was calculated for the population as a whole, and for individual sampling stations as described by Wise *et al.* (1996). Genetic distance was calculated as described by Nei (1972) with distance D equal to $-\log_e I$, where I , the normalized identity, is equal to $J_{xy}/(J_{xx} J_{yy})^{0.5}$. If there are three alleles at a given locus in population X with frequencies of p_1 , p_2 and p_3 , then $J_{xx} = \Sigma p_1^2 + p_2^2 + p_3^2$. If the equivalent allele frequencies in population Y are q_1 , q_2 and q_3 , then $J_{yy} = \Sigma q_1^2 + q_2^2 + q_3^2$, and $J_{xy} = \Sigma p_1 q_1 + p_2 q_2 + p_3 q_3$. The arithmetic mean of each J value over all loci is used to calculate I . All data manipulations and statistical analyses were performed using Microsoft Excel 8.0 and MINITAB, except for the modified Mantel test, which was performed using DLX.FOR (Stevens & Tibayrenc, 1995).

RESULTS

Using PCR to determine the multilocus genotype of filaments

The PCR products associated with each allele at the three loci studied are shown in Fig. 2. Filaments with PC-IGS AS-PCR products of 336 bp and 526 bp were designated types III (BC-*Nod* 9402 lane 1) and I/II (BC-*Nod* 9401 and 9427 lane 1) respectively: the AS-PCR cannot distinguish the closely related sequence types I and II (Hayes & Barker, 1997; Barker *et al.*, 1999). Filaments able to produce a PCR product with primers 9408ITSR or 9427ITSR were designated rDNA-ITS types I (BC-*Nod* 9401 lane 2) and II (BC-*Nod* 9427 lane 3) respectively, and filaments producing a product with both primers were designated type III (BC-*Nod* 9402 lanes 2 and 3). The latter class produces a PCR product with both 9408ITSR and 9427ITSR because it contains two non-identical copies of the rDNA region (see Barker *et al.*, 1999, for detailed explanation). Filaments yielding 280 bp and 380 bp *gvpA*-IGS PCR products were designated *gvpA*-IGS types I (BC-*Nod* 9401 and 9427 lane 4) and II (BC-*Nod* 9402 lane 4) respectively. All AS-PCR reactions yielded products of the expected sizes only. The full-length PCR products of the first-stage amplification of the PC-IGS and rDNA-ITS (Fig. 1) were

occasionally present along with the smaller AS-PCR products. These are not visible in Fig. 2, but, where present, were easily resolved from the AS-PCR products.

Genetic structure of the population

Five hundred filaments were collected at each of 20 sampling stations across the Baltic Sea (Table 1, Fig. 3), and multilocus genotyping was carried out on a total of 4143 single filaments, 141 to 310 filaments from each sampling location. It was not possible to determine the genotype at all three loci for every filament: 56% (2336) gave results at all three loci and a further 38% (1562) at only one or two loci, with only 6% (245) failing to produce any amplification products (Table 4). The three loci studied yielded results with varying degrees of success: the *gvpA*-IGS PCR primers gave products for 87% (3601) of filaments analysed, the rDNA-ITS primers for 75% (3122) and the PC-IGS primers for 65% (2690) (Table 4). At each locus a single allele was found to be dominant within the total population: at the *gvpA*-IGS locus allele I was carried by 97% of the 3601 filaments that produced a PCR product, for the rDNA-ITS locus it was allele II that dominated (92% of the 3122 successful amplifications) and at the PC-IGS the population was dominated by allele III (87% of the 2690 successful amplifications). There are small, but significant differences in the PC-IGS and rDNA-ITS allelic ratios in those filaments that amplified at all three loci and those that amplified at less than three loci (Table 5) (PC-IGS $\chi^2 = 30.52$, d.f. = 1, $P < 0.001$; rDNA-ITS $\chi^2 = 7.01$, d.f. = 2, $P < 0.05$; *gvpA*-IGS $\chi^2 = 0.01$, d.f. = 1, $0.95 > P > 0.90$). The origin of these differences is uncertain. It is possible that limiting amounts of template DNA were liberated from those filaments where one or more amplifications failed and that under these conditions small differences in primer annealing characteristics resulted in the underdetection of the most abundant allelic variants at the PC-IGS and rDNA-ITS loci. For the following population analysis we only consider the 2336 filaments that gave results at all three loci, where there is a reduced likelihood of any bias in the data.

Spatial distribution

The distribution of the most common genotype, type 9 (Table 1), was examined to see if it was evenly distributed across all sampling stations. Chi-squared analysis was used to evaluate the null hypothesis that the relative abundance of this genotype was the same at each station. The null hypothesis of equal relative abundance was rejected ($\chi^2 = 126$, d.f. = 19, $P < 0.005$); therefore we conclude that the distribution of this genotype is not random. Neither parametric (Pearson) nor non-parametric (Spearman rank) tests revealed any correlation between either the individual allele frequencies, or the proportion of genotype 9 at each sampling station, and either temperature, salinity, depth or nutrients including nitrate, nitrite, ammonia, phosphate and silica. The remaining 11 genotypes were too scarce for statistical analyses of their spatial distribution to be performed.

Fig. 2. Agarose gel showing AS-PCR products for cultured *Nodularia* isolates BC-Nod 9401, BC-Nod 9402 and BC-Nod 9427 (Barker *et al.*, 1999). The four lanes for each isolate are (1) PC-IGS, (2) rDNA-ITS using primer 9408R, (3) rDNA-ITS using primer 9427R and (4) *gvpA*-IGS. Size marker lanes (M) are loaded with 100 bp ladder: the bright band in the marker lanes has a length of 500 bp.

Fig. 3. Sampling stations used in this study. *Alexander von Humboldt* and *Litorina* sampling stations are shown as open and filled boxes respectively. Station 1, where *Nodularia* was not found is shown as a filled circle. This figure was redrawn from an area map with marked sampling stations created at <http://pubweb.parc.xerox.com/>

The hypothesis that the genetic distance between sampling stations might correlate with the map distance was tested by calculating Nei's normalized identity I and thus the genetic distance D for each pairwise combination of sampling stations. Using linear regression, we found no significant relationship between these

variables ($R^2 = 1.2\%$), but although there is no significant trend, the scatter plot (Fig. 4) shows that at the greatest distances compared, the genetic distance between stations is always low. In order to test the hypothesis that this effect was caused by low genetic diversity at the extreme Northern and Southern limits of

Table 4. Results from the 4143 PCR reactions carried out in this study

+, positive reaction; –, negative. Note that rDNA-ITS types I, II and III are amplifiable with 9408ITSR, 9427ITSR, or both primers, respectively.

PC-IGS	rDNA 9408ITSR	rDNA 9427ITSR	<i>gvpA</i> -IGS	<i>n</i>
+	+	+	+	126
+	+	–	+	53
+	–	+	+	2157
+	–	–	+	122
–	+	+	+	44
–	+	–	+	11
–	–	+	+	472
–	–	–	+	616
+	+	+	–	11
+	+	–	–	4
+	–	+	–	179
+	–	–	–	38
–	+	+	–	5
–	+	–	–	9
–	–	+	–	51
–	–	–	–	245
4143				

Table 5. Allelic ratios

The number (*n*) and proportion (*p*) of filaments in each allelic class derived from (A) filaments that gave amplification products at all three loci, and (B) filaments that gave an amplification product at only one or two loci are shown.

	A		B	
	<i>n</i>	<i>p</i>	<i>n</i>	<i>p</i>
PC-IGS I/II	275	0.12	80	0.23
PC-IGS III	2061	0.88	274	0.77
PC-IGS total	2336	1.00	354	1.00
rDNA-ITS I	53	0.02	24	0.03
rDNA-ITS II	2157	0.92	702	0.89
rDNA-ITS III	126	0.05	60	0.08
rDNA-ITS total	2336	1.00	786	1.00
<i>gvpA</i> -IGS I	2261	0.97	1223	0.97
<i>gvpA</i> -IGS II	75	0.03	42	0.03
<i>gvpA</i> -IGS total	2336	1.00	1265	1.00

the sampling area, the gene diversity, averaged over all three loci, was calculated for each of the 20 stations, and compared to the distance between each station and the exact centre of the area sampled. The results (Fig. 5) show that there is a significant relationship between gene diversity and the logarithm of the distance from the centre of the sampling locations ($P = 0.037$, $R^2 =$

Fig. 4. Scatter plot and regression line showing the relationship between the genetic and physical distance between all pairwise combinations of sampling stations. **Fig. 5.** The relationship between the gene diversity at each sampling station, and its distance from the exact centre of all sampling stations.

22.1%), following the equation gene diversity = $0.253 - (0.587 \log \text{distance})$.

Linkage analysis

Taken as a whole, the population was found to be dominated by filaments of genotype 9 (Table 1), which accounted for 81 % of the population and at least 59 % at individual stations. All of the 12 possible PC-IGS/rDNA-ITS/*gvpA*-IGS allele combinations were found to be present in the population, although several were extremely rare (Table 1). The index of association (I_A) calculated from all possible pairwise combinations of individuals was found to be 0.203. The standard error for I_A under the assumption of no linkage disequilibrium was calculated to be 0.0008; therefore we are confident that the observed I_A is significantly different from zero, which implies that there is some degree of linkage disequilibrium among the loci studied. The I_A values for the individual population samples were not significantly different from zero at 4 out of the 20 sample stations, but at 3 of these 4 stations (L2, L19 and AH8) gene diversity was exceptionally low (in the bottom four of the 20 stations studied). As allele frequencies tend towards fixation, the ability of I_A to discriminate

Table 6. Summary of randomized reassortment of *Nodularia* genotypes

Subset	No. of combinations with significantly more linkage disequilibrium than the simulated randomly mixed data		No. of combinations with no more linkage disequilibrium than randomly mixed data
	$P < 0.01$	$P < 0.05$	
1	1	3	0
2	2	3	0
3	1	1	2
4	0	2	1
5	0	0	3
6	2	2	1
7	2	3	0
8	1	2	1
9	3	3	0
10	1	1	2
11	3	3	0
12	0	0	3
13	2	3	0
14	2	3	0
15	1	1	2
16	1	2	1
17	2	3	0
18	3	3	0
19	2	2	1
20	0	0	3

clonality and panmixis is much diminished, and thus these results must be interpreted with caution. Most importantly, the fact that the majority of I_A values from individual stations are non-zero shows that the linkage disequilibrium detected in the whole dataset is not the result of pooling genetically discrete subpopulations, each of which is in linkage equilibrium.

Modified Mantel tests for linkage disequilibrium within 20 randomly selected subsets of 100 individuals of *Nodularia* were performed. Analysis showed significant linkage disequilibrium between some loci combinations in 17 out of 20 subsets of data (Table 6). Having detected linkage disequilibrium in the data, we compared the observed number of filaments with each possible pair-wise combination of alleles at any two loci to the number expected for a population in linkage equilibrium. The combination *gvpA*-IGS II + rDNA-ITS II was found to occur less often (29 individuals) than expected by chance (69 individuals): this difference is highly significant ($\chi^2 = 25$, $P = < 0.01$).

Gene diversity, a convenient measure of the probability that any two filaments picked at random from the population will have different alleles at a given locus, was calculated for the population as a whole. Values

were found to range from 0.06 for *gvpA*-IGS to 0.21 for PC-IGS. The total gene diversity value summed for all three loci was 0.41, giving a mean value across all three loci of 0.14.

DISCUSSION

Studies of bacteria, including fluorescent pseudomonads (Haubold & Rainey, 1996) and members of the *Enterobacteriaceae* (Gordon & FitzGibbon, 1999), have demonstrated a link between population genetic structure and properties of the local environment or host respectively. In contrast, the spatial distribution of *Nodularia* genotypes could not be explained by any of the environmental variables examined in this study, including concentrations of key nutrients, salinity, temperature and depth of the water column. The lack of correlation between the physical/chemical conditions at each sampling location and the relative abundance of particular genotypes may be explained in two ways. First, there may not be sufficient variation in conditions across the study region to differentially select between *Nodularia* genotypes; and second, gene exchange may well be common enough to randomize any association between the non-coding loci studied, and those actually under selection in the environment. The difference between our findings and those of Haubold & Rainey (1996) and Gordon & FitzGibbon (1999) may reflect the fact that these earlier studies were based on samples collected from discrete habitats, individual leaves (Haubold & Rainey, 1996) or mammalian hosts (Gordon & FitzGibbon, 1999), whereas we collected samples from a large but contiguous area of the Baltic Sea.

The finding that the most distantly spaced sampling stations were genetically similar (Fig. 4) was initially surprising. A simple explanation for this observation seems to be that for the greatest distances, one is comparing stations at the extreme southwest and northeast of the sampling area. The northern limit of the population was not explored in this study, but our records show that *Nodularia* was not found at *Litorina* station 1 (54°53' N, 10°51' E; Fig. 3), just 44 km west of station 2 (Table 1). This indicates that station 2, the station with the lowest gene diversity, may have been close to the southwestern limit of the *Nodularia* population at the time of sampling. It is possible that the extremes of *Nodularia*'s range are tolerable only to a subset of genotypes, resulting in lower genetic diversity in these areas. The significant negative correlation between the level of gene diversity at each station and its distance from the centre of the study (Fig. 5) provides some limited support for this hypothesis.

The most significant finding of this study is that genetic exchange almost certainly occurs between *Nodularia* filaments in the natural population. This is revealed by the fact that all 12 possible allele combinations (genotypes) were found to exist in our sample of the population. An alternative explanation for the occurrence of the 12 observed genotypes is that parallel

mutation along each of several purely asexual lineages has by chance generated the same alternative alleles (but no others) many times. Suppose, for example, that the ancestor of all *Nodularia* in the Baltic Sea today first diverged into two clonal lineages, each of which accumulated its own unique mutations in the PC-IGS region. We must then assume that by chance, the same suite of point mutations (and no others) occurred in the rDNA-ITS of these two independent lineages, and that finally a parallel insertion/deletion event in each of the six resulting lineages gave rise to the two *gvpA*-IGS alleles, and thus to the 12 genotypes (see Barker *et al.*, 1999, for an exact description of the differences between the alternative alleles, but note that including one unpublished PC-IGS sequence, we have at least three sequences for each allele type at each locus used in this study). Given the staggering odds against the sequence of events described above actually occurring, genetic exchange clearly provides the most parsimonious explanation of the data. We have also considered the possibility that either gene conversion or gene loss could produce the observed distribution of genotypes. The *gvpA* and rDNA coding regions are present in multiple copies in at least some *Nodularia* isolates (all studied so far in the case of *gvpA*), and gene conversion could therefore explain changes in allele patterns without the need for gene exchange. If we take the rDNA locus first, there are three allele types: type I amplifiable with primer 9408R, type II amplifiable with 9427R and type III amplifiable with both primers. Although it might seem as though a gene conversion event could transform a type III isolate into a type I or II, sequence analysis of type III isolates has shown that the rDNA copies present are not identical to those found in either type I or type II isolates (Barker *et al.*, 1999). Gene conversion can not therefore account for the interconversion of any known rDNA types. Similarly, gene conversion is unlikely to explain differences in the *gvpA*-IGS genotype of our samples: although there are two copies of *gvpA*, there is just a single copy of the intergenic spacer that separates them. Of course we can not rule out the possibility that there was an ancestral *Nodularia* which contained all possible alleles at each locus, and that differential loss by gene conversion or deletion gave rise to the 12 types found today, but again gene exchange offers the most parsimonious explanation of the data.

This is the first population-level study suggestive of genetic exchange in a natural cyanobacterial population. The index of association is significantly different from zero, which means that whilst gene exchange could occur in *Nodularia*, the population is not panmictic. This conclusion is also supported by the randomization tests, which showed that whilst some combinations of filaments showed significantly more linkage disequilibrium than expected by chance, other combinations could not be distinguished from simulated random mating data (Table 6). Unfortunately it is not possible to estimate actual levels of recombination by analysing linkage disequilibrium (Maynard Smith *et al.*, 1993; Cohan, 1994), as this would require other factors

such as the mutation rate to be quantified. What is clear from the data is that recombination is not commonplace enough to completely randomize alleles in the population; if that were the case the index of association would not differ significantly from zero, and the randomization tests would show no significant combinations. There are three possible explanations for the non-zero index of association between loci. First, the loci studied may be closely linked (i.e. lie in close physical proximity); second, certain allelic combinations may be selectively advantageous; and third, the population may comprise a set of clones which have undergone many divisions without any genetic exchange with other genotypes.

Whilst analysis of observed and expected numbers for each pairwise combination of loci showed that rDNA-ITS II + *gvpA*-IGS II occurred less frequently than expected, we have no reason to suspect that this particular allele combination should have a low fitness value, especially as in the case of the rDNA region, sequence variation has only been detected in the ITS, and not in coding regions (Barker *et al.*, 1999). It is possible, however, that selection could be acting on regions tightly linked to those studied. It is unlikely that the three loci chosen are closely linked, but since we do not have any information about their physical location in the *Nodularia* genome we can not rule out this possibility. The most likely explanation for the observed degree of linkage disequilibrium seems to be that *Nodularia* undergoes periods of purely asexual growth, during which time certain genotypes grow better than others, but that gene exchange, by unknown mechanisms, occurs frequently enough to ensure that no particular allele at one locus ever occurs exclusively in combination with another from a separate genomic location: such a population structure is described as epidemic by Maynard Smith *et al.* (1993). There have been no other population genetic studies on cyanobacteria with which to compare the results obtained for *Nodularia*; therefore we have no idea whether the population structure observed for this taxon is typical of filamentous cyanobacteria or of the cyanobacteria as a whole. Given the variation in I_A in the genus *Bacillus* (Istock *et al.*, 1992) it seems unwise to extrapolate widely from our findings. Other studies have addressed genetic diversity in aquatic micro-organisms; for example Wise *et al.* (1996) used data from seven polymorphic allozyme loci to show that the mean gene diversity for a natural lotic population of *Burkholderia cepacia* was 0.52. Unfortunately a useful comparison of genetic diversity based on allozyme and DNA markers respectively is impossible. We hope that our data can form a baseline with which to compare future DNA-marker studies of other aquatic microbes.

That genetic exchange can occur between bacteria in the marine environment is not a novel finding; intra-generic conjugative transfer and transformation at frequencies of around 10^{-2} and 10^{-7} respectively have been observed in many *in vitro* studies with marine water and sediments (see review by Ashelford *et al.*, 1997) but all of

the studies to date have involved pseudomonads, *Escherichia coli* or *Vibrio* spp.

Comparison with data from a research cruise in 1996 (Hayes & Barker, 1997) shows that the frequency of the two PC-IGS alleles in the Baltic Sea changed dramatically over 2 years. In 1996, PC-IGS types I/II and III were present in almost equal numbers ($n = 91$ and 88 respectively), but in 1998 type I/II was rare in comparison to type III ($n = 355$ and 2335 respectively). Undoubtedly the abundance of filaments of genotype 9 (Table 1) is the major contributing factor in accounting for this shift in the allelic composition of the *Nodularia* population. We will not be able to explain such temporal shifts in population structure until we know more about the relative fitness of different *Nodularia* genotypes under various environmental conditions, and what impact cyanophage and other pathogens may have on dominant genotypes.

In terms of the taxonomy of *Nodularia* in the Baltic Sea, this study, as with our previous study (Barker *et al.*, 1999) indicates that the Baltic has only a single species, if one chooses to apply a biological species concept.

The data presented here lend support to the hypothesis of Rudi *et al.* (1998) that the occurrence of genetic exchange best explains the incongruent pattern of inheritance for the small-subunit rRNA and *rbclX* regions of many cyanobacteria. What our study shows is that gene exchange occurs in some cyanobacteria at a frequency great enough not just to show up over lengthy evolutionary time scales as demonstrated already, but also to shape the genetic structure of extant populations. We are now studying temporal changes in allele frequencies in greater detail, and starting to explore the mechanisms which facilitate genetic exchange in *Nodularia*.

ACKNOWLEDGEMENTS

We thank Annette Richer, Anja Engel, Klaus von Bröckel and the captains and crews of the FS *Litorina* and the RV *Alexander von Humboldt* for their assistance. This work was supported by the EC Environment programme (ENV4-CT97-0571), and the Natural Environment Research Council of the United Kingdom (GR9/3851, GR3/12415). J.R.S. is supported by the Wellcome Trust. This is ELOISE publication number 164.

REFERENCES

- Ashelford, K. E., Fry, J. C., Day, M. J., Hill, K. E., Learner, M. A., Marchesi, J. R., Perkins, C. D. & Weightman, A. J. (1997). Using microcosms to study gene transfer in aquatic habitats. *FEMS Microbiol Ecol* **23**, 81–94.
- Barker, G. L. A., Hayes, P. K., O'Mahony, S. L., Vacharapiyasophon, P. & Walsby, A. E. (1999). A molecular and phenotypic analysis of *Nodularia* (cyanobacteria) from the Baltic Sea. *J Phycol* **35**, 931–937.
- Barten, R. & Lill, H. (1995). DNA-uptake in the naturally competent cyanobacterium, *Synechocystis* sp. PCC-6803. *FEMS Microbiol Lett* **129**, 83–88.
- Bolch, C. J. S., Blackburn, S. I., Neilan, B. A. & Grewe, P. M. (1996). Genetic characterisation of strains of cyanobacteria using PCR-RFLP of the *cpcBA* intergenic spacer and flanking coding regions. *J Phycol* **32**, 445–451.
- Bolch, C. J. S., Orr, P. T., Jones, G. J. & Blackburn, S. I. (1999). Genetic, morphological and toxicological variation among globally distributed strains of *Nodularia* (cyanobacteria). *J Phycol* **35**, 339–355.
- Chiang, G. G., Schaefer, M. R. & Grossman, A. R. (1992). Transformation of the filamentous cyanobacterium *Fremyella diplosiphon* by conjugation or electroporation. *Plant Physiol Biochem* **30**, 315–325.
- Cohan, F. M. (1994). Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* **9**, 175–180.
- Eardly, B. D., Materou, L. A., Smith, N. H., Johnson, D. A., Rumbaugh, M. D. & Selander, R. K. (1990). Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. *Appl Environ Microbiol* **56**, 187–194.
- Feizabadi, M. M., Robertson, I. D., Cousins, D. V., Dawson, D. J. & Hampson, D. J. (1997). Use of multilocus enzyme electrophoresis to examine genetic relationships amongst isolates of *Mycobacterium intracellulare* and related species. *Microbiology* **143**, 1461–1469.
- Gordon, D. M. & FitzGibbon, F. (1999). The distribution of enteric bacteria from Australian mammals: host and geographical effects. *Microbiology* **145**, 2663–2671.
- Hagen, M. J. & Hamrick, J. L. (1996). Population level processes in *Rhizobium leguminosarum* bv *trifolii*: the role of founder effects. *Mol Ecol* **5**, 707–714.
- Haubold, B. & Rainey, P. B. (1996). Genetic and ecotypic structure of a fluorescent *Pseudomonas* population. *Mol Ecol* **5**, 747–761.
- Hayes, P. K. & Barker, G. L. A. (1997). Genetic diversity within Baltic Sea populations of *Nodularia* (cyanobacteria). *J Phycol* **33**, 919–923.
- Istock, C. A., Duncan, K. E., Ferguson, N. & Zhou, X. (1992). Sexuality in a natural population of bacteria – *Bacillus subtilis* challenges the clonal paradigm. *Mol Ecol* **1**, 95–103.
- Komárek, J., Hübel, M., Hübel, H. & Šmarda, J. (1993). The *Nodularia* studies. 2. Taxonomy. *Arch Hydrobiol Algal Stud* **68**, 1–25.
- Lorenz, M. G. & Wackernagel, W. (1994). Bacterial gene-transfer by natural genetic-transformation in the environment. *Microbiol Rev* **58**, 563–602.
- Mantel, N. (1967). The detection of disease clustering and a generalised regression approach. *Cancer Res* **27**, 209–220.
- Maynard Smith, J. (1995). Do bacteria have population genetics? In *Population Genetics of Bacteria* (Society for General Microbiology Symposium no. 52), pp. 1–12. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington & J. R. Saunders. Cambridge: Cambridge University Press.
- Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384–4388.
- Muropastor, A. M., Kuritz, T., Flores, E., Herrero, A. & Wolk, C. P. (1994). Transfer of a genetic-marker from a megaplasmid of *Anabaena* sp. strain PCC-7120 to a megaplasmid of a different *Anabaena* strain. *J Bacteriol* **176**, 1093–1098.
- Nei, M. (1972). Genetic distance between populations. *Am Nat* **106**, 283–292.
- Nei, M. (1987). Estimation of average heterozygosity and genetic distance from small numbers of individuals. *Genetics* **89**, 583–590.

- Neilan, B. A., Jacobs, D. & Goodman, A. E. (1995). Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphism within the phycocyanin locus. *Appl Environ Microbiol* **61**, 3875–3883.
- Rudi, K., Skulberg, O. M. & Jakobsen, K. S. (1998). Evolution of cyanobacteria by exchange of genetic material among phyletically related strains. *J Bacteriol* **180**, 3453–3461.
- Silva, C., Eguiarte, L. E. & Souza, V. (1999). Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol Ecol* **8**, 277–287.
- Singh, D. T., Bagchi, S. N., Modi, D. R. & Singh, H. N. (1987). Evidence for intergeneric transformation in filamentous, diazotrophic cyanobacteria. *New Phytol* **107**, 347–356.
- Sode, K., Tataru, M., Takeyama, H., Burgess, J. G. & Matsunaga, T. (1992). Conjugative gene-transfer in marine cyanobacteria *Synechococcus* sp., *Synechocystis* sp. and *Pseudanabaena* sp. *Appl Microbiol Biotechnol* **37**, 369–373.
- Spratt, B. G., Smith, N. H., Zhou, J., O'Rourke, M. & Feil, E. (1995). The population genetics of the pathogenic *Neisseria*. In *Population Genetics of Bacteria* (Society for General Microbiology Symposium no. 52), pp. 143–160. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington & J. R. Saunders. Cambridge: Cambridge University Press.
- Stevens, J. R. & Tibayrenc, M. (1995). Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterised by RAPD analysis and isoenzymes. *Parasitology* **110**, 181–186.
- Stevens, J. R. & Tibayrenc, M. (1996). *Trypanosoma brucei* s.l.: evolution, linkage and the clonality debate. *Parasitology* **112**, 481–488.
- Tandeau de Marsac, N., Mazel, D., Bryant, D. A. & Houmard, J. (1985). Molecular-cloning and nucleotide-sequence of a developmentally regulated gene from the cyanobacterium *Calothrix*-PCC-7601 – a gas vesicle protein gene. *Nucleic Acids Res* **13**, 7223–7236.
- Thiel, T. & Wolk, C. P. (1987). Conjugal transfer of plasmids to cyanobacteria. *Methods Enzymol* **153**, 232–243.
- Trott, D. J., Oxberry, S. L. & Hampson, D. J. (1997). Evidence for *Serpulina hyodysenteriae* being recombinant with an epidemic population structure. *Microbiology* **143**, 3357–3365.
- Walsby, A. E., Hayes, P. K. & Boje, R. (1995). The gas vesicles, buoyancy and vertical distribution of cyanobacteria in the Baltic Sea. *Eur J Phycol* **30**, 87–94.
- Whittam, T. S. (1995). Genetic population structure and pathogenicity in enteric bacteria. In *Population Genetics of Bacteria* (Society for General Microbiology Symposium no. 52), pp. 217–245. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington & J. R. Saunders. Cambridge: Cambridge University Press.
- Wilmotte, A. (1994). Molecular evolution and taxonomy of the cyanobacteria. In *The Molecular Biology of Cyanobacteria*, pp. 1–25. Edited by D. A. Bryant. Dordrecht: Kluwer.
- Wilson, W. H. & Mann, N. H. (1997). Lysogenic and lytic viral production in marine microbial communities. *Aquat Microb Ecol* **13**, 95–100.
- Wise, M. G., McArthur, J. V., Wheat, C. & Shimkets, L. J. (1996). Temporal variation in genetic diversity and structure of a lotic population of *Burkholderia* (*Pseudomonas*) *cepacia*. *Appl Environ Microbiol* **62**, 1558–1562.

Received 12 April 2000; revised 21 July 2000; accepted 31 July 2000.