Species-sorting and mass-transfer paradigms control managed natural metacommunities

Susanne Günther[1], Karoline Faust[2, 3], Joachim Schumann[1], Hauke Harms[1], Jeroen Raes[3], Susann Müller[1]

[1] Department of Environmental Microbiology, Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany

[2] VIB Center for the Biology of Disease, KU Leuven, O&N 4, Herestraat 49, 3000 Leuven, Belgium

[3] Department of Microbiology and Immunology, KU Leuven, O&N 4, Herestraat 49, 3000 Leuven, Belgium

Corresponding author: Susann Müller, Department of Environmental Microbiology, Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany. Telephone: +49 341 235-1318. Fax: +49 341 235-1351. E-mail: susann.mueller@ufz.de

Key words: species-sorting, mass effects, metacommunity, microbial community cytometry, individual based observation (IBO)

Running Head: Species-sorting and mass-transfer paradigms in metacommunities

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13402

Summary

A complex microbial system consisting of six different interconnected localities was thoroughly investigated at full scale for over a year. The metacommunity concept originating from macro-ecology was used to uncover mechanisms of community assembly by observing microbial interrelationships in and between the different localities via correlation and network analysis. The individual based observation approach (IBO) was applied using high-throughput microbial community cytometry in addition to next generation sequencing (NGS).

We found robust α-diversity values for each of the six localities and high β-diversity values despite directed connectivity between localities, classifying for endpoint assembly of organisms in each locality. Endpoint characteristics were based on subcommunities with high cell numbers whereas those with lower cell numbers were involved in dispersal. Perturbation caused abiotic parameters to alter local community assembly with especially the rare cells announcing community restructuration processes. The mass-effect paradigm as part of the metacommunity concept was identified by an increase in inter-locality biotic correlations under perturbation which, however, did not unbalance the predominant species-sorting paradigm in the studied full scale metacommunity. Data as generated in this study might contribute to the development of individual based models for controlling managed multi-species natural systems in future.

Introduction

Microbial communities are everywhere. They are the main drivers of biogeochemical cycles, interact intimately with higher organisms and are widely used in biotechnological production processes. However, the mechanisms of underlying community assembly and functioning are still unresolved. In ecology, manifold, often contradicting concepts were developed over decades to explain interactions between species from the viewpoint of community macro-ecology. Now, with the advent of NGS technologies researchers aim to test the validity of these concepts for microbial communities. In this study we evaluate well accepted ecological paradigms of community assembly and functioning using microbial community cytometry (MCC). The metacommunity of a full-scale wastewater treatment plant (WWTP) was studied at various spatial scales for a year with the aim to uncover principles of community assembly and progression with the perspective to use the knowledge for community control.

A metacommunity is defined as a set of local communities that are interlinked by dispersal or multiple interacting species (Wilson, 1992; Leibold *et al.*, 2004). The habitat containing and supporting multiple local communities is called the metacommunity region while the habitat holding a local community is

known by the term locality (Leibold *et al.*, 2004). The community structure in a certain locality is described by its α-diversity and the structural variation between localities by their β-diversity. The total diversity in a metacommunity region is called γ-diversity (Fukami, 2004a). Interactions between localities can be manifold. An example is dispersal describing movement of individuals between localities by immigration or emigration (Leibold *et al.*, 2004). Movement of individuals from one locality to another can result simply from different population sizes between localities or is supported by resource gradients. As to dispersal, several types are known, e. g. source-sink effects leading to the directed increase in a local community due to immigration, whereas colonization describes the local settlement of previously absent species (Leibold *et al.*, 2004). More specifically, niche opportunities are accepted by organisms if resource gradients correspond to their requirements (Shea and Chesson, 2002). Extinction events are also important for the establishment of local communities (Hubbel, 2001; Veresoglou *et al.*, 2015). Whereas the rescue effect helps to counteract extinction by ensuring constant immigration from a source community (Weinbauer and Rassoulzadegan, 2007), species can die or leave the locality due to adverse resource gradients or hostile community compositions (i.e. deterministic extinction) or for other reasons (i.e. stochastic extinction) (Pulliam, 1988). Leibold *et al.* (2004) emphasized that stochastic extinction is often associated with stochastic environmental perturbation affecting predominantly low abundant populations.

Generally, movement between communities can be restricted or unobstructed. Communities can live in closed systems such as pure cultures being the catalysts for biotechnological production processes. In nature communities live most often in open systems such as in the ocean where they are constantly shaped by dispersal and environmental fluctuation. Metacommunities of WWTPs, as used in this study, show characteristics of both. Reactors of different design and operational control reflect closed systems but they are also connected to each other in several ways thus allowing dispersal of individuals. The metacommunity of a WWTP has to fulfill different tasks. The most important one is the purification of wastewater to environmentally acceptable levels. Several localities are involved in the process such as the communities of the primary sludge (A), the primary clarifier (B), the aeration tank (C), and the excess sludge (D). A second task can be the production of methane by a biogas community (E), a process which is increasingly included in modern WWTPs. In Germany 1150 of the 2000 WWTPs with a capacity above 10,000 population equivalents employ a digester (Kolisch, 2011). In our study both tasks are interconnected because the fluids (F) of E are back-directed into C. Moreover, E is fed from B and D. Thus, the metacommunity of our study consists of six localities allowing easy dispersal between connected reactors.

A WWTP receives a source community (Shanks *et al.*, 2013) coming from the inflow which will disperse foreign individuals. Many of them will become extinct due to competitive exclusion (Louca and Doebeli, 2015), while some may be able to use niche opportunities (Shade *et al.*, 2012; Vuono *et al.*, 2016). Others may serve as a seed bank for the successive localities and their constant immigration may prevent extinction within local communities (i.e. rescue effect; Wells *et al.*, 2014; Low-Décarie *et al.*, 2015). Once established, physiologically and functionally very similar species can coexist for a long time. Contrarily, in closed systems without a vital rescue effect some species are likely to randomly outgrow others causing a decrease in diversity whereas in open system diversity will vary depending on fluctuating biomass dispersal and environmental gradients (Petraitis *et al.*, 1989; Hixon *et al.*, 2002). Open systems are therefore unsaturated, meaning that they are able to take up, maintain, and distribute invaders, which is contrary to saturated systems, which do not permit structural changes over long time periods (Cornel and Lawton, 1992; Fukami, 2004b). Such saturated communities that have reached an endpoint assembly configuration are called endpoint communities (Leibold *et al.*, 2004). The balance between those mechanisms of community assembly is unclear so far for WWTPs. Leibold *et al.* defined 4 different metacommunity paradigms including from i) the patch dynamics perspective where localities are identical and isolated, ii) the species-sorting perspective with abiotic factors strongly influencing community assembly in localities, iii) the mass-effect perspective where immigration and emigration determine the local community structure, and finally, iv) the neutral perspective, which declares all species similar and community assembly random. From the biotechnological point of view entirely neutral mechanisms would make a process control of the WWTP metacommunity demanding. Stochastic events would contradict any effort towards process control. On the other hand we would also not expect patch dynamics to dominate the WWTP metacommunity because we have a source community and all reactors are interconnected. Even the distinct tasks of water purification and methane production rely on interwoven local communities. Thus we postulate that WWTP metacommunities are assembled according to species-sorting and mass-effect paradigms. We further postulate that the detailed knowledge of their contribution would provide the chance to optimize and stabilize the processes either by abiotic or biotic parameter control.

Results

Bacterial communities sampled at 6 different localities of a full scale WWTP, i.e. A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, and the F: anaerobic digester centrate (for details see S1 and Figure 1a), were studied over a time frame of 389 days. During this time

at least one perturbation was disturbing the biogas production, whereas the water purification preceded undisturbed (S1). A total of 308 samples was collected and subjected to microbial community cytometry (MCC, Figure 2). In this study, MCC addresses single cells in a community by analyzing intrinsic cell features like cell size and density as well as DNA quantities per cell by fluorescent labeling via the A-T specific dye DAPI. MCC reveals metrics of community dynamics (Koch *et al.*, 2013a; Koch *et al.*, 2014) and of variations between localities (Zimmermann *et al.*, 2016) which can be automatically evaluated using the R-based tools flowCybar (www. bioconductor.org/packages/release/bioc/html/flowCyBar.html) and flowCHIC (http://www.bioconductor.org/packages/release/bioc/html/flowCHIC.html; Koch *et al.*, 2013b; Koch *et al.*, 2013c). Within the MCC histograms, 64 regions for community structure affiliation were chosen. This created a gate-template (i.e. gates comprising specific subcommunities, SC) based on at least 250.000 cells measured per sample. The distribution of cells between the gates in the histogram is given in percent. The 308 samples included 16 time points sampled for A, 60 for B, 72 for C, 49 for D, 71 for E, and 40 for F over the 389 day time period and were analyzed by MCC. A total of 19,712 fixed gate numbers were defined (noise and beads removed) which ranged between 0% (B_G40, day 378) and a maximum of 22.47% (B_G5, day 378, for individual SC amounts see S2) with an average fraction of cells per gate of 1.19%. Peripheral cells that did not cluster within SCs thus being outside the gate-template were defined as off-gate cells. The outcome is visualized as a flowCyBar heatmap representing the 64 regions per measured sample and a grey-color heatmap for the off-gate cells (Figure 2). Repeatability of MCC measurements was tested using triplicates of test samples (S3).

Community structure and diversity of the six localities

The six localities of our WWTP metacommunity were characterized by different grades of connectivity permitting dispersal. The primary clarifier (B) is connected with the primary sludge (A) and with the aeration tank (C), which in turn delivers the sludge for the excess sludge (D). Both sludges A and D are used for methane production in the digester (E). The fluid centrate (F) remaining after dewatering of the digestate from E is back-passed into C (Figure 1a). Cytometric α-diversity was calculated by counting SCs with average cell fractions >1.19% per day, based on data of the flowCyBar. It became apparent that the number of occupied SCs per locality and day established with minimum 18.31 (±4.53) and maximum 22.07 (± 3.06) SCs and the α-diversity Bray-Curtis dissimilarity values of 0.29 (±0.12) and 0.15 (±0.06), respectively, did not suggest highly varying compositions in each of the six localities (Table 1). Cytometric β-diversity values give information on dissimilarity between localities and were calculated by counting SCs that were not present in another locality chosen for comparison (unique SCs). This calculation was done for all locality pairs on a daily basis for all 308 samples. Low β-diversity values were estimated for

the localities C and D (5.96 SCs, Bray-Curtis 0.16) and a high one for e.g. B and D (21.95 SCs, Bray-Curtis 0.43; Table 1, S4). The γ-diversity was estimated as number of SCs in all localities on a given day with abundances over the 1.19% threshold which amounted to 86.45 SCs (±22.82) for the whole time frame. MultiCoLA analysis (Multivariate Cutoff Level Analysis, Gobet *et al.*, 2010) was implemented to test for the contribution of dominant and rare SCs to the α-diversity of the six localities. From the α-diversity value of each community (all SCs with cell abundances >1.19%) the least abundant SCs were removed step by step and the remaining SCs-distributions were tested for similarity and correlation with the initial SC assembly. All localities showed a positive correlation up to a cutoff-level of 50% of the initial SCs assembly and even at a cutoff-level of 65% the correlation coefficients were still high (Table 1, S5). The results suggest that the α-diversity is based mainly on the upper 50% of the most abundant SCs and that rare SCs do not contribute much to these values, a result that was also found elsewhere (cutoff-level 50%, Vuono *et al*., 2016). The fluctuations in community structure occurring both in and between localities are visualized by nonmetric multidimensional scaling (nMDS-Plot, Figure 1b, Bray-Curtis dissimilarities, stress: 0.12). The Bray-Curtis dissimilarity values were higher between the localities (β-diversity, Table 1) but also strongly suggest that some locality pairs were more similar as expressed by lower β-diversities than others. The α-diversity was highest in localities C, D, and E while β-diversity was highest among communities B/C and B/D.

Next to the structural similarities analyzed by MCC and successive ecological (α-, β-, and γ-diversities, MultiCoLA) and statistical (nMDS, Bray-Curtis) measures, correlations within and between the localities were calculated. To detect linked groups (i.e. abundance of cells in SCs) within the 308 samples can assist in deciphering relationships between SCs. Spearman's rank-order-correlation-coefficient, rhoS, was used to search for significant correlations within and between the localities ($P < 0.05$ after multiple-testing correction with Benjamini-Hochberg (Benjamini and Hochberg, 1995)). Correlating a total of 19,712 fixed gate numbers (i.e. SCs of all localities per 308 sampling days, S2) 147,456 correlations were found, 15,954 (10.82% of all correlations) of them being significant with 10,954 positive (mean 0.58) and 5,000 negative correlations (mean -0.5, Figure 1d, S6). The highest density of correlations was found within localities B (12%) and E (12.7%, average: 10.2%) while between localities maxima of not more than 3.53 and 2.46 % were found for the pairs C/D and C/E (average for all locality pairs: 1.28%).

We also used constructed association networks (Williams *et al.*, 2014; Faust *et al.*, 2015; for the 64 gates of the gate-template per locality) implemented for instance in the Cytoscape plugin CoNet (Faust *et al.*, 2012). MCC data were subjected to CoNet, where co-occurrence or mutual exclusion is tested using 5 measures: correlation (Spearman and Pearson), dissimilarity (Bray-Curtis and Kullback-Leibler) and similarity (mutual information). When at least two of the measures were significant ($P < 0.05$ after

multiple-testing correction with Benjamini-Hochberg (Benjamini and Hochberg, 1995)) and their scores are above a threshold selected to include 2000 initial edges for each measure, then the association is shown in form of lines (edges) between the SCs (nodes) within the network. The result of network analysis within the metacommunity is presented in Figure 1c. All six localities showed a high number of significant association, which were mostly positive (green, co-presence). Only a minor edge number was negative (red, mutual exclusion). While the number of nodes within each locality was similar with A: 54, B: 59, C: 61, D: 55, E: 58, and F: 51 the number of edges differed markedly: A: 118 (with 2 mutual exclusions), B: 130, C: 105, D: 85 (with 9 mutual exclusions), E: 90 (with 1 mutual exclusion), and F: 83. CoNet edges between the localities were very rare: only 3 out of 605 edges between localities (or 5 out of 338 nodes).

Thus, the robust structural α-diversity in the six localities is supported by correlational MCC data analysis which indicated strong *intra*-locality networks. In addition, the low density of correlations between localities supported the high values determined for structural β-diversity of the metacommunity. Thus, even under steadily changing inflow conditions of a municipal WWTP and the (restricted) dispersal between localities, the local communities within the metacommunity retained their characteristics over the whole time period of the study. Therefore, the respective cytometric communities qualify as endpoint communities.

Perturbation drives community behavior

WWTPs are prone to perturbations caused by varying loads of carbon and nutrients or other anthropogenic chemicals. During the time of the study there was one perturbation which affected the two functions of the WWTP differently. The water purification, relying mainly on the functionality of locality C (C, N, and P removal) but to a minor extent also on D, B and A, was efficient over the whole time period (see WWTP performance, S1). The biogas production process, however, relying mainly on the functionality of the digester E subsided severely for about 2 months starting at day 50 (Figure 2a, biogas production). Simultaneously, the FosTac values indicated a massive accumulation of acids (S1). FosTac is a measure for the amount of volatile organic acids compared to the buffering capacity of the system and has predictive value for anaerobic digester malfunctions (Rieger and Weiland, 2006). An irreversible digester breakdown was prevented by a series of different operational measures, including a complete stop of sludge feeding (day 66-73) and by stabilizing the pH of the digester with lime (to day 72) and 0.01 N sodium hydroxide (until day 135, S1). The intervention was concluded on day 135 when the biogas production recovered gradually. Subsequently, biogas production was stable again from day 248 on. Peak biogas production occurred at days 289 and 317 due to a two-fold sludge loading with 6.8

and 7.4 m^3h^{-1}, respectively, instead of the average of 3.6 m^3h^{-1} (black points in Figure 2a). Based on the functionality of E and the operational measures initiated we subdivided the process into 6 phases defined as *phase 1*: normal operation and biogas production (54.6 $m^3h^{-1}CH_4$, day 1-44); *phase 2*: acute perturbation with strong decrease in biogas productivity (7.2 $m^3h^{-1}CH_4$, pH stabilization and complete stop of sludge feeding, day 50-78); *phase 3*: receding perturbation with slowly increasing biogas production (17.3 $m^3h^{-1}CH_4$, pH stabilization and low amounts of sludge feeding, day 79-143); *phase 4*: start of reactor recovery with intermediate biogas production (39.4 $m^3h^{-1}CH_4$, day 149-240); *phase 5*: reactor recovery with increased biogas production (79.8 $m^3h^{-1}CH_4$, day 248-368); *phase 6*: normal operation and biogas production (52.3 $m^3h^{-1}CH_4$, day 371-389).

Endpoint communities such as those defined for the six localities (see above) are supposed to be robust assemblies not prone to easy collapse when experiencing perturbation. However, the degree of vulnerability can be different depending on factors such as dispersal, functional redundancy, or locality characteristics. The community biomass productivity is a further factor that seems to influence the vulnerability: higher the productivity was found to go along with higher resistance to perturbations (Kondoh, 2001; Chase, 2010). We see this effect also in our WWTP where the highly productive locality C (hydraulic retention time (HRT) of 0.412 d) was not affected in its function for water purification, whereas the low productive community E (HRT of 23.256 d) lost its function for biogas production almost completely for several weeks. To examine, why especially E reacted vulnerable with a near biogas production breakdown we determined the phylogenetic composition of the digester community E during acute and receding perturbation in *phases 2* and *3* (2 samples, respectively) and analyzed its member bacteria and archaea by pyrosequencing (Figure 2b). We found the phylogenetic diversity differing between acute and receding perturbation. During acute perturbation the bacterial community was dominated by the acid producing *Propionibacteriaceae* comprising 80.5 and 49.3 % of all sequences and by *Thermotogaceae* (5.7 and 20.8%) which are frequently found in digesters where they participate in the utilization of complex carbohydrates (Maus *et al.*, 2015; days 71 and 77, *phase 2*, Figure 2b). During receding perturbation, Clostridiales Incertae Sedis XIII, *Caldilineaceae,* and unclassified *Bacteroidetes* increased their abundance 4.9, 2.1, and 16 fold, respectively, whilst the previously abundant *Propionibacteriaceae* decreased 1.3 fold and *Thermotogaceae* nearly disappeared (from 20.8% to 0.4%; day 128, *phase 3*). The archaeal community was less diverse and comprised of almost only methane producers clustering in 4 main families: *Methanosaetaceae*, *Methanomicrobiaceae, Methanospirillaceae* and *Methanosarcinaceae* in a ratio of 1/0.74/0.23/0.16 (day 71, *phase 2*). In *phase 3* the *Methanomicrobiaceae* declined to only minor amounts (< 3.2%) and the *Methanosarcinaceae* nearly disappeared (< 0.14%). The *Methanospirillaceae* were relatively unaffected (< 10.5 % in *phase 2*, < 12.3 %

in *phase 3*) whilst the *Methanosaetaceae* outcompeted all other methanogens with 79.4 (± 2.04) % of all archaeal sequences. This expansion was not surprising because *Methanosaetaceae* are well known to be the dominant methanogens in WWTP digesters with low concentrations of volatile fatty acids (VFA) due to their high affinity for acetate which is the prevalent VFA in municipal WWTPs (Smith and Ingram-Smith, 2007; Demirel and Scherer, 2008). Thus, as the acute perturbation recedes and the VFA concentration normalizes (represented by decreasing FosTac values: 0.49 (d 71), 0.32 (d 77), 0.06 (d 121) and 0.08 (d 128)) *Methanosaetaceae* dominate over other acetoclastic and hydrogenotrophic methanogens. To support the phylogenetic data, the cytometric structural dissimilarities between the four sampled days were additionally compared by the flowCHIC approach (CHIC, Koch *et al.*, 2013c) in the nMDS plot (Figure 2c, Bray-Curtis, stress = 0.095). The dissimilarities between the samples in *phases 2* and *3* (0.59 and 2.19) suggest an increasing structural dissimilarity after perturbation, thus confirming the sequencing data of the bacterial community.

When analyzing the MCC data we made an interesting observation. The clustering efficiency (see methods) of the SCs in E was lower during perturbation (e.g. 73.01% at day 84) in comparison to the undisturbed community (e.g. 77.9% at day 17, Figure 2a, right side). The average clustering efficiency for all localities was 76.01%. Decrease in clustering efficiency caused increased cell numbers outside the gate template. Both events point to either incoming cells from preceding localities or an increased evenness between the SCs of a sample caused by the upcoming of rare taxa. We did not find very strong *inter*-locality correlations between SCs and off-gate cell numbers of precedent and ensuing localities in each of the six phases (four significant correlations by Spearman, Pearson, Kendall testing, Figure 3). Instead, *intra*-locality correlations were more transpiring (seven significant correlations) and a gradual increase in Sheldon-evenness was observed for locality E after the perturbation period. Such data endorse an upcoming of rare taxa (S6, 7).The sequencing data suggest that only the bacteria might be responsible for the effect instead of the emergence of *intra*-locality archaea. The increased evenness can be understood as part of the species-sorting paradigm where cells use the niche opportunity to establish themselves out of the core community. In macro-ecology such events are known as typical co-occurrences during perturbation (McCabe *et al.*, 2000; Elmquist *et al.*, 2003).

Paradigms vital in localities

Species-sorting is one of four paradigms of metacommunities (Leibold *et al.*, 2004). Our data, showing a high number of *intra*-locality correlations between SCs (Figure 1) and, in addition, between SCs and off-gate cells in a respective locality conform to this paradigm. Furthermore, abiotic factors can be anticipated to shape a locality, therefore *intra*- and *inter*-locality correlations with abiotic data would

support the prevalence of species-sorting. Mass-transfer is another paradigm which should be considered because a WWTP qualifies as an open system to some extent with high probability of dispersal between connected localities. Mass-transfer can be sought by correlating the cell abundance data in SCs *between* localities. To test both assumptions - the influence of the abiotic data and the occurrence of dispersal between localities - now biotic data along with 165 measured abiotic parameters (S2) were subjected, again, to the 3 association (Spearman, Pearson, Kendall), to the network (Cytoscape plugin CoNet) and, in addition, to bioenv (R package vegan) analyses. From the 3 correlation measures were only those values accepted where at least two measures had significances of $P < 0.05$ (after multiple-testing correction with Benjamini-Hochberg (Benjamini and Hochberg, 1995)). Results of the bioenv and network analyses were integrated in cases where significant correlations were found (marked in Figure 3 with asterisks). Data from each of the 6 phases were tested after creating subsets of almost equal length (11-13 sampling points per phase) to avoid any bias due to varying sample numbers.

The outcome is presented in Figure 3 for C and E (and for D and F in S8). The six phases differed from each other in two ways: the extent of community *intra- vs. inter*-locality correlations (Figure 3, pie charts) and the number of locality-abiotic correlations (Figure 3, bars, CoNet maps). We found nearly no abiotic factors shaping locality-specific communities in *phases 1-2* which changed in *phases 3-5* for the localities C and E. Especially in *phase 4* an increase in significant correlations between abiotic parameters and SCs was obvious, with mainly *intra*-locality correlations in C (6 out of 10) but only *inter*-locality correlations in E originating from precedent localities (8 of 8; bars, Figure 3). In *phase 6* the number ceased to almost zero similar to *phase 1*. Similar trends are shown by the CoNet analyses were the number of co-occurring lines between locality nodes and abiotic parameters followed the same course. The more the perturbation was retracted, the less abundant were the correlations between abiotic parameters and SCs in all localities (Figure 3, and S8). The strong impact of abiotic parameters among them total solids, nitrogen and phosphate species in *phase 4* and partly in *phase 5* points to a vital species-sorting paradigm in the localities C and E, although for E the abiotic shaping came mainly from C.

The mass-transfer paradigm, however, requires transfer of cells rather than spillover of abiotic factors between localities. We generally expected the mass-transfer responsible for the strong presence of endpoint communities which can only establish in connected localities if a rescue effect is acting to avoid competitive extinction. However, not all SCs contributed equally to the endpoint communities as was seen from the MultiCoLA analysis. Some of the SCs may thus subsidize to mass-transfer without substantial consequences for endpoint community characteristics. Therefore, we tested mass-transfer

between localities by counting biotic correlations using the following SC groups: complete gate-template (64 SCs), the higher abundant SCs (SCs >1.19%), and the dominant SCs (SCs dominant MultiCoLA). In addition, we tested for a possible contribution of the low and medium abundant SCs to mass-transfer for all but dominant SCs (SCs rare MultiCoLA) and rare SCs (SCs <1.19%, Table 2). *Intra*-community correlations between SCs were prevailing in all phases (Figure 1) and vital for all SC groups defined (Table 2). But we also counted strong *inter*-locality correlations especially in *phases 4* and *5* for C and E (Figure 3, pie charts, based on cell numbers in the 64 SC gate-template). A strong increase in *inter*-locality correlations for C and E in *phase 4* was verified for the higher abundant SCs (SC >1.19%), the rare SCs (SCs <1.19%), and the rare MultiCoLA SCs (Table 2). The importance of the rare SCs and the rare MultiCoLA SCs in the transfer process became obvious in an nMDS similarity analysis and by calculating the Gower distances (Gower, 1971) for these SCs which were closest between C and E in *phase 4* (Table 2), suggesting close *inter*-locality connections of those SCs in those two localities. The number of *inter*-locality correlations was still high in *phase 5* for the same selected SC groups. But different from all others, dominant MultiCoLA SCs did not show noteworthy *inter*-locality correlation numbers with other dominant MultiCoLA SCs, which again verifies their importance for stabilizing the endpoint communities. It could be assumed that rare SCs respond simultaneously to the altered environmental conditions during the restructuration process in *phases 4* and *5*, however, such an exclusive synchronized species sorting mechanism, which was not demonstrated for the dominant MultiCoLA SCs, seems less likely. An *inter*-locality transfer of organisms between the connected localities seems more plausible thus corroborating agreement with the mass-transfer paradigm.

After recovery of the biogas production, it remained still unclear whether species-sorting or mass-transfer were more important for the restructuration of the metacommunity after the perturbation. When visualizing the numbers of correlations found between all SCs (19,712 final gates) and all abiotic parameters (165), a trend to both *intra*- and *inter*-locality interactions became apparent (Figure 4, nMDS plot, Bray-Curtis, Stress: 0.02). In *phase 1* all localities showed mainly *intra*-community correlations (average distance between localities C, D and E: 3.6). As the disturbance progressed the type of correlations changed towards *inter*-locality and in case of C, D and E also towards locality-abiotic correlations. The nMDS comparison of all available biotic and abiotic data suggests agreement with the mass-transfer paradigm due to the shift to *inter*-locality correlations starting with phase 3, but the clear impact of the abiotics and the return to starting states classifies for species–sorting characteristics. The maximum distances for the localities in comparison to *phase 1* were found in *phase 4* for C (48) and *phase 5* for D (53) and E (46). In *phase 6* localities E and F were almost back at their initial positions in the

nMDS plot (distances to initial points: 9.95 (E) and 4.13 (F)), whereas C and D were still dominated by *inter*-locality correlations (distances to initial points: 45.6 (C) and 32 (D)).

The presence of locality-bound endpoint communities and the impact of abiotic parameters suggest species-sorting as the dominant paradigm in our WWTP metacommunity but the permanence of the endpoint communities and the transfer of cells between localities during perturbation qualified for mass-transfer.

Discussion

To conform to increasing demands of a circular economy, WWTPs need to make optimal use of their metacommunities (Shade *et al.*, 2012) e.g. by combining the water purification process with biogas production. Those operational functions will require organisms of different phylogenetic origin and metabolic traits in connected localities. WWTP microorganisms are well described for their diversity and metabolic functions (e.g. Ye and Zhang, 2013) and a high diversity seems to ensure functionally stable WWTP operation (Johnson *et al.*, 2014). However, the knowledge on full scale WWTP community-level properties and on responses to ecological changes is still scarce. In this study, we used the metacommunity concept (Logue *et al.*, 2011; Winegardner *et al.*, 2012; Leibold *et al.*, 2004) to investigate the assembly and dynamics of communities in connected localities and to uncover perturbation-associated symptoms in community dynamics.

Using single-cell analytics we found strong locality-specific α-diversities in our WWTP which suggests assembly of endpoint communities. The numbers of core SCs of the six endpoint communities (A-F) were low (determined by MultiCoLA (Gobet *et al.*, 2010)), and their respective Bray-Curtis dissimilarity values for given localities suggest high similarity (i.e. high stability) for the whole time period (389 days, Table 1). The findings match typical characteristics of endpoint communities which are saturated in their diversity and not prone to shifts (Morton and Law, 1996; Leibold *et al.*, 2004). In addition, it can be suggested that their stability is maintained by the steady supply of source communities which serve as seed bank for respective receiving localities (i.e. rescue effect (Wells *et al.*, 2014; Low-Décarie *et al.*, 2015)). Endpoint communities are not common in closed or open ecosystems. Closed systems have no access to source communities and frequently show community evolutions based on competitive exclusion or by experiencing deterministic extinction (Frentz *et al.*, 2015). Open systems are not seasoned by stable source communities and subjected to fluid environmental conditions (Fukami, 2004b). Our WWTP, in contrast, is neither closed nor fully open but operates differently connected localities. Exchange occur uni-directional between two or at maximum three localities (Figure 1a). The

only exception is B where the fluctuating inflow provides the source community. All successive communities are fed by source communities that already represent endpoint assembly. Thus, although the immigration schemes depended on either one flexible or several different endpoint source communities, all localities established their own individual endpoint community, thus creating a metacommunity of high β-diversity (Table 1).

Not only the cytometric α-diversity was different between the six endpoint communities but also the intrinsic biomass forming capacity. While B and C experienced high biomass productivities due to the steady supply of carbon and nutrients at short HRTs (HRT_B= 0.042 d and HRT_C=0.412 d), the community in E continued with low biomass productivity at long HRT_E (23.256 d). Short HRTs select for fast growing and competitively metabolizing organisms which has been suggested to be a precondition for an assumed high functional redundancy between organisms (Kondoh, 2001; Chase, 2010). The insurance hypothesis (Yachi and Loreau, 1999) applies to highly productive communities, which ensure their stability by the buffering effect of the redundancy of functions such as carbon degradation (Yin *et al.*, 2000) and phosphate accumulation (Nguyen *et al.*, 2011). The microbial cytometric fingerprints did not reveal functionally competitive dynamics in localities with short HRTs, however, organisms with similar growth characteristics may present themselves in similar cytometric distribution dynamics. This can only be verified by NGS technologies, which was not applied routinely to the 308 samples that were investigated in this study. Immigrating organisms inevitably came from source communities with higher productivity such as from B (HRT_B=0.042 d) to C (HRT_C= 0.412 d) and from C (via D) to E (HRT_E=23.256 d) or B (via A) to E. One might assume that the receiving slower productive communities would be responsive to the migration of fast growing source organisms by competitive extinction. But cytometric distribution dynamics clearly showed that none of these source communities established itself in a receiving community. Thus, it can be suggested that the localities are saturated and characterized by conditions that predefine biomass productivities – characteristics referred to as carrying capacity in macro-ecology (Arrow *et al.*, 1995). This concept defines a maximum possible population density limited by factors which are usually habitat size and available resources. Here, HRT seems to be such a factor. Interestingly, the different biomass forming capacity of the localities supported the assumption that immigration processes did not much influence the α- and β-diversities in our WWTP. Strong *intra*-locality and nearly absent *inter*-locality correlations between SCs (determined by CoNet) confirmed locality α-diversity in our WWTP (Figure 1b). Hence, mechanisms like colonization (Kirmer *et al.*, 2008), speciation (Arnegard *et al.*, 2014), or source-sink effects (Gravel *et al.*, 2010) did not come to pass at the first glance. Even competitive extinction is obviously prevented or limited by the rescue effect which ensures a steady supply of those source organisms which can take advantage of the resource gradients in the

receiving localities. It can be concluded that immigration in our WWTP is necessary but not as dominating and influential as is known for completely open systems. The data indicated that neither patch nor mechanisms underlying the neutral paradigm contributed to the assembly of endpoint communities. Thus, the huge differences between the locality-specific α-diversities point to species-sorting with efficient dispersal as prevailing paradigms in our WWTP (Winegardner *et al.*, 2012).

Nevertheless, the digester nearly broke down despite the presence of strong and stable endpoint communities in each locality for the time period of 389 days. Generally, permanent theories predict invasion-resistant endpoint communities in case of a given species pool (Morton and Law 1996 in Fukami 2004a). Thus, invasion-resistance of endpoint communities contradicts interaction between localities. This strong relation, however, seems to dissolve during perturbation. When biogas production ceased in *phases 2* and *3* still no correlations between any of the measured biotic or abiotic parameters emerged before the values for FosTac, pH, and methane indicated process failure (Figure 3 and S8). In addition, even the onset of operational measures to restore the methane production did not uncover immediate responses from the metacommunity. But eventually, an increase in off-gate cells during *phase 4* became conspicuous (Figure 2; S7). Off-gate cells are not part of the gate-template (64 SCs per sample) and can be easily analyzed and quantified cytometrically and thus qualify as key-marker for instabilities in otherwise stable endpoint communities. We could show by testing for Sheldon-evenness that they possibly emerged from within their locality qualifying as conditionally rare taxa in response to digester recovery.

In addition to off-gate cells we searched for further organisms that might emerge in gates outside of dominant SCs (i.e. SC dominant MultiCoLA, Table2) during phases 2-4 and thus might have contributed to or result from the perturbation. In fact, *inter*-locality connections were detected, though still at much lower numbers in comparison to the frequent *intra*-locality correlations. *Inter*-locality correlations revealed mainly rare MultiCoLA SCs as suspects in E and C (Table 2). The low Gower distance values for rare MultiCoLA SCs for localities C and E (*phase 4*, Table 2) suggest transfer of organisms either by the feeding of the digester from C via D to E or/and by back-directing sludge from E via F into C. Invasions (i.e. mass-transfer) occurring at high dispersal rates are observed to prevent resistance of endpoint communities (Lockwood in Fukami 2004a). However, our data suggest medium or low dispersal rates as was obvious by the types of SCs involved in the mass transfer and by the stable endpoint communities. Thus, if the perturbation was caused by invading organisms they must have been part of the rare SCs.

If we could prove *inter*-locality correlations for rare organisms (rare MultiCoLA SCs and off-gate cells) in *phase 4*, why couldn't we see any of those in *phases 2* and *3* where the perturbation transpired? To sustain or retrieve the water purification and biogas producing functions, operational stability of the WWTP was preserved by a restricted sludge transfer in our WWTP. Thus, a measureable mass-effect was confronted by disconnecting A and D from E in *phase 2* to prevent further transport of the unknown perturbation. In this situation, the upcoming of new bacterial phylotypes analyzed by NGS (Figure 2b) together with the decrease in pH and the increase in FosTac levels in E suggest locality-imminent competitive acidogenic organisms which would even further prevent recovery of methane production (Weiland, 2010; Koch *et al.*, 2013c). At the same time the archaea community became a monodominant community (Figure 2b). The shutdown of HRT_E might have further favored the species-sorting perspective for this locality.

Eventually, as intended, the obstruction of the sludge-transfer led to recovery of methane production in E. Since the digester was disconnected from localities A and B in *phases 2* and partly in *3* the inflow was directly transferred into C which was operated with the identical HRT_C as before. Thus, it is not surprising that significant correlations between abiotic parameters (especially total solids and phosphate species) and SCs within this locality increased to higher numbers, and also the SCs in E correlated but with abiotic parameters from the precedent localities after re-start of sludge feeding. Such events can either indicate upcoming of new or adaptation of indigenous species to new environments. Instead, the biotic *inter*-locality data indicate increased invasion of the lower abundant SCs (rare MultiCoLA SCs) but only during the restructuration process. Thus, according to our cytometric analyses the WWTP metacommunity acts foremost according to the species-sorting paradigm but mass-transfer by '*efficient* dispersal' also contributes to this paradigm. The term '*efficient* dispersal' was introduced by Winegardner *et al.* (2012), who strongly contradicted an exclusive classification of metacommunities according to the four paradigms, and this term seems to perfectly meet the mechanisms vital in the WWTP studied.

The paradigms species-sorting and mass-transfer demand for different control strategies of WWTPs, when perturbed or stable. Dispersal between localities should be within the boundaries of the locality-specific carrying capacity in WWTPs and in case of perturbation forestalled. Species-sorting effects should be supported by e.g. creating an environment that favors the survival of bottle-neck species. In our study these were the archaea which were promoted by a shift in pH to higher values in E, and zero HRT_E to prevent competitive exclusion. All abiotic parameters that support productive methanogens need to be kept in place (Graef and Andrews, 1974). Operators need to know if changes in the operational regime (e.g. in HRT) will help to mend imbalances in the different localities, in order to avoid

process malfunction or failure. Currently, schemes for a safer process control are still not available and individual based models (IBM) face problems related to the representation of complex systems at large scale (Zomorrodi and Segre, 2016). Although models are available (reviewed by Haanemeier *et al.*, 2015; Widder *et al.*, 2016) based on programs such as ModelSEED (http://seed-viewer.theseed.org/seedviewer.cgi?page=ModelView), Raven Tool (http://biomet-toolbox.org/index.php?page=downtools-raven), COMETS (http://www.bu.edu/segrelab/comets/), and recently, MetaFlux (component of the SRI's Pathway Tool software, Karp *et al.*, 2015) and iDynoMiCs 2 (Clegg *et al.*, 2014), those models are still being able to simulate the behavior of only a few organisms at a time (pers. Information P. Karp). Thus, individual based observation (IBO) will be the current alternative to collect data on such communities and possibly contribute in this way to further developments of IBMs.

Experimental procedures

Origin of the microbial communities

Wastewater samples were obtained from a full-scale WWTP in Eilenburg (Germany) that treats domestic and industrial (brewery and beverage industries) wastewater. Phosphate removal is achieved via enhanced biological phosphate removal (EBPR) with chemical precipitation as a backup. Samples were taken over a time period of 389 days at six localities throughout the WWTP. The ground plan, all technical parameters, and the WWTP operational scheme are described in S1.

Flow Cytometry

Cells were fixated directly after sampling with a ratio of 1:4 (volsample / volfixative) in 8% paraformaldehyde for 30 min at room temperature (RT). After centrifugation (3,200 x g; 20 min, 15°C) the cells were resuspended in 70% ethanol at a ratio of 1:8 (volsample / volfixative) and stored at -20°C until measurement. DNA staining of cells was performed as described (Koch *et al.*, 2013b), except the concentration of the DAPI staining solution used was 0.24 µM instead of 0.68 µM. Stained cells were stored in the dark at RT in sealed glass tubes for 24 hours and then measured flow cytometrically as logarithmically scaled 2Dplots according to DAPI fluorescence for DNA content and forward scatter for cell size related information. For every 2Dplot at least 250,000 cells were measured and beads were amended into every sample for adjustment. Samples were analyzed with a prototype of a CyFlow-Space (Partec, Görlitz, Germany) which is equipped with a 355 nm laser (Genesis CX355-150-STM-OPSLaser-Diode System, Coherent, CA, USA). Details of instrumental and analysis set up are documented in S3. Instrument calibration was done prior to measurement using 0.5 µm blue and 1 µm green fluorescent

beads (vendors in S3) as well as a DAPI stained bacterial standard with a known DNA-pattern. Analysis of measured samples and creation of the gate-template (see Koch *et al.*, 2013b; S3) was done using FlowJo V10 (FlowJo, LLC, Oregon USA). All raw data can be accessed at the FlowRepository (https://flowrepository.org/) under accession numbers FR-FCM-ZZQX and FR-FCM-ZZG6.

Analytical Methods

Parameters were analyzed at all sampling points according to the German DIN guidelines (S2). Some of the parameters were measured continuously; while others were measured intermittently, such as pH; electrical conductivity (EC); oxidation-reduction-potential (ORP); temperature; chemical-oxygen-demand (COD); biological-oxygen-demand within 5 days (BOD_5); FosTac; P_{-sup}, $PO_4^{3-}{}_{-sup}$, $P_2O_{5\text{-}sup}$ (based on sample supernatant (-sup)); P_{-tot}, $PO_4^{3-}{}_{-tot}$, $P_2O_{5\text{-}tot}$ (based on the heated, original sample containing cells (-tot)); NH_4-N, NH_4^+, NH_3, NO_3-N, NO_3^-, NO_2-N, NO_2^- (based on sample supernatants). Supernatants were obtained after centrifugation at 6,400 x g, 15 min, and 15°C.

Statistical Analysis

If not stated otherwise, all calculations and graphs were generated with R version 3.0.2 (R Development Core Team, 2013) using the packages flowCyBar (Schumann *et al.*, 2014), Hmisc (Harrell *et al.*, 2014), gplots (Warnes *et al.*, 2014), psych (Revelle, 2015) and vegan (Oksanen *et al.*, 2015). Statistical analysis was performed in 7 subsequent steps (S6 I). For additional statistical analysis CoNet (Faust *et al.*, 2012) and bioenv (vegan) were used (for details see S6).

Clustering efficiency

Clustering efficiency was calculated as the amount of cells that can be detected inside of the gate-template (S3) in relation to all measured cells (100%).

Pyrosequencing

DNA extraction, PCR and pyrosequencing reactions of 4 samples (days: 71, 77, 121, 128) from locality E were performed at LGC Genomics GmbH, Berlin, Germany. Details of sample preparation and sequencing are listed in S9. Raw sequences can be accessed at the sequence read archive (SRA; http://www.ncbi.nlm.nih.gov/sra) using the accession number: SRP074185

Supporting Information

	type	size (MB)
S1. General information on the WWTP sampled in this study containing the ground plan and WWTP characteristics.	pdf	0.24
S2. Contains a table of all measured parameters, the respective measurement method and their mean, minimum and maximum values.	pdf	0.40
S3. Details of MCC analysis, gate template and repeatability test.	pdf	1.54
S4. Cytometric α-, β- and γ-diversities and Bray-Curtis dissimilarities for all localities.	pdf	0.27
S5. Results of the multivariate cutoff level analysis (MultiCoLA).	pdf	0.08
S6. Detailed description of statistical approaches used.	pdf	0.16
S7. Sheldon's evenness for SCs and off-gate cells of each locality over time.	pdf	0.22
S8. Results of the correlation, network and bioenv analyses for localities D and F and overview of all results obtained by correlation, network and bioenv analyses.	pdf	0.92
S9. Details of the pyrosequencing methodology.	pdf	0.08

Acknowledgements

We acknowledge the support of the European Fond of Regional Development, the Sächsische Aufbaubank (PN 100072351) and the Helmholtz Association within the RP Renewable Energies. We also thank Christine Süring for technical assistance, the operators of the WWTP Eilenburg and the Abwasserzweckverband Mittlere Mulde.

Conflict of Interest

The authors declare that this research was conducted without any relationships that could be considered a potential conflict of interest.

Refererences

Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S. *et al.* (2014) Genetics of ecological divergence during speciation. Nature 511: 307–311.

Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C.S. *et al.* (1995) Economic growth, carrying capacity, and the environment. Science 268: 520-521.

Benjamini, Y., and Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57: 289-300.

Chase, J.M. (2010) Stochastic community assembly causes higher biodiversity in more productive environments. Science 328: 1388-1391.

Clegg, R.J., Dyson, R.J., and Kreft, J.-U. (2014) Repair rather than segregation of damage is the optimal unicellular aging strategy. BMC Biol 12: 52.

Cornell, H.V., and Lawton, J.H. (1992) Species interactions, local and regional processes, and limits to the richness of ecological communities - a theoretical perspective. J Anim Ecol 61: 1-12.

Demirel, B., and Scherer, P. (2008) The roles of acetotrophic and hydrogenotrophic methanogens during anaerobic conversion of biomass to methane: a review. Rev Environ Sci Biotechnol 7: 173-190.

Elmqvist, T., Folke, C., Nyström, M., Peterson, G., Bengtsson, J., Walker, B., and Norberg, J. (2003) Response diversity, ecosystem change, and resilience. Front Ecol Environ 1: 488-494.

Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J. *et al.* (2012) Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol 8: e1002606.

Frentz, Z., Kuehn, S., and Leibler, S. (2015) Strongly deterministic population dynamics in closed microbial communities. Phys Rev X 5: 041014.

Fukami, T. (2004a) Community assembly along a species pool gradient: implications for multiple-scale patterns of species diversity. Popul Ecol 46: 137-147.

Fukami, T. (2004b) Assembly history interacts with ecosystem size to influence species diversity. Ecology 85: 3234-3242.

Gobet, A., Quince, C., and Ramette, A. (2010) Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. Nucleic Acids Res 38: e155.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. Biometrics 27: 857-871.

Graef, S.P., and Andrews, J.F. (1974) Stability and control of anaerobic digestion. Journal WPCF 46: 667-682.

Gravel, D., Guichard, F., Loreau, M., and Mouquet, N. (2010) Source and sink dynamics in meta-ecosystems. Ecology 91: 2172-2184.

Hanemaaijer, M., Röling, W.F.M., Olivier, B.G., Khandelwal, R.A., Teusink, B., and Bruggeman F.J. (2015) Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. Front Microbiol 6: 213.

Harrell, F.E., Dupont, C. *et al*. (2014) Hmisc: Harrell Miscellaneous. R package version 3.14-3., [WWW document]. URL http://CRAN.R-project.org/package=Hmisc.

Hixon, M.A., Pacala, S.W., and Sandin, S.A. (2002) Population regulation: historical context and contemporary challenges of open vs closed systems. Ecology 83: 1490-1508.

Hubbell, S.P. (2001) The unified neutral theory of biodiversity and biogeography. Princeton, NJ, USA: Princeton University Press.

Johnson, D.R., Lee, T.K., Park, J., Fenner, K., and Helbling, D.E. (2014) The functional and taxonomic richness of wastewater treatment plant microbial communities are associated with each other and with ambient nitrogen and carbon availability. Environ Microbiol 17: 4851-4860.

Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R. *et al.* (2015) Pathway tools version 19.0 update: software for pathway/genome informatics and systems biology. Brief Bioinform doi: 10.1093/bib/bbv079.

Kirmer, A., Tischew, S., Ozinga, W.A., Von Lampe, M., Baasch, A. and Van Groenendael, J.M. (2008) Importance of regional species pools and functional traits in colonization processes: predicting re-colonization after large-scale destruction of ecosystems. J Appl Ecol 45: 1523-1530.

Koch, C., Fetzer, I., Schmidt, T., Harms, H., and Müller, S. (2013a) Monitoring functions in managed microbial systems by cytometric bar coding. Environ Sci Technol 47: 1753-1760.

Koch, C., Günther, S., Desta, A.F., Hübschmann, T., and Müller, S. (2013b) Cytometric fingerprinting for analysing microbial intra-community structure variation and identifying sub-community function. Nat Protoc 8: 190-202.

Koch, C., Fetzer, I., Harms, H. and Müller, S. (2013c) CHIC—an automated approach for the detection of dynamic variations in complex microbial communities. Cytometry 83A: 561-567.

Koch, C., Harms, H., and Müller, S. (2014) Dynamics in the microbial cytome – single cell analytics in natural systems. Curr Opin Biotech 27: 134-141.

Kolisch, G. (2011) Studie zur Aufbereitung und Einspeisung von Faulgas auf kommunalen Kläranlagen. WiW, Wupperverband, Stadtwerke Solingen GmbH.

Kondoh, M. (2001) Unifying the relationships of species richness to productivity and disturbance. Proc R Soc Lond B 268: 269-271.

Lardon, L.A., Merkey, B.V., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., and Smets, B.F. (2011) iDynoMiCS: next-generation individual-based modelling of biofilms. Environ Microbiol 13: 2416-2434.

Law, R., and Morton, R.D. (1996). Permanence and the assembly of ecological communities. Ecology 77: 762-775.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett 7: 601-613.

Logue, J.B., Mouquet, N., Peter, H., and Hillebrand, H. (2011) Metacommunity working group. Empirical approaches to metacommunities: a review and comparison with theory. Trends Ecol Evol 26: 482-491.

Louca, S. and Doebeli, M. (2015) Transient dynamics of competitive exclusion in microbial communities. Environ Microbiol doi:10.1111/1462-2920.13058.

Low-Décarie, E., Kolber, M., Homme, P., Lofano, A., Dumbrell, A., Gonzalez, A. *et al.* (2015) Community rescue in experimental metacommunities. Proc Natl Acad Sci 112: 14307-14312.

Maus, I., Cibis, K.G., Wibberg, D., Winkler, A., Stolze ,Y., König, H. *et al.* (2015) Complete genome sequence of the strain *Defluviitoga tunisiensis* L3, isolated from a thermophilic, production-scale biogas plant. J Biotechnol 203: 17-18.

McCabe, D.J., and Gotelli, N.J. (2000) Effects of disturbance frequency, intensity, and area on assemblages of stream invertebrates. Oecologia 124: 270-279.

Nguyen, H.T., Le, V.Q, Hansen, A.A., Nielsen, J.L., and Nielsen, P.H. (2011) High diversity and abundance of putative polyphosphate-accumulating *Tetrasphaera*-related bacteria in activated sludge systems. FEMS Microbiol Ecol 76: 256-267.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2015) vegan: Community Ecology Package. R package version 2.3-2., [WWW document]. URL http://CRAN.R-project.org/package=vegan.

Petraitis, P.S., Latham, R.E., and Niesenbaum, R.A. (1989) The maintenance of species diversity by disturbance. Q Rev Biol 64: 393-418.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. Am Nat 132: 652-661.

R Development Core Team. (2013) R: A language and environment for statistical computing, 3.0.1 edn. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2015) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, [WWW document]. URL http://CRAN.R-project.org/package=psych Version = 1.5.8.

Rieger, C., and Weiland, P. (2006) Prozessstörungen frühzeitig erkennen. Biogas J 4: 18-20.

Schumann, J., Koch, C., Günther, S., Fetzer, I., and Müller, S. (2014) flowCyBar: Analyze flow cytometric data. R package version 0.99.0.; [WWW document]. URL http://www.ufz.de/index.php?de=16773.

Shade, A., Peter, H., Allison, S.D., Baho, D.L., Bergam, M., Bürgmann, H. *et al.* (2012) Fundamentals of microbial community resistance and resilience. Front Microbiol 3: 417.

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., and McLellan, S.L. (2013) Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl Environ Microbiol 79: 2906-2913.

Shea, K., and Chesson, P. (2002) Community ecology theory as a framework for biological invasions. Trends Ecol Evol 17: 170-176.

Smith, K.S., and Ingram-Smith, C. (2007) Methanosaeta, the forgotten methanogen? Trends Microbiol 15: 150-155.

Veresoglou, S.D., Halley, J.M., and Rillig M.C. (2015) Extinction risk of soil biota. Nat Commun 6: 8862.

Vuono, D.C., Munakata-Marr, J., Spear, J.R., and Drewes, J.E. (2016) Disturbance opens recruitment sites for bacterial colonization in activated sludge. Environ Microbiol 18: 87-99.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., *et al.* (2014) gplots: Various R programming tools for plotting data. R package version 2.14.2., [WWW document]. URL http://CRAN.R-project.org/package=gplots

Weiland, P. (2010) Biogas production: current state and perspectives. Appl Microbiol Biotechnol 85: 849-860.

Weinbauer, M.G., and Rassoulzadegan, F. (2007) Extinction of microbes: Evidence and potential consequences. Aquat Microb Ecol 3: 205-215.

Wells, G.F., Wu, C.H., Piceno, Y.M., Eggleston, B., Brodie, E.L., DeSantis, T.Z. *et al.* (2014) Microbial biogeography across a full-scale wastewater treatment plant transect: evidence for immigration between coupled processes. Appl Microbiol Biot 98: 4723-4736.

Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T. *et al.* (2016) Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J doi: 10.1038/ismej.2016.45

Williams, R.J., Howe, A., and Hofmockel, K. (2014) Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Front Microbiol 5: 358.

Wilson, D.S. (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. Ecology 73: 1984-2000.

Winegardner, A.K., Jones, B.K., Ng, I.S., Siqueira, T., and Cottenie, K. (2012) The terminology of metacommunity ecology. Trends Ecol Evol 27: 253-254.

Yachi, S., and Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc Natl Acad Sci USA 96: 1463-1468

Ye, L., and Zhang, T. (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. Appl Microbiol Biot 97: 2681-2690.

Yin, B., Crowley, D., Sparovek, G., De Melo, W.J., and Borneman, J. (2000) Bacterial functional redundancy along a soil reclamation gradient. Appl Environ Microb 66: 4361-4365.

Zimmermann, J., Hübschmann, T., Schattenberg, F., Schumann, J., Durek, P., Friedrich, M. *et al.* (2016) High-resolution microbiota flow cytometry reveals dynamic colitis-associated changes in fecal bacterial composition. Eur J Immunol 46: 1300-1303.

Zomorrodi, A.R., and Segre, D. (2016) Synthetic ecology of microbes: Mathematical models and applications. J Mol Biol 428: 837-861.

Figures

Figure 1. a) Scheme of the WWTP with the 6 localities and the respective water and sludge flows (A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, F: anaerobic digester centrate). b) NMDS plot of structural community changes within the 6 localities (Bray-Curtis dissimilarity, stress: 0.12). Each day is shown as a dot and the dot size is representative of the day of sampling (day 1 is the smallest and day 389 the biggest). c) Associations found by network analysis (CoNet): The 6 localities and the associations between each of the 64 individual subcommunities therein over a time span of 389 days. Green lines stand for co-presence and red for mutual exclusion. Inter-locality associations are also shown as green lines. d) Associations found with Spearman's correlation coefficient after multiple testing correction of the respective *P* values and significance screening. Significant positive correlations are shown in dark red, negative in dark grey.

Figure 2. a) Left side: Cytometric barcode of the 6 localities (A-F, A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, F: anaerobic digester centrate) and respective biogas production over 389 days of analysis. There are 64 subcommunities (SCs) per sampling point, their abundances changes are color marked: red bars indicate increasing and blue bars decreasing SC abundances. White background indicates days where no sample could be taken from the respective sampling point. 6 phases were found for biogas production: phase 1 and 6 with normal biogas production (white background), phase 2 and 3 with acute and receding disturbance (dark grey background), phase 4 and 5 recovery (grey and slight grey). Biogas values and methane content are given as mean values for the respective periods. Black points indicate peak biogas production at days 289 and 317. Right side: Clustering efficiency of the cells within the 6 sampling points. Clustering efficiency was calculated as the amount of cells [%] that are clustered in defined gates in relation to all measured cells. Dark colors indicate a higher DNA and cell size related diversity (i.e. a higher amount of peripheral cells). b) Phylogenetic changes in bacterial and archaeal communities of the anaerobic digester (E) are shown during phase 2 (day 71 and 77) and phase 3 (day 121 and 128) with the number of sequences given in % of all found sequences. c) The MCC structural dissimilarity (determined by the CHIC approach (Koch *et al.*, 2013c) of the 4 days selected for sequencing is shown in the nMDS plot.

Figure 3. Pie charts: Intra- and inter-locality correlations within basins C and E during the 6 phases. The colors in the pie charts correspond to the colors defined in Figure 1. Bar charts: Significant correlations of the locality specific SCs with abiotic factors, which were confirmed by at least two of the tested correlation measures (Spearman, Pearson and Kendall). Values are given as percentages of the absolute possible number of correlations per parameter (64 for all SCs in one locality). The correlations marked

with '*' indicate associations also found either with all 3 correlation measures, or with network analysis or bioenv. Correlations marked with '**' indicate associations that were found with all 3 correlations measures and network analysis and bioenv. Results of the network analysis for all phases are shown on the right side. Co-presence interactions are shown in green and mutual exclusion in red, the localities are depicted as circles: B in red, C in light green, D in dark green, E in blue, F in purple and abiotics in grey. The higher the number of significant interactions (edges) is the bigger are the circles (number of nodes).

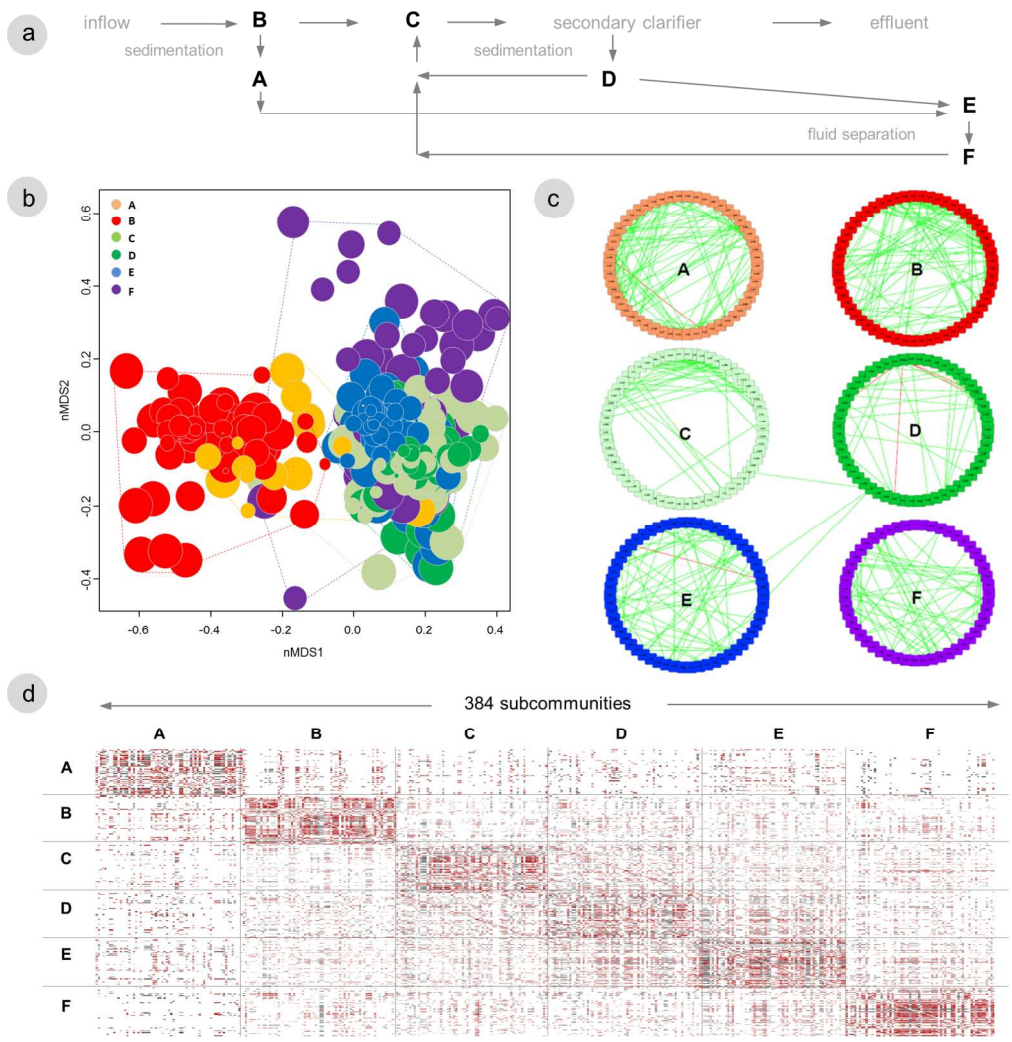
Figure 4. nMDS analysis (Bray-Curtis dissimilarities) of the mean association number obtained for the 6 phases in terms of intra- and inter-locality associations, locality-abiotic associations and abiotic-abiotic associations obtained by correlation, network and bioenv analysis. Phases are size-encoded (smallest points for phase 1 and biggest for phase 6). To highlight the clusters/groups of all samples (intra, inter) we used the function ordiellipse of the R package "vegan". Ordiellipse draws a polygon for dispersion ellipses around a certain group of samples using the standard deviation of the point scores. The (weighted) correlation defines the direction of the principal axis of the ellipse. To connect the points of each cluster/group (inter, intra) we used the function ordispider of the R package "vegan". Ordispider draws a "spider" diagram where each point is combined to the cluster/group center calculated as centroid (average of cluster/group points).

Table 1. Overview of results from different diversity measures used in this study. The complete tables can be found in S4 for cytometric diversities and Bray-Curtis dissimilarities and in S5 for MultiCoLA analysis. For MultiCoLA analyses only SCs with a mean abundance above 1.19 were regarded (A: 21 SCs, B: 18 SCs, C: 22 SCs, D: 23 SCs, E: 24 SCs, F: 18 SCs); the 50% cutoff level removes all SCs which have a summed up abundance below 50% of sum of all abundancies. SC: subcommunities, MultiCoLA: multivariate cutoff level analysis, A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, F: anaerobic digester centrate

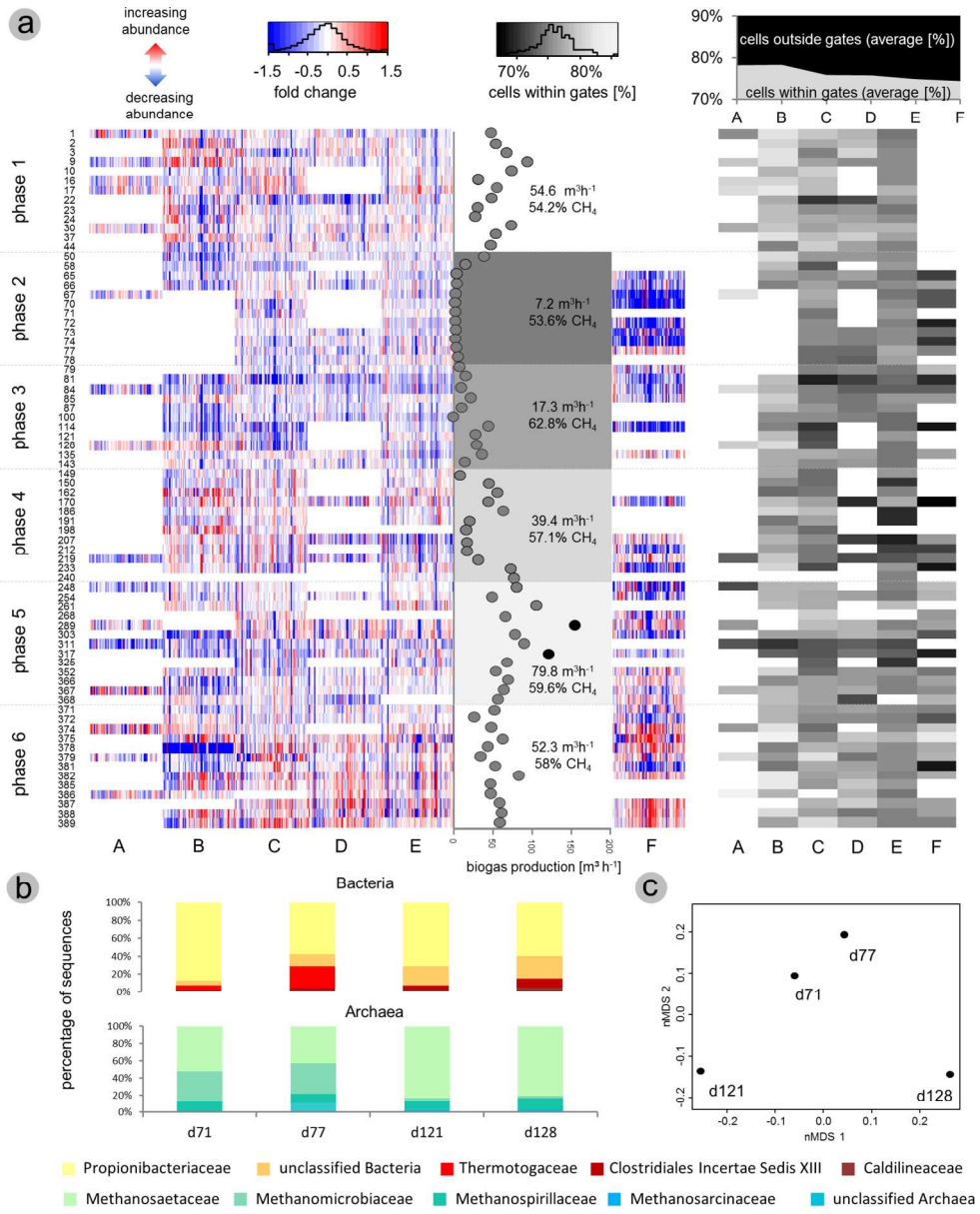
	α-diversity	α-diversity Bray-Curtis dissimilarity	β-diversity		β-diversity Bray-Curtis dissimilarity		Contribution abundant SC to α-diversity MultiCoLA			
	number SCs	dissimilarity	lowest number of unique SCs	highest number of unique SCs	lowest dissimilarity	highest dissimilarity	50% cutoff correlation coefficient	50% cutoff Procrustes	65% cutoff correlation coefficient	65% cutoff Procrustes
A	21.06 (±3.84)	0.24 (±0.1)	11.07 (±6.43, A/B)	14.53 (±4.63, A/C)	0.26 (A/B)	0.34 (A/D and F)	0.94	0.96	0.9	0.96
B	18.41 (±4.14)	0.19 (±0.07)	11.07 (±6.43, B/A)	21.95 (±5.56, B/D)	0.26 (B/A)	0.43 (B/D)	0.93	0.96	0.91	0.95
C	22.07 (±3.06)	0.16 (±0.06)	5.96 (±4.12, C/D)	20.56 (±6.24, C/B)	0.16 (C/D)	0.42 (C/B)	0.89	0.95	0.75	0.78
D	21.80 (±2.96)	0.15 (±0.06)	5.96 (±4.12, (D/C)	21.95 (±5.56, D/B)	0.16 (D/C)	0.43 (D/B)	0.91	0.89	0.8	0.58
E	21.77 (±2.02)	0.15 (±0.06)	9.56 (±3.1, E/D)	18.47 (±5.74, E/B)	0.2 (E/C and D)	0.39 (E/B)	0.92	0.97	0.9	0.97
F	17.90 (±5.27)	0.29 (±0.12)	11.34 (±3.77, F/E)	19.43 (±4.77, F/B)	0.26 (F/E)	0.42 (F/B)	0.94	0.95	0.8	0.75

Table 2. Different SC-sets of the metacommunity were tested for intra- and inter-community correlations and the resulting numbers subjected to measurements of the respective Gower distances. Numbers of within metacommunity correlations for localities C and E were counted, when at least 2 correlation measures (Spearman, Pearson, and, Kendall) were significant (P <0.05 after multiple testing correction). SCs dominant MultiCoLA: Multivariate cutoff level analysis (MultiCoLA) of gates with mean abundances >1.19 was performed, yielding 6 SCs for C and 7 SCs for E which were dominant at a cutoff-level of 50% (correlation values for Spearman and Procrustes above 0.8). SCs rare MultiCoLA: SCs that were excluded within SCs dominant MultiCoLA (C: 58 SCs, E: 57 SCs). SCs mean <1.19%: All SCs mean abundances within the metacommunity that are below 1.19% (C: 42 SCs, E: 40 SCs), where 1.19% is the average of the overall metacommunity SC abundance. SCs mean >1.19%: All SCs mean abundances within the metacommunity that are above 1.19% (C: 22 SCs, E: 24 SCs) where 1.19% is the average of the overall metacommunity SC abundance.

correlation numbers							
locality: community part	type	P1	P2	P3	P4	P5	P6
C: SCs rare MultiCoLA	intra	232	100	348	180	230	114
	inter	1	6	2	69	28	27
C: SCs dominant MultiCoLA	intra	4	2	0	10	2	2
	inter	1	0	0	2	0	0
C: SCs mean <1.19%	intra	23	27	15	104	12	14
	inter	0	0	1	18	8	4
C: SCs mean >1.19%	intra	170	56	278	110	154	84
	inter	1	6	3	39	21	13
E: SCs rare MultiCoLA	intra	298	283	204	236	124	286
	inter	5	7	2	68	36	26
E: SCs dominant MultiCoLA	intra	8	28	8	13	3	6
	inter	0	1	0	0	0	2
E: SCs mean <1.19%	intra	85	200	106	66	34	83
	inter	7	0	11	17	20	11
E: SCs mean >1.19%	intra	144	138	118	118	78	156
	inter	2	4	1	34	24	22
Gower distances C-E							
SCs rare MultiCoLA	distance	0.16	0.38	0.29	0.12	0.27	0.35
SCs mean <1.19%	distance	0.34	0.46	0.49	0.13	0.36	0.36



a) Scheme of the WWTP with the 6 localities and the respective water and sludge flows (A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, F: anaerobic digester centrate). b) NMDS plot of structural community changes within the 6 localities (Bray-Curtis dissimilarity, stress: 0.12). Each day is shown as a dot and the dot size is representative of the day of sampling (day 1 is the smallest and day 389 the biggest). c) Associations found by network analysis (CoNet): The 6 localities and the associations between each of the 64 individual subcommunities therein over a time span of 389 days. Green lines stand for co-presence and red for mutual exclusion. Inter-locality associations are also shown as green lines. d) Associations found with Spearman's correlation coefficient after multiple testing correction of the respective P values and significance screening. Significant positive correlations are shown in dark red, negative in dark grey.
178x183mm (300 x 300 DPI)



a) Left side: Cytometric barcode of the 6 localities (A-F, A: primary sludge, B: primary clarifier, C: aeration tank, D: excess sludge, E: anaerobic digester, F: anaerobic digester centrate) and respective biogas production over 389 days of analysis. There are 64 subcommunities (SCs) per sampling point, their abundances changes are color marked: red bars indicate increasing and blue bars decreasing SC abundances. White background indicates days where no sample could be taken from the respective sampling point. 6 phases were found for biogas production: phase 1 and 6 with normal biogas production (white background), phase 2 and 3 with acute and receding disturbance (dark grey background), phase 4 and 5 recovery (grey and slight grey). Biogas values and methane content are given as mean values for the respective periods. Black points indicate peak biogas production at days 289 and 317. Right side: Clustering efficiency of the cells within the 6 sampling points. Clustering efficiency was calculated as the amount of cells [%] that are clustered in defined gates in relation to all measured cells. Dark colors indicate a higher DNA and cell size related diversity (i.e. a higher amount of peripheral cells). b) Phylogenetic changes in bacterial and archaeal communities of the anaerobic digester (E) are shown during phase 2 (day 71 and 77)

and phase 3 (day 121 and 128) with the number of sequences given in % of all found sequences. c) The MCC structural dissimilarity (determined by the CHIC approach (Koch et al., 2013c) of the 4 days selected for sequencing is shown in the nMDS plot.
185x227mm (300 x 300 DPI)

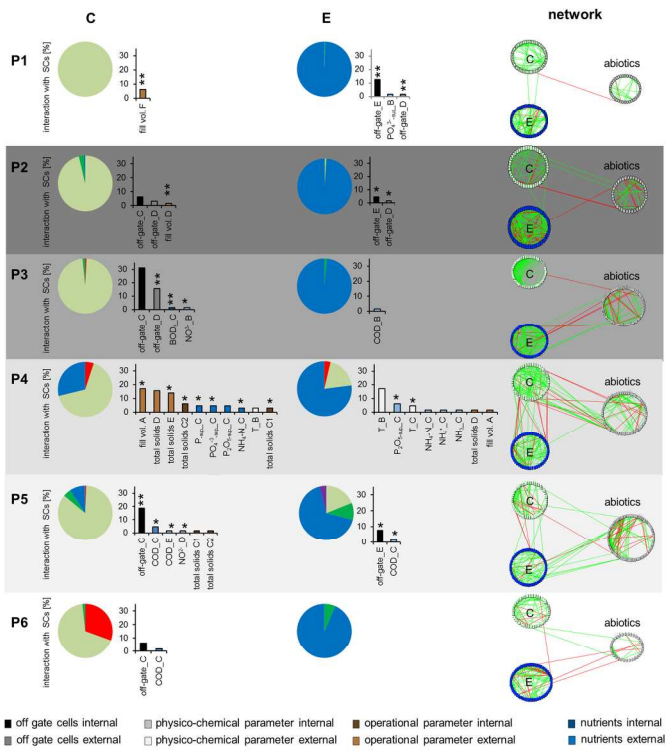
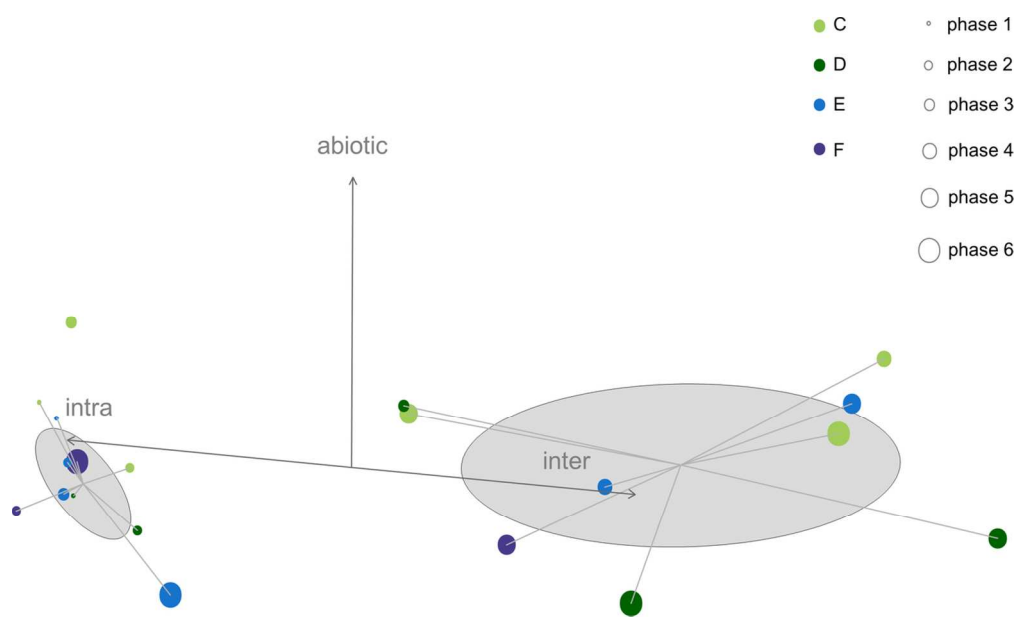


Figure 3. Pie charts: Intra- and inter-locality correlations within basins C and E during the 6 phases. The colors in the pie charts correspond to the colors defined in Figure 1. Bar charts: Significant correlations of the locality specific SCs with abiotic factors, which were confirmed by at least two of the tested correlation measures (Spearman, Pearson and Kendall). Values are given as percentages of the absolute possible number of correlations per parameter (64 for all SCs in one locality). The correlations marked with '*' indicate associations also found either with all 3 correlation measures, or with network analysis or bioenv. Correlations marked with '**' indicate associations that were found with all 3 correlations measures and network analysis and bioenv . Results of the network analysis for all phases are shown on the right side. Co-presence interactions are shown in green and mutual exclusion in red, the localities are depicted as circles: B in red, C in light green, D in dark green, E in blue, F in purple and abiotics in grey. The higher the number of significant interactions (edges) is the bigger are the circles (number of nodes).
190x142mm (300 x 300 DPI)



nMDS analysis (Bray-Curtis dissimilarities) of the mean association number obtained for the 6 phases in terms of intra- and inter-locality associations, locality-abiotic associations and abiotic-abiotic associations obtained by correlation, network and bioenv analysis. Phases are size-encoded (smallest points for phase 1 and biggest for phase 6). To highlight the clusters/groups of all samples (intra, inter) we used the function ordiellipse of the R package "vegan". Ordiellipse draws a polygon for dispersion ellipses around a certain group of samples using the standard deviation of the point scores. The (weighted) correlation defines the direction of the principal axis of the ellipse. To connect the points of each cluster/group (inter, intra) we used the function ordispider of the R package "vegan". Ordispider draws a "spider" diagram where each point is combined to the cluster/group center calculated as centroid (average of cluster/group points).
120x72mm (300 x 300 DPI)