

Technical University of Denmark



Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers

Applied metapopulation theory for marine microbes

Lindh, Markus V.; Sjöstedt, Johanna; Ekstam, Börje; Casini, Michele; Lundin, Daniel; Hugerth, Luisa W.; Hu, Yue O. O.; Andersson, Anders F.; Andersson, Agneta; Legrand, Catherine; Pinhassi, Jarone

Published in:

Environmental Microbiology

Link to article, DOI:

[10.1111/1462-2920.13650](https://doi.org/10.1111/1462-2920.13650)

Publication date:

2017

[Link back to DTU Orbit](#)

Citation (APA):

Lindh, M. V., Sjöstedt, J., Ekstam, B., Casini, M., Lundin, D., Hugerth, L. W., ... Pinhassi, J. (2017). Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers: Applied metapopulation theory for marine microbes. *Environmental Microbiology*, 19(3), 1222-1236. DOI: 10.1111/1462-2920.13650

DTU Library

Technical Information Center of Denmark



General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers

Markus V. Lindh,^{1†} Johanna Sjöstedt,^{1†¶}
Börje Ekstam,² Michele Casini,³ Daniel Lundin ¹,
Luisa W. Hugerth,⁴ Yue O. O. Hu,⁴
Anders F. Andersson,⁴ Agneta Andersson,⁵
Catherine Legrand¹ and Jarone Pinhassi ^{1*}

¹Centre for Ecology and Evolution in Microbial model Systems - EEMiS, Linnaeus University, Kalmar SE-39182, Sweden.

²Department of Biology and Environmental Science, Linnaeus University, Kalmar SE-39182, Sweden.

³Department of Aquatic Resources, Institute of Marine Research, Swedish University of Agricultural Sciences, Lysekil SE-45330, Sweden.

⁴Science for Life Laboratory School of Biotechnology, KTH Royal Institute of Technology, Stockholm SE-10691, Sweden.

⁵Department of Ecology and Environmental Science, Umeå University, Umeå SE-90187, Sweden.

†

Summary

Metapopulation theory developed in terrestrial ecology provides applicable frameworks for interpreting the role of local and regional processes in shaping species distribution patterns. Yet, empirical testing of metapopulation models on microbial communities is essentially lacking. We determined regional bacterioplankton dynamics from monthly transect sampling in the Baltic Sea Proper using 16S rRNA gene sequencing. A strong positive trend was found between local relative abundance and occupancy of populations. Notably, the occupancy-frequency distributions were significantly bimodal with a satellite

mode of rare endemic populations and a core mode of abundant cosmopolitan populations (e.g. *Synechococcus*, SAR11 and SAR86 clade members). Temporal changes in population distributions supported several theoretical frameworks. Still, bimodality was found among bacterioplankton communities across the entire Baltic Sea, and was also frequent in globally distributed datasets. Datasets spanning waters with widely different physicochemical characteristics or environmental gradients typically lacked significant bimodal patterns. When such datasets were divided into subsets with coherent environmental conditions, bimodal patterns emerged, highlighting the importance of positive feedbacks between local abundance and occupancy within specific biomes. Thus, metapopulation theory applied to microbial biogeography can provide novel insights into the mechanisms governing shifts in biodiversity resulting from natural or anthropogenically induced changes in the environment.

Introduction

Marine microorganisms regulate ecosystem services through their influence on fluxes of matter and energy in the ocean (Falkowski *et al.*, 2008). Conspicuously, these fluxes are ever dynamic due to the pronounced potential of microbial communities to rapidly respond to environmental change, both through adjustments in species composition and metabolic activity (Allison and Martiny, 2008; Gómez-Consarnau *et al.*, 2012; Logue *et al.*, 2015). Historically, local communities have been regarded as independent ecological units with individual integrity, determined by local interactions among coexisting species and environmental factors (Lawton, 1999). However, over the past decades, the profound influence of regional factors, e.g. dispersal, in structuring local species assemblages has been established (Harrison and Cornell, 2008; Ricklefs, 2008; Östman *et al.*, 2010; Logue *et al.*, 2011; Lindström and Langenheder, 2012). Yet, despite the recognized importance of microbial biogeographical patterns (see e.g. Ghiglione *et al.*, 2012; Sul *et al.*, 2013; Zinger *et al.*, 2014),

Received 14 September, 2016; revised 10 December, 2016; accepted 18 December, 2016. *For correspondence. E-mail jarone.pinhassi@lnu.se; Tel. +46-480-446212; Fax +46 480 447305. Present addresses: [†]Department of Biology, Lund University, Lund SE-22362, Sweden; [‡]Department of Biology/Aquatic Ecology, Lund University, Lund SE-22362, Sweden; [¶]Centre for Ocean Life, Institute for Aquatic Resources, Technical University of Denmark, Charlottenlund 2900, Denmark

the mechanistic processes involved in structuring bacterioplankton communities remain largely unconstrained (Martiny *et al.*, 2006; Hanson *et al.*, 2012; Lindström and Langenheder, 2012).

Metapopulation theory has mainly been used to characterize the dynamics of macroorganisms in patchy environments, i.e. how local colonization and extinction rates interact with species distribution at a regional scale. Colonization (C) and extinction (E) rates are typically calculated from the temporal change (dt) in fraction of sites occupied (P) of populations in a given ecosystem. Changes in occupancy (dP) is thus calculated as $dP/dt = C - E$ (Hanski and Gyllenberg, 1993). When applied in terrestrial ecology, these theoretical models give explanatory power to identify the drivers of distribution patterns among insects, plants and animals (McGeoch and Gaston, 2002), providing knowledge critical for developing strategies in conservation biology. Two major models have attracted important attention; Levin's model and the core-satellite hypothesis (CSH) (Levin, 1974; Hanski, 1982). The two original models differ in how extinction rates are calculated (linear according to Levin's model and quadratic according to the CSH model; Supporting Information Fig. S1A and B) to predict occupancy-frequency distributions (the number of species that occupy different number of sites). Further, Levin's model predicts a unimodal occupancy-frequency distribution characterized by a skewed pattern where most species in a region occupy a single site and the number of species decreases with increasing number of sites (Levin, 1974; Supporting Information Fig. S1A). The decrease in number of species occupying increasing number of sites is predicted from a linear relationship between extinction rates and occupancy (Supporting Information Fig. S1A). Accordingly, as the frequency of extinction is calculated across all sites in a given ecosystem and all patches are assumed to have an equal risk of extinction (Levin, 1974), common species with high occupancy are more likely to go extinct due to the high number of patches they occupy, i.e. the risk to go extinct increases with occupancy.

In contrast, the CSH predicts a bimodal pattern where the decrease in number of species with increasing number of study sites is followed by an increase in species occupying all or most sites (Hanski, 1982; Supporting Information Fig. S1B). Hanski's metapopulation model (Hanski, 1982; Brown, 1984; Gaston *et al.*, 2000) predicts a quadratic function of both colonization and extinction rates and takes into account positive feedback mechanisms, i.e. a rescue effect, between local abundance and occupancy (Supporting Information Fig. S1B). Extinction rates are therefore predicted to be low for populations with high and low occupancy but high for populations with intermediate occupancy (Supporting Information Fig. S1B). Variations of both Levin's model and the CSH have been elaborated

over time to explore whether colonization rates are independent of occupancy (Gotelli, 1991; Hanski and Gyllenberg, 1993). In particular, (Gotelli, 1991) proposed a mechanism where colonization can be independent of occupancy due to a constant propagule rain or dormant seedbank, thus having linear relationships between both colonization- and extinction rates and occupancy (Supporting Information Fig. S1C and D). Still, there are few, if any, empirical tests of these models using high-throughput molecular data in aquatic habitats in general and for microorganisms in particular. Taken together, application of metapopulation theory may resolve the underlying mechanisms of microbial species distributions, and could improve predictions and interpretations of patterns in biodiversity responding to environmental change.

Here we aimed to investigate the importance of regional effects in shaping monthly sampled local bacterioplankton communities (operational taxonomic units [OTUs], conservatively defined at 97% 16S rRNA gene sequence identity) in the Baltic Sea Proper. Our objectives were to determine (i) the relationship between local relative abundance and occupancy, (ii) the type of occupancy-frequency distribution patterns, and (iii) colonization and extinction rates. We further aimed to identify OTUs showing distinct distributional and colonization/extinction patterns. To validate observed distribution patterns, the analysis of occupancy-frequency distributions was extended to complementary datasets spanning the entire Baltic Sea as well as global International Census of Marine Microbes (ICoMM) datasets. We also evaluated observed OTU distributions with beta-diversity patterns and variability in environmental conditions available as metadata in the ICoMM database.

Material and methods

Field sampling

In 2010–2011 during the productive season (April–October), spatiotemporal dynamics of bacterioplankton communities were investigated at 16 stations off the east coast of Sweden in the Baltic Sea Proper (Supporting Information Fig. S2). The area and sampling of physicochemical parameters are detailed in Bertos-Fortis *et al.*, (2016) and supporting information.

DNA extraction, 454 PCR, sequence processing and analysis

Collection of DNA and amplicon processing from the 117 collected samples was carried out as previously described (Bertos-Fortis *et al.*, 2016), and see supporting information. The study area (Supporting Information Fig. S2) was chosen as part of a collaboration between the projects PLANFISH and EcoChange to study food-web dynamics in a coastal to open ocean gradient. For details on hydrological conditions in the study area please see (Díaz-Gil *et al.*, 2014; Legrand *et al.*, 2015; Bertos-Fortis *et al.*, 2016). Seawater from each

station was collected in acid washed Milli-Q rinsed polycarbonate bottles, at discrete depths (2, 4, 6, 8, and 10 m) that were pooled and filtered shipboard. All 16 stations were represented in 2010 and classified according to their spatial distribution (coastal-offshore) as coastal (C), middle (M) and open (O) sea stations. During 2011, only coastal (C) and open (O) stations were sampled.

DNA from biomass collected onto 0.2 µm-pore-size filters was extracted according to (Riemann *et al.*, 2000). Amplicon processing was performed as in (Lindh *et al.*, 2015). Rarefaction curves for all samples are provided in Supporting Information Fig. S3. The final OTU table delineated at 97% 16S rRNA gene sequence identity contained 100 samples, each with 3000 sequences, and a total of 10951 different OTUs (excluding singletons and chloroplast sequences). Delineation at 97% 16S rRNA gene sequence identity represents a conservative estimate of phylogenetic diversity in a sample. For the Baltic Sea dataset we obtained 16S rRNA gene bacterial sequence data (delineated at 99% 16S rRNA gene fragment sequence identity) from (Hu *et al.*, 2016), (see also Supporting Information). Thirty-one OTU tables (delineated at 97% 16S rRNA gene fragment sequence identity) of bacterial sequences from ICoMM projects were downloaded from <http://vamps.mbl.edu/index.php>, see details in Supporting Information. To evaluate the robustness of using phylogenetic definition of populations, we tested the effect of clustering our Baltic Sea Proper data at different levels of 16S rRNA gene sequence identities. It is here important to note that a portion of rare OTUs may, from purely technical OTU assembly and clustering reasons, be influenced by sequences derived from closely related abundant OTUs (see e.g. He *et al.*, 2015). However, the large variation in observed patterns of a majority of rare OTUs indicates that such methodological biases had limited effect on our conclusions in the present study.

Statistical analyses and graphical outputs

In the present paper we have defined abundant and rare OTUs as having relative abundances >1% and <0.1%, respectively (Pedros-Alio, 2006). Furthermore, we defined 'core' and 'satellite' populations as having average local relative abundance >0.01 and average occupancy (P) >0.9, and average local relative abundance <0.001 and average $P = 0.01$, respectively). To elucidate how our observed data fit with current models, we calculated rates of colonization and extinction by comparing occupancies of OTUs between succeeding months, i.e. the number of new sites an OTU had colonized compared with the number of previously occupied sites from which it had disappeared (Fig. 3). As in previous work (Wardle *et al.*, 2011), we considered the terms colonization and extinction in a broad sense as analogous to gain and loss of OTUs, i.e. local extinction is going from presence to absence. Accordingly, absence was indicated by local relative abundance <0.006%, given the detection limit of our sequencing methodology. Conversely, colonization represented the shift from absence (i.e. below detection limit) to presence, irrespectively of whether this is the result of on-site growth or from immigration of populations from neighboring waters. An equivalent to Tokeshi's test of bimodality was performed using Mitchell-Olds' and Shaw's test (Mitchell-Olds and Shaw, 1987)

for the location of quadratic extremes. Non-linear least squares analysis was performed using the equations of Levin's model, the CSH hypothesis and Gotelli's propagule rain model (Levin, 1974; Hanski, 1982; Gotelli, 1991).

Levin's original model (Levin, 1974) is as follows:

$$dP/dt = CP(1 - P) - EP \quad (1)$$

where P is the fraction of occupied sites, C is colonization rate and E is extinction rate. When $P = 1$, occupancy is 100% and all sites are occupied and when $P = 0$, occupancy is 0%, meaning regional extinction. Colonization rate is the number of colonized empty sites over time. If C is greater than E plus the variance in E ($C > E + \sigma^2_E$), then Levin's model predicts a unimodal distribution. Hanski's model (Hanski, 1982) is as follows:

$$dP/dt = CP(1 - P) - EP(1 - P) \quad (2)$$

Thus, if the variance in S ($C - E$) is greater than S (i.e. $\sigma^2_S > S$) the model predicts a bimodal species distribution.

Variations in Levin's and Hanski's models were further tested by including propagule rain into the non-linear least squares analyses according to (Gotelli, 1991), calculated as follows for Levin's and Hanski's models, respectively:

$$dP/dt = C(1 - P) - EP \quad (3)$$

$$dP/dt = C(1 - P) - EP(1 - P) \quad (4)$$

In Gotelli's framework (Gotelli, 1991) the model assumption is that colonization does not originate from local populations, but from a constant propagule rain, or dormant seedbank. Colonization is thus independent of P and based on a linear relationship between colonization rate (C) and occupancy.

All statistical tests and graphical outputs were performed in R 3.2.2 (R Core Team, 2014), using the package Vegan (Jari Oksanen *et al.*, 2010) and ggplot2 (Wickham, 2009). Supporting Information Figs S2 and S5A were made using ODV (Version 4.5.0; Schlitzer 2015).

Results and discussion

Differences in community structure among Baltic Sea Proper sites

We investigated bacterioplankton population dynamics to elucidate the role of regional as compared with local dynamics in shaping biogeographical patterns (Supporting Information Fig. S2). Samples of bacterioplankton community composition in the grid of 16 stations in the Baltic Sea Proper over two years (2010-2011) clustered according to season, with a tendency towards differentiation between coastal and offshore sites (Supporting Information Fig. S4). This agreed with the recognized importance of spatio-temporal shifts in environmental conditions for shaping community composition of marine bacteria in the Baltic Sea and elsewhere (Fuhrman *et al.*, 2006; Gilbert *et al.*, 2012; Lindh *et al.*, 2015; Bunse *et al.*, 2016).

Distribution of rare and abundant species

Across all monthly transect samplings, between 201 and 401 OTUs were detected in each sample. Among those OTUs, between 8–25 OTUs on each sampling date were abundant whereas a grand majority (typically 96–343 OTUs) was rare (Supporting Information Fig. S5). This pattern is typical of aquatic microbial communities, and widespread also among macroorganisms in terrestrial environments (Morse *et al.*, 1988; Gaston *et al.*, 2000; Rappé *et al.*, 2000; Magurran and Henderson, 2003; Pedros-Alio, 2006; Sogin *et al.*, 2006; Unterseher *et al.*, 2011).

The pattern of many rare compared with a few abundant populations for organisms in a wide variety of habitats may be inherently linked with strong regional dynamics, i.e. a positive relationship between local abundance and regional occupancy. Such regional dynamics have been established also for bacterioplankton (Blackburn *et al.*, 2006; Heino and Virtanen, 2006; Östman *et al.*, 2010). Accordingly, we found a positive relationship between average local relative abundance of OTUs with average occupancy (i.e. the number of sites occupied by specific OTUs) (Fig. 1). Given this pattern characteristic for a wide range of organisms and ecosystems, including both macroorganisms and highly diverse phytoplankton and bacterioplankton assemblages (Hanski, 1982; Brown, 1984; Gaston *et al.*, 2000), our results suggest that mechanisms shaping regional dynamics typically observed among macroorganisms outside the marine realm could potentially shape also marine bacterioplankton assemblages.

Evaluation of metapopulation models in the Baltic Sea Proper

The relationship between local relative abundance and regional occupancy may be derived from and be directly linked to OTU abundance patterns. Analysis of distribution patterns among OTUs showed a recurring occupancy-frequency distribution that was significantly bimodal in all eleven sampled months (Mitchell-Olds & Shaw's test, $P < 0.05$, $n = 6$ –12; Fig. 2). The observed bimodal pattern was remarkably stable and persisted throughout all levels of clustering of OTUs, and including or excluding singletons, but with a weaker tendency toward bimodality at 100% OTU cut-offs (Supporting Information Fig. S6). In fact, differences in the number of OTUs found at all stations compared with the number only found at a single station were not significantly different between 97% and 99% 16S rRNA gene sequence delineation nor with or without singletons (Supporting Information Fig. S7; Repeated measures ANOVA, Tukey's test, $P < 0.05$, $n = 100$). Yet, there were significant differences in the number of core populations between 100% and 97% OTU

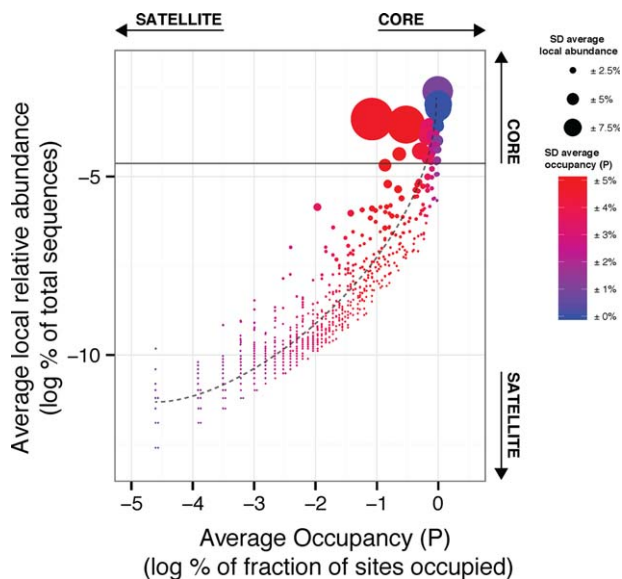


Fig. 1. Average relative local abundance (normalized sequence abundance) against average occupancy. Color denotes standard deviation in occupancy where red is high and blue is low variation. Size denotes standard deviation in local relative abundance where large size is high and small size is low variation. By moving towards left and down on the x - and y -axis, respectively, OTUs display satellite characteristics schematically indicated by arrows. By moving towards right and up on the x - and y -axis, respectively, OTUs display core characteristics. Grey dotted line is smoothing curve (LOESS). Grey line indicates relative abundance above 1% (log -4.6).

clustering cut-offs (Supporting Information Fig. S7; Repeated measures ANOVA, Tukey's test, $P = 0.01$, $n = 100$) but not in the number of satellite populations (Supporting Information Fig. S7; Repeated measures ANOVA, Tukey's test, $P < 0.05$, $n = 100$).

These results match predictions of bimodality in the CSH proposed by (Hanski, 1982) rather than predictions of unimodality in Levin's model (Levin, 1974). The observed bimodal pattern was remarkably stable and persisted throughout several OTU clustering levels and irrespectively of including or excluding singletons, (see Supporting Information Figs S6 and S7). To our knowledge, this observed bimodal pattern represents the first finding of bimodality both for prokaryotic communities in general, and for any organism in marine environments.

Observed patterns in regional colonization rates followed a quadratic curve, where colonization (C) at first increased with fraction of sites occupied (P) until about 50–65% occupancy, followed by a decrease in C until maximum P (Fig. 3). Extinction rates followed a similar quadratic pattern, where extinction (E) first increased with P , but until a higher occupancy compared with C (between 65 and 75%), followed by a decrease of E until maximum P (Fig. 3). We examined how metapopulation models could explain variations in colonization and extinction rates

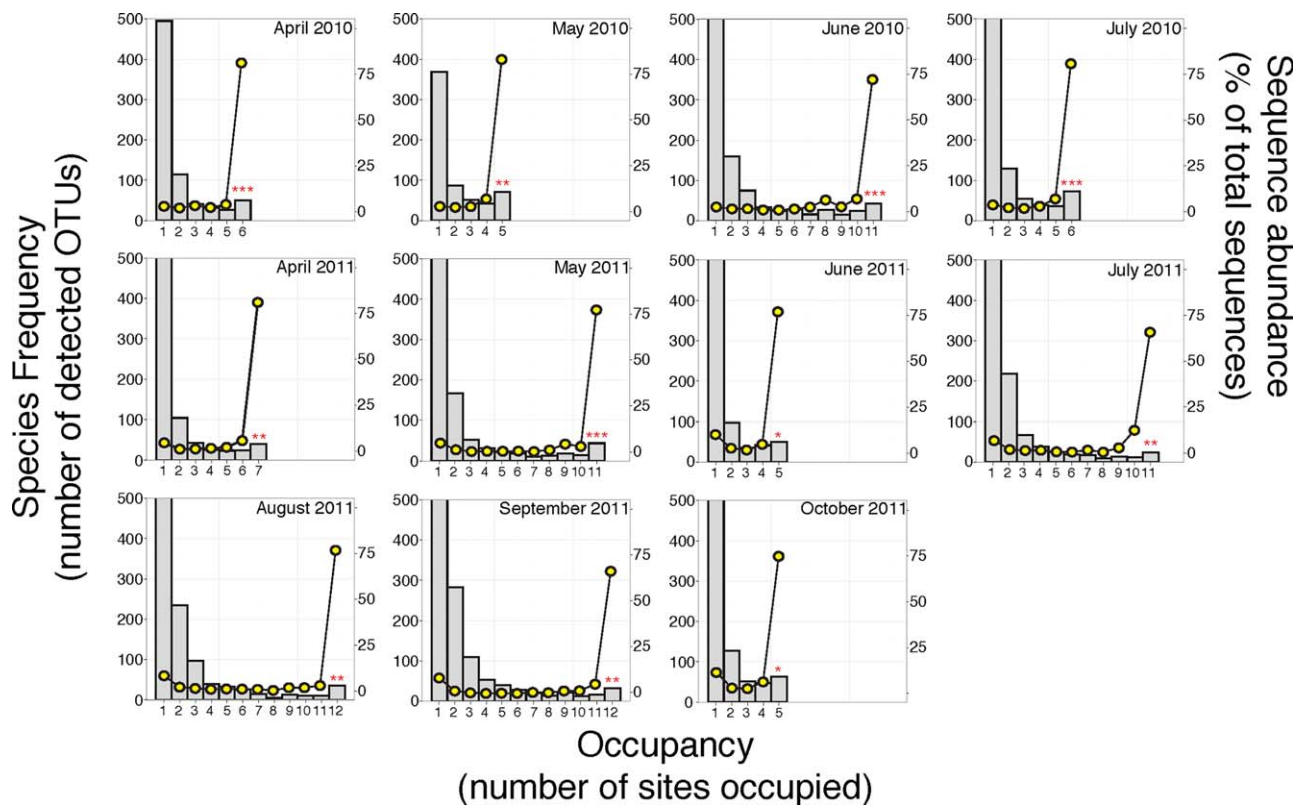


Fig. 2. Occupancy-frequency distribution of OTUs in our spatiotemporal study of the Baltic Sea Proper (Western Gotland basin and Kalmar sound). Filled yellow circles indicate relative sequence abundance (percentage of total sequences in each sample). Asterisks denote significance levels (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) for Mitchell-Olds & Shaw's test. The maximum number of sites sampled each month and at each station (i.e. maximum occupancy) is given by the x-axes. [Colour figure can be viewed at wileyonlinelibrary.com]

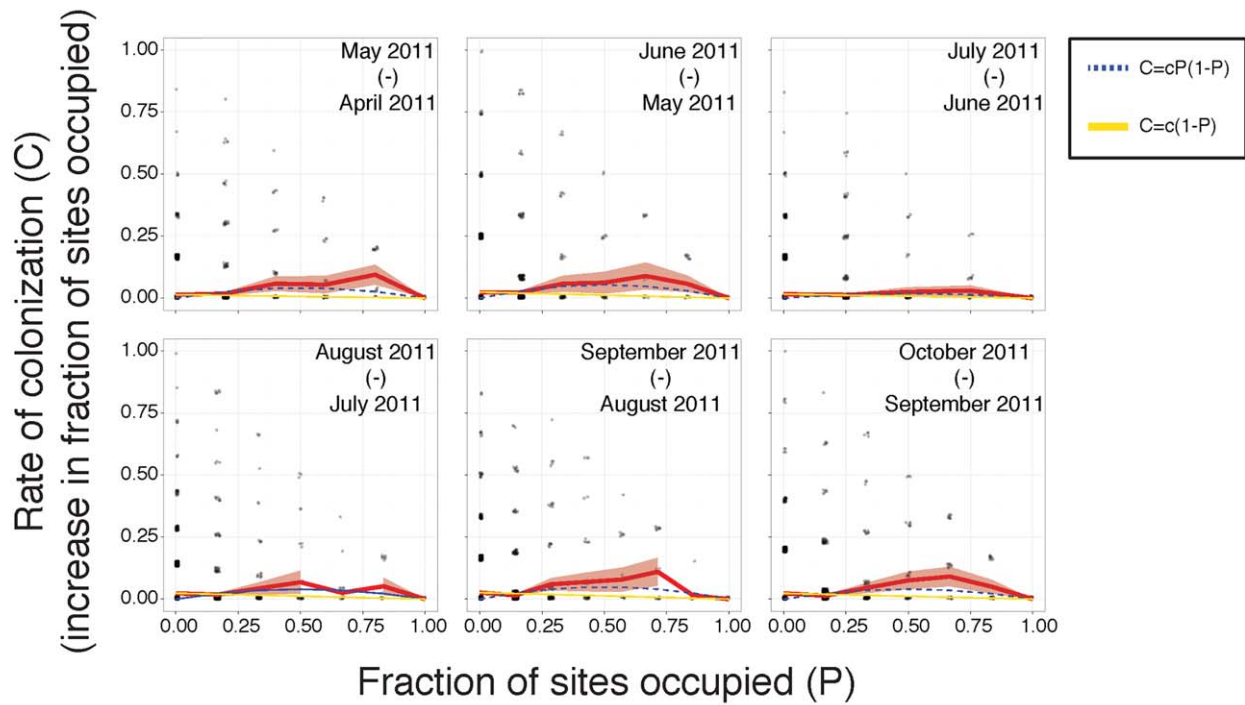
by using non-linear least squares analysis. Predictions based on the CSH (Hanski, 1982) agreed well with the field results for extinction rates (See blue dashed lines in Fig. 3), with on average low residual standard error, whereas colonization rates had higher residual standard error (Supporting Information Table S1). The CSH significantly explained both colonization and extinction rates. Predictions based on Levin's model (Levin, 1974) showed overall higher residual standard errors for extinction rates compared with the CSH (Supporting Information Tables S1 and S2). Notably, both models significantly explained observed extinction rates, although extinction rates predicted by Levin's model had lower parameter estimates compared with the CSH predictions (Supporting Information Tables S1 and S2). In addition, linear prediction of colonization rates suggested by Gotelli (1991), also resulted in significant fits with observed colonization rates (Supporting Information Table S3). Gotelli's propagule rain model in fact had lower residual standard error compared with both Levin's and Hanski's quadratic models of colonization rates. Thus, all models produced significant fits for quadratic and linear colonization and extinction rates. However, since the CSH prediction of bimodal occupancy-

frequency patterns matched our data on OTU distributions, whereas Levin's and Gotelli's predictions of unimodality did not, the CSH yielded a better overall agreement with the observed data. The bimodality predicted in the CSH arises because colonization and extinction rates depend on occupancy (Hanski, 1982). In agreement, our results show that colonization and extinction rates are coupled with the regional occurrence of bacterioplankton populations. This agreement of theoretical predictions and measured bacterioplankton population dynamics favored the CSH for interpreting regional effects of OTU distributions in the Baltic Sea Proper.

Linking metapopulation dynamics and specific populations in the Baltic Sea proper

In the Baltic Sea Proper dataset, the number of endemic satellite bacterioplankton populations (i.e. OTUs only found in one site each month) varied between 380 and 2201 OTUs. In contrast, the much fewer cosmopolitan core populations (i.e. found in all sites each month) ranged between 25 and 70 OTUs. A striking consequence of the steep rank abundance distribution patterns was that the

A



B

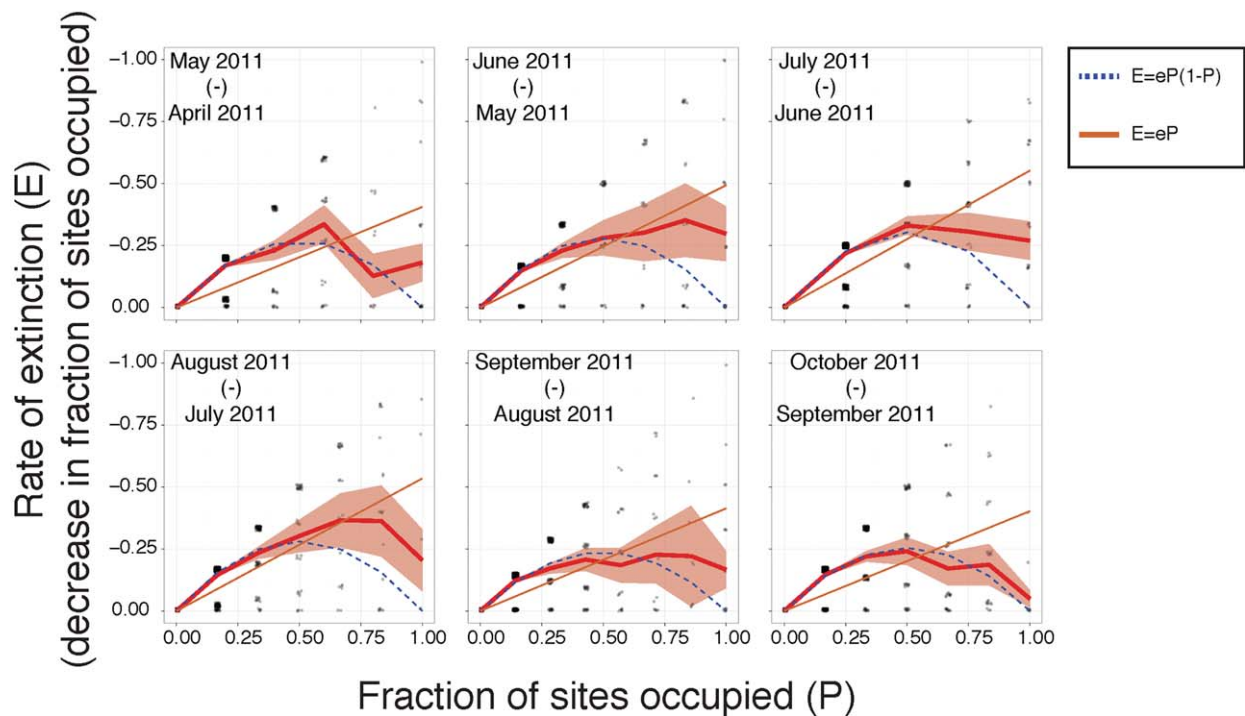


Fig. 3. Colonization rates (A), and extinction rates (B) against occupancy (P). Each point represents the change in fraction of sites occupied from one month to the next for individual OTUs, in total 4437 OTUs for all months. The red line indicates mean and the red shaded area around the mean is standard error. Blue dashed, yellow and brown lines indicate the curve of observed data fitted by non-linear least squares to the metapopulation model by Hanski (1982), Gotelli (1991) and Levin (1974), respectively, (Supporting Information Table S1). Examples are from April–October 2011. OTU positions are jittered to reduce overplotting and colored in grey scale according to OTU density.

core OTUs found at all sites during one sampling month contributed to between 60 and 80% of the sequence abundance that month (see yellow filled circles in Fig. 2). The sequence abundance patterns described in Fig. 2 were observed also at 99% and 100% OTU clustering of 16S rRNA gene sequences, and both when including and excluding singletons (Supporting Information Fig. S8).

Given the pronounced range between satellite to core populations, we aimed to link the identity of particular OTUs with observed metapopulation dynamics. First, abundant populations with high relative abundance and average occupancy (i.e. average local relative abundance >0.01 and average $P > 0.9$, respectively) were classified as 'core'; corresponding to Hanski's core populations, ($n = 17$) (Supporting Information Table S4). 'Core' populations such as those belonging to the *Synechococcus* (PF_000001), CL500-29 (PF_000000), SAR11 (PF_000002), SAR86 (PF_000007), hgcl (PF_000007), NS3a clades (PF_000008) and *Spartobacteria* (PF_000010) were almost never found in the rare fraction in our study, and were at nearly all times part of the rightmost column in each of the panels in Fig. 2 (Supporting Information Table S4). The presence of core populations affiliated with the SAR11 and SAR86 clades and *Synechococcus* is in agreement with the cosmopolitan distribution of these lineages in marine environments (Pommier *et al.*, 2007; Ghiglione *et al.*, 2012; Gibbons *et al.*, 2013). It is generally recognized that 'core' populations often have advantages in utilization of widespread resources (Brown, 1984; Hanski and Gyllenberg, 1993). Hence, 'core' OTUs, even though they adhere to an oligotrophic life strategy, may be linked to the concept of generalist dynamics (Gómez-Consarnau *et al.*, 2012; Teeling *et al.*, 2012). Irrespectively of the precise physiological mechanisms, our data suggest that core OTUs depend largely on maintaining high levels of occupancy, implying an increased probability of colonizing new sites and sustaining large population sizes through positive feedback mechanisms.

Second, populations fluctuating in occupancy and local relative abundance (SD $P > 0.05$ and SD relative abundance >0.05 , respectively) were classified as 'transient' ($n = 13$) (Supporting Information Table S4). For example, two *Planctomycetaceae* OTUs (PF_000021; PF_000073), a *Roseibacillus* (PF_000037) and a *Comamonadaceae* OTU (PF_000206) were transient. Interestingly, all transient populations displayed intermediate occupancy around 50%. Temporal transitions are common in bacterioplankton communities and appear to occur between all abundance classes (Campbell *et al.*, 2011; Alonso-Sáez *et al.*, 2015; Lindh *et al.*, 2015).

Third, rare populations ($<0.1\%$ of average local relative abundance) with high occupancy ($P > 0.5$) were classified as 'continuously rare' ($n = 11$) (Supporting Information Table S4). It is noteworthy that these continuously rare

populations, e.g. *Fluviicola* (PF_000123), *Pedobacter* (PF_000205), and *OM182 clade* (PF_000144) exhibited high occupancy. Similarly, Galand *et al.*, (2009) observed that rare phylotypes remained rare when surface seawater samples were compared with deep waters from the Arctic, highlighting that most rare phylotypes were continuously rare within that ecosystem. This suggests that some populations may have a permanent and widespread ecological strategy of being rare.

Fourth, rare populations with low occupancy were classified as 'satellite'; (i.e. average local relative abundance <0.001 and average $P = 0.01$) equivalent to Hanski's satellite populations, ($n = 9987$) (Supporting Information Table S4). We exemplify the 20 most extreme cases of satellite populations with OTUs only detected at a single site (Supporting Information Table S7). Thus, for example 'satellite' populations were affiliated with SAR11 (PF_001188), SAR86 (PF_000937), *Psychrobacter* (PF_000615) and *Owenweeksia* (PF_000426). In accordance, rare OTUs in the Arctic have highly restricted geographic ranges despite the potential for high dispersal and low loss rates (Galand *et al.*, 2009). Satellite characteristics such as these indicate a pronounced biogeography of the rare biosphere, likely caused by the high extinction rates observed.

We note that extinction rates were always higher than colonization rates ($>12.5\%$ and $<12.5\%$, respectively). Higher rates of extinction compared with colonization imply that a few populations with higher colonization compared with extinction rates cause bimodal occupancy-frequency patterns but most populations are headed towards local extinction and regional rarity. The quadratic relationships for colonization and extinction rates at different occupancies highlight the importance of regional dynamics for local populations with a decreasing probability for extinction with increasing occupancy. Overall, these results substantiate the hypothesis that the effect of stochastic variation is highest for populations with intermediate occupancy. A corollary of this scenario is that it is far more likely for common populations than for rare ones to become more common, and more likely for rare populations to become even more rare. However, this particular scenario is only possible under the assumption that dispersal occurs equally for all species. Studies in the macroecological context have shown that this is not a correct assumption for plants and animals (see e.g. Mehranvar and Jackson, 2001). Still, it has been suggested that microbial eukaryotes have ubiquitous dispersal given their abundance (Finlay and Clarke, 1999; Finlay, 2002). In fact, different microbial traits might significantly alter their dispersal abilities; for example, dormancy mechanisms may facilitate dispersal of certain bacterial groups (Lennon and Jones, 2011). However, further information on the dispersal capacity of particular microbial species is needed. Nevertheless, we validated

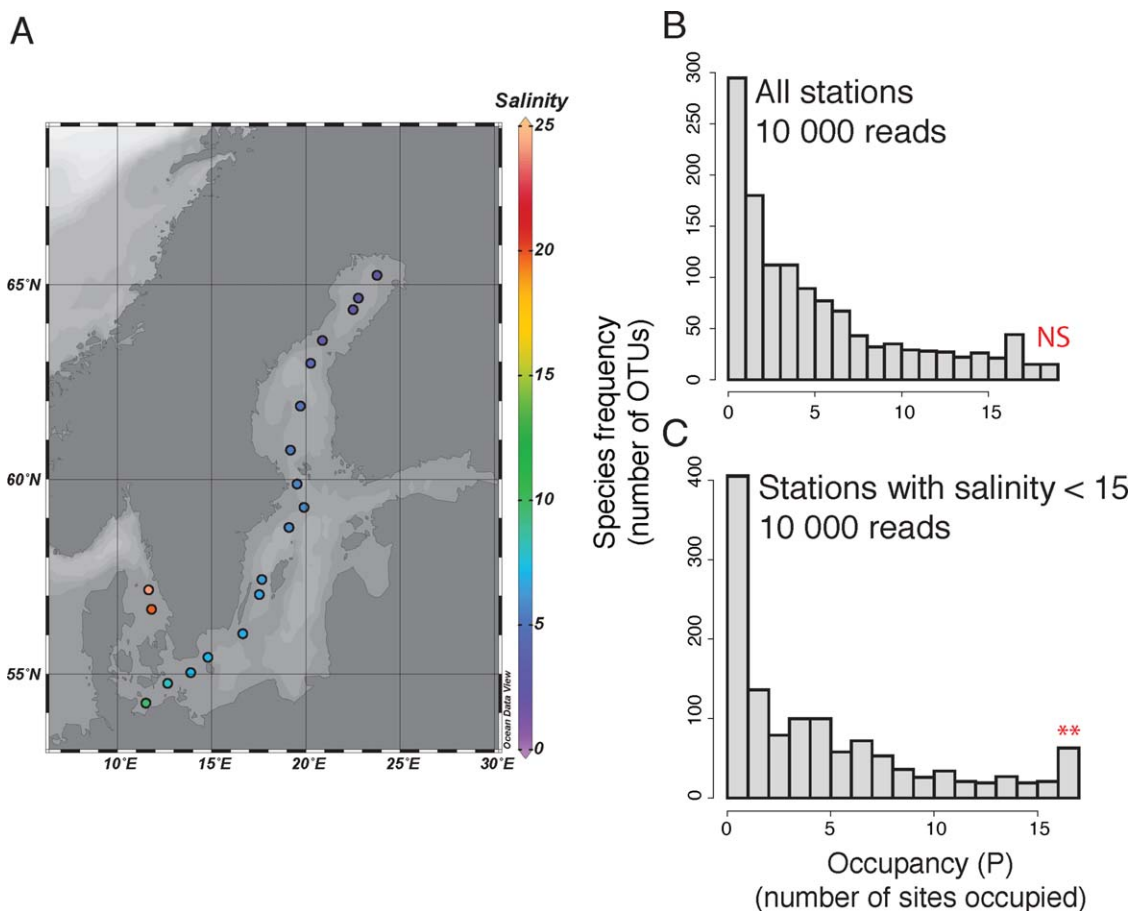


Fig. 4. A Baltic Sea transect covering the entire 2000 km salinity gradient (A), with the occupancy-frequency distribution of OTUs for all sites of the transect (B), and the occupancy-frequency distribution of OTUs for brackish water sites only (i.e. salinity < 15) (C). Asterisks denote significance levels ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$) for Mitchell-Olds & Shaw's test. NS denote non-significant bimodal patterns.

the core-satellite hypothesis, showing that positive feedback mechanisms regulate bacterioplankton population dynamics in the Baltic Sea Proper.

In the Baltic Sea Proper dataset we also observed particular cases of population dynamics characterized by high colonization or extinction rates. Firstly, several OTUs showed a pattern of being absent one month, and present at nearly all sites the following month, thus having a large colonization rate (Fig. 3; Supporting Information Table S4, $n = 11$). We denote this as 'microbial rain' pattern, as a reformulation of the concept 'propagule rain' coined by (Gotelli, 1991) for dispersal patterns of seed propagules from trees. The detected 'microbial rain' populations included *Pseudomonas* (PF_000313), AEGEAN-169 marine group (PF_000040) and FukuN18 freshwater group OTUs (PF_000183). A reciprocal pattern was observed for extinction rates, where some OTUs were present at all sites one month, followed by absence from all sites the following month - we term this 'microbial evanescence', indicating large extinction rates (Fig. 3;

Supporting Information Table S4, $n = 9$). 'Microbial evanescence' populations were exemplified by *Polaribacter* (PF_000009), *Rhodobacteraceae* (PF_000427) and *Fluviicola* OTUs (PF_000134). We suggest that 'microbial rain' and 'microbial evanescence' types of OTU distributions are likely to result from linkages between strong seasonal shifts in environmental conditions and niche differentiation.

OTU abundance distribution in a Baltic Sea-to-Kattegat transect

To extend the exploration of the applicability of the CSH and its consequences for interpreting bacterioplankton distribution patterns, we investigated complementary datasets that covered a range of marine environments and geographic distances. Thus, we first analysed a separate dataset of 16S rRNA genes from the Baltic Sea, spanning the entire 2000 km salinity gradient of this semi-enclosed brackish system from salinity 3 to 32 (Fig. 4A; 19 sites). No

significant bimodal occupancy-frequency pattern was found when including marine sites with salinity ~ 30 from the Kattegat Sea (Fig. 4B; Mitchell-Olds & Shaw's test, $P > 0.05$, $n = 19$). However, exclusion of the marine sites uncovered a significant bimodal pattern among the brackish sites (Fig. 4C; Mitchell-Olds & Shaw's test, $P < 0.05$, $n = 17$). This pattern was consistent when subsampling to 5000 sequences but not at 1000 sequences (Supporting Information Fig. S9A and B; Mitchell-Olds & Shaw's test, $P > 0.05$ and $P < 0.05$, 1000 and 5000 sequences, respectively, $n = 17$). The differentiation between marine and brackish water sites indicated that changes in environmental conditions, in this case likely the salinity shift (from 5.62 ± 2.06 to 22.00 ± 3.10 , Baltic Sea and Kattegat Sea respectively), influenced the metapopulation dynamics - separating bacterioplankton assemblages into distinct biomes in the Baltic Sea compared with the Kattegat Sea. Although the precise mechanisms remain unknown, salinity influences the distribution of bacterial populations and their functional potential in the Baltic Sea (Herlemann *et al.*, 2011; Dupont *et al.*, 2014; Hugerth *et al.*, 2015). Curiously, strong general shifts in marine biota occur across the Darss Sill at the border between Baltic Sea brackish-marine conditions, as shown in e.g. palaeoecological analyses of diatoms (Witkowski *et al.*, 2005). We conclude that bacterioplankton communities over major ranges of the Baltic Sea salinity gradient exhibit core and satellite population dynamics as observed in the regional analyses of Baltic Sea Proper data. Thus, species-abundance patterns among Baltic Sea bacterioplankton may provide clues to how bimodality is linked to specific biomes.

OTU abundance distribution in global ICoMM datasets

To further extend the analyses of bimodal occupancy-frequency distributions, we analysed distribution patterns in broad global biogeographic datasets provided by ICoMM (Amaral-Zettler *et al.*, 2010). In 16 of 31 oceanic regions bacterial communities displayed a significant bimodal pattern (Supporting Information Table S5). In Fig. 5A we exemplify some of the statistically significant bimodal patterns by the samples from the Arctic Chukchi and Beaufort Seas (ACB), Deep Arctic Ocean (DAO), New Zealand sediments (NZS) and open ocean off Cape Cod (SSD) (Mitchell-Olds & Shaw's test, $P < 0.05$, $n = 8-16$; Fig. 5A; Supporting Information Table S5). It is noteworthy that bacterial community composition in several vast oceanic regions (e.g. ACB and NZS, spanning thousands of kilometers) displayed bimodality. Still, in other regions, OTU occupancy-frequency distribution patterns were not significantly bimodal (Supporting Information Table S5), such as the northwest Mediterranean (BMO) and Sponge communities from the Great Barrier Reef (SPO) (Mitchell-Olds &

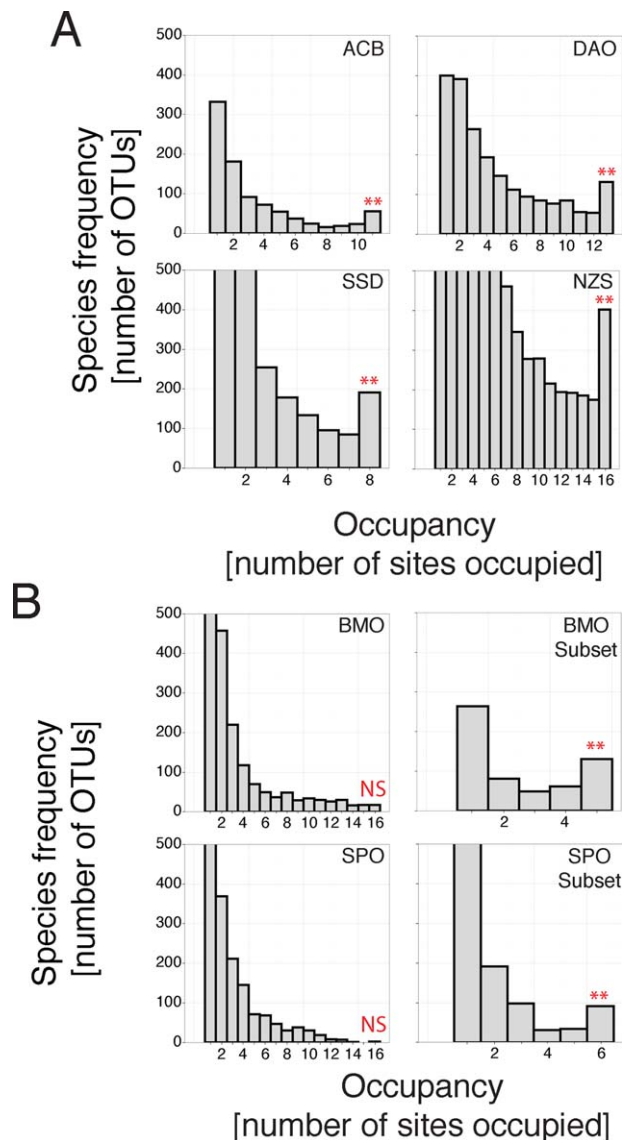


Fig. 5. Examples of occupancy-frequency distribution of OTUs from global ICoMM datasets displaying significant bimodal patterns (A), examples not showing significant bimodality unless applying the analysis on a subset of samples within datasets (B). Oceanic regions are abbreviated; ACB, Arctic Chukchi Beaufort; DAO, Deep Arctic Ocean; NZS, New Zealand Sediment; SSD, spatial scaling diversity (Cape Cod Atlantic Ocean); BMO, Blanes Bay Microbial Observatory (North West Mediterranean Sea); SPO; Sponge Bacteria (Great Barrier Reef). Please note that maximum 500 OTUs are plotted for the leftmost bar in each graph. Asterisks denote significance levels ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$) for Mitchell-Olds & Shaw's test. NS denote non-significant bimodal patterns. [Colour figure can be viewed at wileyonlinelibrary.com]

Shaw's test, $P > 0.05$, $n = 11-16$; Fig. 5B; Supporting Information Table S5). However, some of these datasets stand out by the heterogeneity of the samples they encompass. For example, the BMO transect comprised samples from different depths although it is well established that bacterioplankton community composition differs significantly

between depths in a range of ecosystems, including the northwest Mediterranean Sea (see e.g. Pommier *et al.*, 2010). Samples were therefore separated according to depth, and surface waters were analysed separately; we thus found a significant bimodal occupancy-frequency pattern for these surface samples (Mitchell-Olds & Shaw's test, $P = 0.003$, $n = 5$; Fig. 5B; Supporting Information Table S5; too few samples were available to test bimodality at other depths). Similarly, it is established that bacterial community composition differs considerably between sponge species (Schmitt *et al.*, 2012). Thus, when accounting for sponge species, a significant bimodal occupancy-frequency pattern emerged for OTUs associated with the sponge *Rhopaloeides odorabile* in the SPO dataset (Mitchell-Olds & Shaw's test, $P = 0.002$, $n = 6$; Fig. 5B; Supporting Information Table S5). Taken together, analyses of datasets from several oceanic regions provided strong support for a widespread prevalence of bimodality among marine prokaryotic assemblages.

From the above analyses we deduce that colonization of bacterioplankton populations could be regulated by physicochemical factors, structuring community composition into specific biomes. In this scenario, large differences in environmental conditions should lead to distinct patterns of bacterioplankton community structure, i.e. no bimodality. In contrast, within regions where dispersal is high compared with variations in local environmental conditions, bimodality should be detected. We therefore evaluated the relationship between bimodal occupancy-frequency patterns, beta-diversity, and environmental conditions. To analyse such potential dependencies, we separated ICoMM datasets into groups for datasets with significant or non-significant bimodal patterns, and calculated average Bray-Curtis distances for all. This revealed that datasets with significant bimodal patterns on average showed significantly lower Bray-Curtis dissimilarity values than datasets without bimodality (two-sample *t*-test, $P = 2.1 \times 10^{-5}$, $n = 31$) (Supporting Information Fig. S10A). Moreover, the core:satellite ratio (number of OTUs occupying all sites compared with a single site only) was significantly negatively correlated with average Bray-Curtis distance (linear regression, $P = 0.0007$, $R^2 = 0.30$, $n = 31$; Supporting Information Fig. S10B). A generalized linear model fit indicated a saturation where the core:satellite ratio leveled off around 0.05 (generalized linear model, inverse Gaussian distribution, $P = 0.0006$, $n = 31$; Supporting Information Fig. S10B). In terms of absolute changes in environmental conditions – i.e. analyses made on average Euclidean distances – temperature, salinity and depth had an overall tendency toward lower absolute differences for datasets with significant bimodal patterns (Supporting Information Fig. S10C–E). However, only temperature (Student's *t*-test, $P = 0.046$, $n = 20$) was significantly different between datasets with and without significant bimodal patterns. Notably,

absolute shifts in geographic distance were not significantly different between datasets with or without significantly bimodal patterns (Supporting Information Fig. S10F; two-sample *t*-test, $P = 0.292$, $n = 31$). In agreement, Galand *et al.*, (2009) and colleagues found that distance was less important compared with water mass for structuring community composition. These findings suggest that shifts in environmental conditions, and not spatial distance *per se*, are critical for regulating positive feedback mechanisms between local abundance and regional occupancy, thus affecting the occupancy-frequency pattern.

Although the bimodality patterns observed in the current study conforms with Hanski's metapopulation model, alternative hypotheses have been suggested to explain the shape of different occupancy-frequency distributions, including sampling effects (Gotelli, 1991; Tokeshi, 1992; McGeoch and Gaston, 2002). For example, the shape of the occupancy-frequency distribution depends on the size of the sample unit and the number of units sampled. To address potential issues of sampling effects, van Rensburg and colleagues (2000) subsampled the number of individuals for multiple spatial scales. Although spatial scale is an important factor in shaping bimodal occupancy-frequency distributions, this could not be separated from the rescue effect at the patch scale. In this paper, we performed two separate rarefaction tests (one for the Baltic Sea Proper and one for the Baltic Sea transect data, where we subsampled each of the datasets to different number of sequences to elucidate the impact of sampling effects. For the Baltic Sea Proper data, we observed that, logically, the number of core and satellite populations increased with increasing sequencing depth (Supporting Information Fig. S11). Most communities displayed a significant bimodal occupancy-frequency pattern at 1600 sequences and all communities were significantly bimodal at ≥ 3000 sequences. For the Baltic Sea transect data, subsampling to 1000 sequences resulted in a non-significant bimodal occupancy-frequency pattern, whereas subsampling to 5000 sequences resulted in a significant bimodal pattern. These findings suggest that sequencing at 3000–5000 sequences per sample is sufficient to capture bimodal occupancy-frequency distributions.

The pattern of linear increase in the number of core and satellite populations with increasing sequencing depth supports the CSH. Alternatively, the frequency distribution can be a direct consequence of different species-abundance patterns (e.g. log-normal [Preston, 1948] or logarithmic [Williams, 1950] rank-abundance distributions). Nevertheless, when we fitted different distribution models with rank-abundance curves for the Baltic Sea transect data, individual communities followed multiple distribution models, but most were characterized by lognormal distributions (Supporting Information Fig. S12A). However, the marine sites followed the Mandelbrot distribution. As different

distribution models were found both within and across these environmental gradients coupled with bimodal and unimodal patterns, respectively, we judge that sampling effects from such differences in rank abundance profiles were limited. Moreover, when we plotted a rank abundance curve for all sites collectively and compared it with rank abundance curves excluding marine sites, the curves were remarkably similar and both exhibited lognormal distributions (Supporting Information Fig. S12B and C). Another potential concern would be that sequencing errors result in inflated number of OTUs. This bias was, in part, addressed by excluding singletons in the analyses (although we did test the influence of including singletons in our analyses, see above). Further, the bioinformatics pipeline used in the present paper and developed by Quince *et al.*, (2011) aims at resolving sequencing errors. In addition, populations were clustered at 97% 16S rRNA gene identity, which result in a conservative estimation of populations. Future studies addressing the relative importance of artefactual versus biological factors in terms of shaping occupancy-frequency distributions could pinpoint robust mechanisms determining biogeographical distribution patterns among microbial assemblages (McGeoch and Gaston, 2002).

Still, in the datasets studied here, bimodality was found over a variety of temporal-geographical scales, indicating a pronounced robustness of the analyses. Alternatively, neutral mechanisms in combination with dispersal could result in a distribution pattern with locally abundant taxa also occurring frequently (Sloan *et al.*, 2006; Woodcock *et al.*, 2006; Economo and Keitt, 2008). However, colonization and extinction rates predicted by the CSH agreed well with our field data and point towards stochastic variation in rates of local extinction and/or colonization accounting for the observed bimodality. Taken together, our findings on the biogeography of microbial populations at distances ranging from <10 km to >10,000 km indicate that bimodal OTU distributions are an important recurring pattern in several marine ecosystems across the world's oceans, stressing the importance of regional dynamics.

Conclusions

Metapopulation models that include positive feedback mechanisms between local abundance and regional occupancy provide testable predictions of colonization and extinction rates (McGeoch and Gaston, 2002). Explorations of occupancy-frequency patterns have typically been done in communities with <100 species, whereas only few studies have investigated communities with >400 species (McGeoch and Gaston, 2002; Soininen and Heino, 2005). Our findings show that the core-satellite hypothesis can be applied to highly diverse prokaryotic assemblages in marine environments (>10,000 phylogenetically distinct populations) to characterize biogeographical patterns

providing insight into the importance of regional dynamics. Importantly, our results indicate that the core-satellite hypothesis provides a theoretical basis for how the distribution patterns of rare and abundant bacterioplankton populations in the local environment are regulated by stochastic variation in colonization and extinction rates. Given the pivotal role of marine bacteria in determining ocean-wide biogeochemical cycles of essential elements and fluxes of energy, the potential to identify drivers of the distribution of key populations could be central for characterizing the genomic potential of bacterioplankton for processes such as carbon, nitrogen and phosphorous cycling. We infer that analysis of occupancy-frequency patterns (e.g. the extent of bimodal distributions of core and satellite populations) can be a highly valuable approach allowing the definition of microbial biomes and for distinguishing where shifts in biomes occur across environmental gradients. Thus, analyses like those performed here open up for deepened mechanistic understanding of factors determining biogeographical distribution patterns among distinct microbial communities and their key populations. Such understanding could contribute to critically interpreting the connectivity between oceanic regions and how this relates to natural and anthropogenic changes in the environment.

Acknowledgements

We acknowledge Sabina Arnautovic and Emmelie Nilsson for their skillful technical assistance in the processing of samples. We thank Emil Fridolfsson at the Linnaeus University Centre for Ecology and Evolution in Microbial model Systems (EEMiS), Olof Lövgren at the Swedish University of Agricultural Sciences (SLU), and the crew of R/V Mimer for their outstanding efforts in this sampling campaign. We also thank Bengt Karlson at the Swedish Meteorological and Hydrological Institute (SMHI), for the sampling of the trans-Baltic bacterioplankton communities on R/V Poseidon. The computations were performed on resources provided by SNIC's UPPNEX project (CITE doi:10.1186/2047-217X-2-9) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project b2011200. This work was supported by the Swedish Research Council, the Swedish governmental strong research programme EcoChange, the Linnaeus University Centre for Ecology and Evolution in Microbial model Systems (EEMiS), and by the BONUS BLUEPRINT project, which has received funding from BONUS, the joint Baltic Sea research and development programme (Art 185), funded jointly from the European Union's Seventh Programme for research, technological development and demonstration and from the Swedish Research Council FORMAS to J.P. Y.H. was supported by a scholarship from the China Scholarship Council (CSC#201206950024). Further support was provided by the PLAN FISH project, financed by the Swedish Agency for Marine and Water Management (former Swedish Board of Fisheries) and the Swedish Environmental Protection Agency.

Author contributions

M.V.L., C.L. and J.P. conceived the study; M.V.L., M.C., C.L. and J.P. designed research; MB-F, ML, MC, and CL organized the fieldwork and performed the sampling; M.V.L., J.S., L.H. and Y.H. performed research; M.V.L., J.S. B.E., D.L., L.H., A.F.A., A.A. and J.P. analyzed data; M.V.L., J.S. B.E. and J.P. wrote the paper. All authors discussed the results and commented on the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

- Allison, S.D., and Martiny, J.B. (2008) Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105** (Suppl 1): 11512–11519.
- Alonso-Sáez, L., Diaz-Perez, L., and Morán, X.A. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol* **17**: 3766–3780.
- Amaral-Zettler, L., Artigas, L.F., Baross, J., Bharathi, P.A., L., Boetius, A., Chandramohan, D., *et al.* (2010) A Global census of marine microbes, In *Life in the World's Oceans: Diversity, Distribution, and Abundance*. McIntyre, A.D.(ed.), Wiley-Blackwell, Oxford, UK, pp. 221–245.
- Bertos-Fortis, M., Farnelid, H.M., Lindh, M.V., Casini, M., Andersson, A., Pinhassi, J., and Legrand, C. (2016) Unscrambling cyanobacteria community dynamics related to environmental factors. *Front Microbiol* **7**: 625.
- Blackburn, T.M., Cassey, P., and Gaston, K.J. (2006) Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *J Anim Ecol* **75**: 1426–1439.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *Am Nat* **124**: 255–279.
- Bunse, C., Bertos-Fortis, M., Sassenhagen, I., Sildever, S., Sjöqvist, C., Godhe, A., *et al.* (2016) Spatio-temporal interdependence of bacteria and phytoplankton during a Baltic Sea spring bloom. *Front Microbiol* **7**: 517.
- Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108**: 12776–12781.
- Díaz-Gil, C., Werner, M., Lövgren, O., Kaljuste, O., Grzy, A., Margonski, P., and Casini, M. (2014) Spatio-temporal composition and dynamics of zooplankton in the Kalmar Sound (Western Baltic Sea) in 2009–2010. *Bor Environ Res* **19**: 323–335.
- Dupont, C.L., Larsson, J., Yooseph, S., Ininbergs, K., Goll, J., Asplund-Samuelsson, J., *et al.* (2014) Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**: e89549.
- Economou, E.P., and Keitt, T.H. (2008) Species diversity in neutral metacommunities: a network approach. *Ecol Lett* **11**: 52–62.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.
- Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061–1063.
- Finlay, B.J., and Clarke, K.J. (1999) Ubiquitous dispersal of microbial species. *Nature* **400**: 828–828.
- Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* **106**: 22427–22432.
- Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M., and Lawton, J.H. (2000) Abundance–occupancy relationships. *J Appl Ecol* **37**: 39–59.
- Ghiglione, J.F., Galand, P.E., Pommier, T., Pedros-Alio, C., Maas, E.W., Bakker, K., *et al.* (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc Natl Acad Sci USA* **109**: 17633–17638.
- Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., and Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci USA* **110**: 4651–4655.
- Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrueck, L., Reeder, J., Temperton, B., *et al.* (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.
- Gómez-Consarnau, L., Lindh, M.V., Gasol, J.M., and Pinhassi, J. (2012) Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ Microbiol* **14**: 2361–2378.
- Gotelli, N.J. (1991) Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. *Am Nat* **138**: 768–776.
- Hanski, I. (1982) Dynamics of regional distribution - the core and satellite species hypothesis. *Oikos* **38**: 210–221.
- Hanski, I., and Gyllenberg, M. (1993) Two general metapopulation models and the core-satellite species hypothesis. *Am Nat* **142**: 17–41.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Harrison, S., and Cornell, H. (2008) Toward a better understanding of the regional causes of local community richness. *Ecol Lett* **11**: 969–979.
- He, Y., Caporaso, J.G., Jiang, X.T., Sheng, H.F., Huse, S.M., Rideout, J.R., *et al.* (2015) Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. *Microbiome* **3**: 20.
- Heino, J., and Virtanen, R. (2006) Relationships between distribution and abundance vary with spatial scale and ecological group in stream bryophytes. *Freshw Biol* **51**: 1879–1889.
- Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5**: 1571–1579.
- Hu, Y.O.O., Karlson, B., Charvet, S., and Andersson, A.F. (2016) Diversity of pico- to mesoplankton along the

- 2000 km salinity gradient of the Baltic Sea. *Front Microbiol* **7**: 679.
- Hugerth, L., Larsson, J., Alneberg, J., Lindh, M., Legrand, C., Pinhassi, J., and Andersson, A. (2015) Metagenome-assembled genomes uncover a global brackish microbiome. *Gen Biol* **16**: 279.
- Jari Oksanen, F.G., Blanchet, R., Kindt, P., Legendre, R.B., O'hara, G.L., Simpson, et al. (2010) vegan: Community Ecology Package. R package version 1.17-5. <https://cran.r-project.org/web/packages/vegan/index.html>.
- Lawton, J.H. (1999) Are there general laws in ecology? *Oikos* **84**: 177–192.
- Legrand, C., Fridolfsson, E., Bertos-Fortis, M., Lindehoff, E., Larsson, P., Pinhassi, J., and Andersson, A. (2015) Interannual variability of phyto-bacterioplankton biomass and production in coastal and offshore waters of the Baltic Sea. *Ambio* **44**: 427–438.
- Lennon, J.T., and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* **9**: 119–130.
- Levin, S.A. (1974) Dispersion and population interactions. *Am Nat* **108**: 207–228.
- Lindh, M.V., Sjöstedt, J., Andersson, A.F., Baltar, F., Hugerth, L.W., Lundin, D., et al. (2015) Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. *Environ Microbiol* **17**: 2459–2476.
- Lindström, E.S., and Langenheder, S. (2012) Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* **4**: 1–9.
- Logue, J.B., Mouquet, N., Peter, H., and Hillebrand, H. (2011) Empirical approaches to metacommunities: a review and comparison with theory. *Trends Ecol Evol* **26**: 482–491.
- Logue, J.B., Stedmon, C.A., Kellerman, A.M., Nielsen, N.J., Andersson, A.F., Laudon, H., et al. (2015) Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. *ISME J* **10**: 533–545.
- Magurran, A.E., and Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714–716.
- Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.
- McGeoch, M.A., and Gaston, K.J. (2002) Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biol Rev* **77**: 311–331.
- Mehranvar, L., and Jackson, D.A. (2001) History and taxonomy: their roles in the core-satellite hypothesis. *Oecologia* **127**: 131–142.
- Mitchell-Olds, T., and Shaw, R.G. (1987) Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**: 1149–1161.
- Morse, D.R., Stork, N.E., and Lawton, J.H. (1988) Species number, species abundance and body length relationships of arboreal beetles in Bornean lowland rain forest trees. *Ecol Entomol* **13**: 25–37.
- Pedros-Alio, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.
- Pommier, T., Neal, P., Gasol, J., Coll, M., Acinas, S., and Pedros-Alio, C. (2010) Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat Microb Ecol* **61**: 221–233.
- Pommier, T., Canbäck, B., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., et al. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.
- Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology* **29**: 254–283.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
- Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.
- R development core team. (2014) R: A language and environment for statistical computing. <https://cran.r-project.org/>.
- Rappé, M.S., Vergin, K., and Giovannoni, S.J. (2000) Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol Ecol* **33**: 219–232.
- Ricklefs, R.E. (2008) Disintegration of the ecological community. *Am Nat* **172**: 741–750.
- Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 578–587.
- Schlitzer, R. (2015) Ocean Data View, <http://www.odv.awi.de>.
- Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., et al. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.
- Sloan, W., Lunn, M., Woodcock, S., Head, I.M., Nee, S., and Curtis, T.P. (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., et al. (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Soininen, J., and Heino, J. (2005) Relationships between local population persistence, local abundance and regional occupancy of species: distribution patterns of diatoms in boreal streams. *J Biogeogr* **32**: 1971–1978.
- Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, L.A., and Sogin, M.L. (2013) Marine bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci USA* **110**: 2342–2347.
- Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., et al. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.
- Tokeshi, M. (1992) Dynamics of distribution in animal communities - theory and analysis. *Res Pop Ecol* **34**: 249–273.
- Unterseher, M., Jumpponen, A.R.I., Öpik, M., Tedersoo, L., Moora, M., Dormann, C.F., and Schnittler, M. (2011) Species abundance distributions and richness estimations in fungal metagenomics – lessons learned from community ecology. *Mol Ecol* **20**: 275–285.

- van Rensburg, B.J., McGeoch, M.A., Matthews, W., Chown, S.L., and van Jaarsveld, A.S. (2000) Testing generalities in the shape of patch occupancy frequency distributions. *Ecology* **81**: 3163–3177.
- Wardle, D.A., Bardgett, R.D., Callaway, R.M., and Van der Putten, W.H. (2011) Terrestrial ecosystem responses to species gains and losses. *Science* **332**: 1273–1277.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. <https://cran.r-project.org/web/packages/ggplot2/index.html>
- Williams, C.B. (1950) The application of the logarithmic series to the frequency of occurrence of plant species in quadrats. *J Ecol* **38**: 107–138.
- Witkowski, A., Broszinski, A., Bennike, O., Janczak-Kostecka, B., Bo Jensen, J., Lemke, W., et al. (2005) Darss Sill as a biological border in the fossil record of the Baltic Sea: evidence from diatoms. *Quat Int* **130**: 97–109.
- Woodcock, S., Curtis, T.P., Head, I.M., Lunn, M., and Sloan, W. (2006) Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol Lett* **9**: 805–812.
- Zinger, L., Boetius, A., and Ramette, A. (2014) Bacterial taxa-area and distance-decay relationships in marine environments. *Mol Ecol* **23**: 954–964.
- Östman, O., Drakare, S., Kritzberg, E.S., Langenheder, S., Logue, J.B., and Lindström, E.S. (2010) Regional invariance among microbial communities. *Ecol Lett* **13**: 118–127.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig. S1. Conceptual drawing of the models outlined. Levin's model (panel A), Hanski's core satellite hypothesis (CSH) (panel B), Gotelli's adaptations of propagule rain in Levin's model and Hanski's CSH (panel C) and (panel D), respectively. P is the fraction of occupied sites, C is colonization rate and E is extinction rate. The linear and quadratic extinction rates are calculated according to $dP/dt = CP(1 - P) - EP$ (Levin, 1974) and $dP/dt = CP(1 - P) - EP(1 - P)$ (Hanski, 1982), that characterize Levin's model (panel A) and the CSH (panel B), respectively. Levin's model (panel A) predicts a unimodal occupancy-frequency distribution (the number of species that occupy different number of sites; gray columns). Levin (1974) calculated the frequency of extinction across all sites in a given ecosystem where all patches are assumed to have an equal risk of extinction, resulting in common species with high occupancy having a higher risk to go extinct due to the high number of patches they occupy. In contrast, the CSH (panel B) predicts a bimodal pattern, and incorporates positive feedback mechanisms between local abundance and regional occupancy (Hanski, 1982). High rates of colonization in the CSH protect a population from extinction and are known as the rescue effect. In Gotelli's model (Gotelli, 1991), colonization is independent of P and based on a linear relationship between C and P , Levin's model and Hanski's CSH are thus modified according to $dP/dt = C(1 - P) - EP$ and $dP/dt = C(1 - P) - EP(1 - P)$, (panel C) and (panel D), respectively.

Fig. S2. Map showing the 80-km transect in the Baltic Sea Proper, adapted from (Díaz-Gil et al., 2014; Legrand et al., 2015). This transect was performed in the PLANFISH and

EcoChange frameworks, (see (Díaz-Gil et al., 2014; Legrand et al., 2015; Bertos-Fortis et al., 2016)).

Fig. S3. Rarefaction curves for each sample collected in the Baltic Sea Proper.

Fig. S4. Cluster analysis for comparing beta diversity calculated from Bray-Curtis distance estimation based on 454 pyrosequencing data using 97% 16S rRNA sequence similarity. Yellow points indicate distinct seasonal clusters. Station location is abbreviated with C, M and O denoting coastal, middle and open ocean sites, respectively.

Fig. S5. Baltic Sea Proper species rank-abundance curves for all samples.

Fig. S6. Occupancy-frequency distribution of OTUs delineated at 97%, 99% and 100% 16S rRNA gene sequence identity cut-off when clustering OTUs and including or excluding singletons. Samples are as for Figure 2, in our spatiotemporal study of the Baltic Sea Proper (Western Gotland basin and Kalmar sound).

Fig. S7. Boxplots of the number of core and satellite populations detected at different 16S rRNA gene sequence identity cut-off for clustering OTUs (97%, 99% and 100%), and including or excluding singletons. Samples are as for Figure 2, in our spatiotemporal study of the Baltic Sea Proper (Western Gotland basin and Kalmar sound).

Fig. S8. Sequence abundance and occupancy. Filled circles indicate relative sequence abundance (percentage of total sequences in each sample) for OTUs delineated at 97%, 99% and 100% 16S rRNA gene sequence identity, respectively. Samples are as for Figure 2, in our spatiotemporal study of the Baltic Sea Proper (Western Gotland basin and Kalmar sound).

Fig. S9. Occupancy-frequency distribution of OTUs from the Baltic Sea transect for brackish water sites only subsampled at 1000 and 5000 sequences, (A) and (B), respectively. Asterisks denote significance levels (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$) for Mitchell-Olds & Shaw's test. NS denote non-significant bimodal patterns.

Fig. S10. Boxplot of average Bray-Curtis distances from each ICoMM dataset and non-significant or significant bimodality (A), the relationship between average Bray-Curtis distance and the core:satellite ratio (number of OTUs occupying all compared to one site) (B), and the relationship between average Euclidean distances of environmental variables, geographic distance and non-significant or significant bimodality (C-F). Asterisks denote significance levels (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$) for Student's t-tests. Colored lines in (B) indicate a linear regression model and a generalized linear model (inverse Gaussian distribution), red and blue color, respectively.

Fig. S11. Subsampling of bacterioplankton communities found in each month showing the number of core (A), and satellite (B) populations plotted against increasing sequencing depths. OTU tables were rarefied to 100, 200, 400, 800, 1600 and 3000 sequences, respectively.

Fig. S12. Rank-abundance curves from the Baltic Sea transect data for individual sites (A), all sites collectively (B), and sites excluding marine conditions (C), (i.e. only sites with salinity < 15 were included).

Table S1. Non-linear least squares (NLS) analysis and statistics of observed colonization (C) and extinction (E) rates fitted with Hanski's model (Hanski, 1982) of $C = cP(1 - P)$ and

$E=eP(1-P)$. Asterisks denote significance levels (* $P<0.05$, ** $P<0.01$, *** $P<0.001$) for NLS tests.

Table S2. Non-linear least squares (NLS) analysis and statistics of observed extinction (E) rates fitted with Levin's model (Levin, 1974) where $E=eP$. Asterisks denote significance levels (* $P<0.05$, ** $P<0.01$, *** $P<0.001$) for NLS tests.

Table S3. Non-linear least squares (NLS) analysis and statistics of observed colonization (C) rates fitted with Gotelli's model of propagule rain (Gotelli, 1991) where $C=c(1-P)$. Asterisks denote significance levels (* $P<0.05$, ** $P<0.01$, *** $P<0.001$) for NLS tests.

Table S4. Classification of taxa into "core", "transient", "satellite", "microbial rain" and "microbial evanescence" populations. Times detected of each individual OTU average occupancy (P) in percentage and average local abundance. Populations with high average relative abundance $>1\%$ and high occupancy were classified as "core". Populations with high variance in occupancy and relative abundance were classified as "transient". Populations with low relative

abundance but high occupancy and little variance in relative abundance and occupancy were classified as "continuously rare". Populations with low average relative abundance and low occupancy were classified as "satellite". "Microbial rain" and "Microbial evanescence" populations have large variance in occupancy, but have at one or more occasions gone from being undetected at all sites to being present at all sites, and vice versa, respectively.

Table S5. Bimodality in global ICoMM data of marine microbes. For each ICoMM dataset, we have included the number of sites, total number of OTUs and the time frame within which the samples were collected, with p-values for Mitchell-Olds & Shaw's test of quadratic extremes. We further provide percentages of OTUs occupying 1 site, intermediary sites (i.e. average percentage of OTUs occupying 2 or more sites but not all) and all sites. The core:satellite ratio (i.e. the number of OTUs occupying all sites compared to 1 site) is also displayed. For comparison, average Bray-Curtis distances for each dataset is also provided. Asterisks denote significance levels (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).