VERSITA

Central European Journal of **Biology**

Population genetics: the next stop for microbial ecologists?

Commentary

Ramiro Logares

*Institute of Marine Sciences, ICM, CMIMA, CSIC,
ES-08003 Barcelona, Spain*

Abstract: Microbes play key roles in the functioning of the biosphere. Still, our knowledge about their total diversity is very limited. In particular, we lack a clear understanding of the evolutionary dynamics occurring within their populations (*i.e.* among members of the same biological species). Unlike animals and plants, microbes normally have huge population sizes, high reproductive rates and the potential for unrestricted dispersal. As a consequence, the knowledge of population genetics acquired from studying animals and plants cannot be applied without extensive testing to microbes. Next generation molecular tools, like High Throughput Sequencing (e.g. 454 and Illumina) coupled to Single Cell Genomics, now allow investigating microbial populations at a very fine scale. Such techniques have the potential to shed light on several ecological and evolutionary processes occurring within microbial populations that so far have remained hidden. Furthermore, they may facilitate the identification of microbial species. Eventually, we may find an answer to the question of whether microbes and multicellular organisms follow the same or different rules in their population diversification patterns.

Keywords: Diversity • Whole genome amplification • Pyrosequencing • Diversification

1. Microbial *vs*. Macrobial

Microbes represent a large fraction of the total biological diversity on Earth [1], having vital roles in the functioning of the biosphere [2]. Despite their importance, we still know very little about the total extent of microbial diversity [3], how such diversity is structured in space and time, as well as the ecological and physiological factors influencing microbial evolutionary diversification. In multicellular organisms, population size, reproductive rate and dispersal ability can play significant roles in the diversification of lineages [4]. For instance, small multicellular populations are more prone to evolve randomly by genetic drift than large populations, as well as to experience accidental extinctions [5]. Opposite to most multicellular organisms, microbes normally have huge population sizes, high reproductive rates and the potential ability for long-distance dispersal [6-8]. These characteristics could have a considerable influence on the diversification patterns of microbes, which could then differ from the patterns observed in multicellular organisms.

Microbes' potential for long-distance dispersal, along with their usually massive population sizes, have been hypothesized to reduce their chances of evolutionary diversification [9]. Specifically, long-distance dispersal and huge population sizes would be translated into unrestricted gene flow, preventing opportunities for allopatric speciation. However, the arrival of new molecular technologies started to change this view in the last 15 years, by showing that there is a much higher diversity of microbes than previously expected, and that there are endemic as well as cosmopolitan microbial species [10-13]. Thus, it is becoming increasingly apparent that the potential for high dispersal does not prevent microbial diversification and speciation. Most likely, the interplay of high dispersal rates and evolutionary diversification is more complex than what was initially considered. Even if dispersal is high, the environment may show enough heterogeneity to promote diversification through adaptive selection. Adaptive diversification may also be favored by the large genetic diversity that seems to be present in microbial populations [14-20]. Natural selection can act over this vast diversity and fine-

* E-mail: ramiro.logares@gmail.com

tune different microbial populations to their respective environments.

2. The emergence of microbial population genetics

Although during the recent years there has been a substantial advance in knowledge of the phylogenetic diversity and distributions of microbial groups [11-13], the field of microbial population genetics is still in its infancy [21]. To date, the number of genetic studies focusing on microbial populations is very limited in comparison with similar studies in animals and plants. However, the few available studies indicate a number of interesting patterns. For instance, in microeukaryotes, genetically distinct microbial populations can occur in marine environments despite the constant flow of currents [19,22-24]. In addition, different lakes can harbor genetically distinct microbial populations, and in some cases, distinct populations can even coexist within the same water body [20,25]. In bacteria, distinct putative populations of *Vibrio* were evidenced within the same community [26]. In all cases, population differentiation can progress into further diversification and eventually lead to speciation, as appears to have happened in some protists [27,28].

Despite the insights that the mentioned studies have provided, more extensive analyses including a larger number of strains from the same species are needed to comprehend the dynamics of microbial populations. Of particular interest is to assess key population parameters (e.g. population size) as well as the intrapopulation processes that can lead to restriction of gene flow and population divergence. Altogether, this information will be vital for understanding microbial diversification, as this phenomenon starts within populations [see 4,29]. Research in this direction should also unveil important parameters of microbial populations, such as the effective population size (*i.e.* the number of individuals in a population that pass genetic information to the next generation). The effective population size is an important variable that can determine different evolutionary dynamics; still, very few studies have calculated this value for microbes. In one of the few studies available, a large effective population size (7.5×10^7) was estimated for *Paramecium* [7]. These estimates were an order of magnitude larger than other calculations for macroorganisms. Other studies used simulations to investigate the relationships between population size and genetic diversity. For example, Mes [21] simulated population sizes between 10^3 and 10^7, and in most cases obtained large genetic diversity values. Still, the

author admits that natural populations are most likely several orders of magnitude larger [21], suggesting enormous amounts of genetic diversity within microbial species. Interestingly, there are also examples indicating the opposite pattern, as is the case in the marine picoeukaryote group MAST-4. It was estimated that the number of cells belonging to this successful lineage is about 10^{24} [30]. Despite such a huge number of cells, there is no evidence of more than ten species in MAST-4. Furthermore, the intraspecies diversification appears to be very low, as indicated by the 18S and ITS rDNA markers (Rodríguez-Martínez *et al.*, in revision). If we distribute this number of cells evenly between species, we would have population sizes of about 0.1×10^{24}, all of them showing a very limited neutral divergence. In bacteria, it has been found that LD12, the freshwater sister group of the superabundant marine SAR11, can also reach high abundances [$>1 \times 10^8$ cells/L; 31]. Still, LD12 presents low levels of evolutionary diversification across its vast geographical range, in particular when compared to SAR11 [32]. Altogether, the evolutionary scenario in MAST-4 and LD12 would be characterized by either a very recent evolutionary divergence followed by worldwide dispersal, or by a very strong environmental filtering that penalizes any deviation from an optimal cell design. More studies are needed to estimate the sizes of natural microbial populations and their amounts of genetic diversity.

3. Next generation technologies for investigating microbial population genetics

The limited number of population level studies in microbes is likely a consequence of the difficulties to a) define what organisms belong to the same species, and b), once a species is delineated, to obtain a variety of information from each individual cell. In the case of population genetics, once microbes belonging to the same putative species are found, the most common following issue is to obtain enough DNA from a single cell for multiple analyses. Opportunely, the relatively new technique called Single Cell Whole Genome Amplification (SCWGA) can generate enough genomic DNA template for multiple downstream analyses [33] by producing Single Amplified Genomes (SAGs). Among the variants for carrying out SCWGA, the Phi29 based Multiple Displacement Amplification (MDA) method appears to be the most popular for working with microbial cells [34,35]. MDA using Phi29 appears to be a good alternative for efficient whole genome amplification with relatively low error and bias [36,37]. SCWGA has

been combined with flow cytometry cell sorting, so cells belonging to specific sizes or even taxonomic groups can be identified and isolated and SAGs can be generated from them [35]. Still, contamination can be an important issue when generating SAGs, as Phi29 will amplify any DNA present in the reaction. Furthermore, there may be biases in the amplification favoring different phylogenetic groups. Nevertheless, SCWGA has the potential to provide template DNA from thousands of single cells from the same microbial species. For example, it is possible to sort *via* flow cytometry thousands of single cells in a few hours and generate SAGs that can later be screened with species-specific markers (e.g. ITS rDNA). Those SAGs of interest can later be separated and used for further population genetics analysis. As mentioned before, population genetics analyses normally require several markers, which can now be obtained from the SAGs using traditional PCR and sequencing or Next Generation Sequencing (NGS).

The use of SAGs for the study of microbial eukaryotes can also help to delineate species, a key issue in population genetics. Multiple molecular markers retrieved from each of several different SAGs may allow using the Phylogenetic Species Concept (PSC) [38], assuming that recombination does not occur within these markers. The concordance between the genealogies constructed with these markers points to the fixation of formerly polymorphic loci after genetic isolation, and these concordant branches would connect different species. Conflicts among these genealogies point to recombination between members of the same species. Thus, the transition from concordance to conflict delineates different phylogenetic species [38]. The PSC can be complemented with other criteria for delineating microbial species. For example, Coleman [39] indicated that the secondary structure of the nuclear ribosomal Internal Transcribed Spacer 2 (ITS-2) can predict mating, and therefore the Biological Species Concept [BSC; 40] could be applied. Altogether, sequencing of SAGs may allow coupling of the PSC with the BSC through the sequencing of multiple markers plus the ITS-2 from single microeukaryotic cells.

High Throughput Sequencing (HTS) techniques also open new research avenues for microbial population genetics. Next generation techniques like pyrosequencing [41], have increased considerably the sequencing speed and decreased its costs [42]. For instance, one 454 GS FLX Titanium/FLX +-pyrosequencing run can potentially generate 1 million sequences of about 400-700 base pairs respectively (http://www.my454.com/). Other sequencers such as Illumina are becoming increasingly popular, and can produce billions of sequences in a period of days (http://www.illumina.com/). Furthermore,

there is an array of other techniques which are available or being developed that can be useful in this field [43]. HTS can be applied to SAGs and then a large number of markers from different individuals can be obtained. After that, a subset of orthologous markers that are detected in different individuals can be separated and used for sequence analysis. This approach may still be challenging, but in my opinion, in the short term there will be an increasing number of population genetic studies coupling SCWGA with HTS. At the moment, studies with a similar philosophy are appearing. For example, Yoon *et al*. [44] have isolated single cells from evolutionarily closely related marine microeukaryotes (picobiliphytes) using fluorescence activated cell sorting and generated SAGs. Afterwards, 454 and Illumina shotgun sequencing were carried out using the SAGs as templates. 454 sequencing produced about 230,000 sequences/SAG while Illumina produced about 100 million sequences for all the analyzed SAGs. This study revealed complex interactions between microeukaryotes, bacteria and viruses, and it serves as a proof of concept for this approach.

The rapid increase in sequence data brought by HTS has also increased dramatically the need for powerful computers to analyze it. For example, one traditional run in an ABI 3730xl capillary sequencer would typically produce about 96 sequences of 650 bp each that could be analyzed on a desktop computer. Contrastingly, one Illumina HiSeq2000 can produce about 3000 million sequences (of 100 bp each) in a single run, and these data need to be analyzed in a computer cluster of at least 32 cores and 64 gigabytes of RAM. Many microbiologists are now becoming familiar with the use of high throughput computing and bioinformatics, but this still represents a challenge for the advancement of the field. Many universities and research institutions have now invested in high throughput computing. Furthermore, a large amount of computer power can now be purchased in the cloud (e.g. Amazon Web Services; http://aws.amazon.com/). Interestingly, we have probably reached the point where producing the data is considerably cheaper than its analysis and storage.

Given the remarkable advance in the power of sequencing technologies and computing capabilities [45], it would not be adventurous to think that in the next 20 years, population genetics will be carried out using the entire genome of microbial strains (especially prokaryotic). Today, around 6400 genomes (most of them prokaryotic) are publicly available, and several hundred more are in the waiting list (see http://www.genomesonline.org/). The combination of 454 plus Illumina sequencers makes it now possible

to sequence genomes with a high throughput, and everything indicates that throughput and sequence length will keep increasing during the next years, in particular with the advent of Third Generation Sequencing (e.g. Pacific Biosciences). Analyzing complete genomes from different individuals across populations would virtually be the ultimate population genetics data. Parallel studies using high throughput transcriptomics could also be combined in order to analyze differences in gene expression among microbial populations. For example, at the moment, several transcriptomes of diverse microeukaryote strains are being sequenced as part of an initiative of the Gordon and Betty Moore Foundation (http://marinemicroeukaryotes.org/). This initiative, as well as other similar studies, may start unveiling different or similar genetic expression profiles for diverse microbial populations.

4. Avenues for future research

Microbial population genetics is still a field in its infancy, but most likely it will grow dramatically during the next decades. This advance will be tightly coupled to the progress of molecular and bioinformatics tools as well as computer power. New studies will address important questions, for example, how microbes' huge population sizes affect their diversification, what orders of magnitude microbial populations sizes are, how often speciation occurs in microbial taxa, how microbial populations respond to selection, how much genetic variability microbial populations store, the role of genetic drift in microbial evolution, and whether microbes and macroorganisms present different or similar population diversification dynamics. The knowledge gained will probably bring a much deeper understanding of the factors driving microbial diversification and eventually speciation. Such knowledge may unveil differences and similarities between microbial and multicellular diversification patterns. Thus, during the next 20 years, we should obtain a more comprehensive view of the biological diversification processes across all biological kingdoms.

Acknowledgements

References

[1] Green J. L., Bohannan B.J.M., Spatial scaling of microbial biodiversity, Trends Ecol Evol, 2006, 21, 501-507

[2] Falkowski P.G., Fenchel T., Delong E.F., The microbial engines that drive Earth's biogeochemical cycles, Science, 2008, 320, 1034-1039

[3] Pedros-Alio C., Marine microbial diversity: can it be determined?, Trends Microbiol, 2006, 14, 257-263

[4] Futuyma D.J., Evolutionary Biology, Sinauer Associates, Inc., Sunderland, Massachusetts, 1998

[5] Coyne J.A., Orr H.A., Speciation, Sinauer Associates, Inc., Sunderland, Massachusetts, 2004

[6] Dolan J.R., An introduction to the biogeography of aquatic microbes, Aquat Microb Ecol, 2005, 41, 39-48

[7] Snoke M.S., Berendonk T.U., Barth D., Lynch M., Large global effective population sizes in Paramecium, Mol Biol Evol, 2006, 23, 2474-2479

[8] Newton R.J., Jones S.E., Helmus M.R., McMahon K.D., Phylogenetic ecology of the freshwater Actinobacteria acI Lineage, Appl Environ Microb, 2007, 73, 7169-7176

[9] Finlay B.J., Global dispersal of free-living microbial eukaryote species, Science, 2002, 296, 1061-1063

[10] Pace N.R., A molecular view of microbial diversity and the biosphere, Science, 1997, 276, 734-740

[11] Hughes Martiny J.B., Bohannan B.J.M., Brown J.H., Colwell R.K., Fuhrman J.A., Green J.L., et al., Microbial biogeography: putting microorganisms on the map, Nat Rev Microbiol, 2006, 4, 102-112

[12] Logares R., Does the global microbiota consist of a few cosmopolitan species?, Ecol Austral, 2006, 16, 85-90

[13] Lopez-Garcia P., Moreira D., Tracking microbial biodiversity through molecular and genomic ecology, Res Microbiol, 2008, 159, 67-73

[14] Medlin L.K., Lange M., Nothig E.M., Genetic diversity in the marine phytoplankton: a review and a consideration of Antarctic phytoplankton, Antarct Sci, 2000, 12, 325-333

[15] Rynearson T.A., Armbrust E.V., DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom Ditylum brightwellii, Limnol Oceanogr, 2000, 45, 1329-1340

[16] Shankle A.M., Mayali X., Franks P.J.S., Temporal patterns in population genetic diversity of Prorocentrum micans (Dinophyceae), J Phycol, 2004, 40, 239-247

[17] Evans K.M., Kuhn S.F., Hayes P.K., High levels of genetic diversity and low levels of genetic differentiation in North Sea Pseudo-nitzschia pungens (Bacillariophyceae) populations, J Phycol, 2005, 41, 506-514

[18] Iglesias-Rodriguez M.D., Schofield O.M., Batley J., Medlin L.K., Hayes P.K., Intraspecific genetic diversity in the marine coccolithophore Emiliania huxleyi (Prymnesiophyceae): The use of microsatellite analysis in marine phytoplankton population studies, J Phycol, 2006, 42, 526-536

[19] Nagai S., Lian C., Yamaguchi S., Hamaguchi M., Matsuyama Y., Itakura S., et al., Microsatellite markers reveal population genetic structure of the toxic dinoflagellate Alexandrium tamarense (Dinophyceae) in Japanese coastal waters, J Phycol, 2007, 43, 43-54

[20] Logares R., Boltovskoy A., Bensch S., Laybourn-Parry J., Rengefors K., Genetic Diversity Patterns in Five Protist Species Occurring in Lakes, Protist, 2009, 160, 301-317

[21] Mes T.H.M., Microbial diversity - insights from population genetics, Environ Microbiol, 2008, 10, 251-264

[22] Rynearson T.A., Armbrust E.V., Genetic differentiation among populations of the planktonic marine diatom Ditylum brightwellii (Bacillariophyceae), J Phycol, 2004, 40, 34-43

[23] Lilly E.L., Halanych K.M., Anderson D.M., Phylogeny, biogeography, and species boundaries within the Alexandrium minutum group, Harmful Algae, 2005, 4, 1004-1020

[24] Harnstrom K., Ellegaard M., Andersen T.J., Godhe A., Hundred years of genetic structure in a sediment revived diatom population, Proc Natl Acad Sci USA, 2011, 108, 4252-4257

[25] Wilson A.E., Sarnelle O., Neilan B.A., Salmon T.P., Gehringer M.M., Hay M.E., Genetic variation of the bloom-forming cyanobacterium Microcystis aeruginosa within and among lakes: Implications for harmful algal blooms, Appl Environ Microb, 2005, 71, 6126-6133

[26] Hunt D.E., David L.A., Gevers D., Preheim S.P., Alm E.J., Polz M.F., Resource partitioning and sympatric differentiation among closely related bacterioplankton, Science, 2008, 320, 1081-1085

[27] Logares R., Rengefors K., Kremp A., Shalchian-Tabrizi K., Boltovskoy A., Tengs T., et al., Phenotypically different microalgal morphospecies with identical ribosomal DNA: A case of rapid adaptive evolution?, Microb Ecol, 2007, 53, 549-561

[28] Logares R., Daugbjerg N., Boltovskoy A., Kremp A., Laybourn-Parry J., Rengefors K., Recent evolutionary diversification of a protist lineage, Environ Microbiol, 2008, 10, 1231–1243

[29] Avise J.C., Phylogeography: the history and formation of species, Harvard University Press, Cambridge, Massachusetts, 2000

[30] Massana R., Terrado R., Forn I., Lovejoy C., Pedros-Alio C., Distribution and abundance of uncultured heterotrophic flagellates in the world oceans, Environ Microbiol, 2006, 8, 1515-1522

[31] Salcher M. M., Pernthaler J., Posch T., Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12), ISME J, 2011, 5, 1242-1252

[32] Logares R., Brate J., Heinrich F., Shalchian-Tabrizi K., Bertilsson S., Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11), Mol Biol Evol, 2010, 27, 347-357

[33] Binga E.K., Lasken R.S., Neufeld J.D., Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology, ISME J, 2008, 2, 233-241

[34] Marcy Y., Ishoey T., Lasken R.S., Stockwell T.B., Walenz B.P., Halpern A.L., et al., Nanoliter reactors improve multiple displacement amplification of genomes from single cells, PLOS GENET, 2007, 3, e155

[35] Stepanauskas R., Sieracki M.E., Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time., Proc Natl Acad Sci USA, 2007, 104, 9052-9057

[36] Pinard R., de Winter A., Sarkis G.J., Gerstein M.B., Tartaro K.R., Plant R.N., et al., Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing, Bmc Genomics, 2006, 7, 216

[37] Spits C., Le Caignec C., De Rycke M., Van Haute L., Van Steirteghem A., Liebaers I., et al., Optimization and evaluation of single-cell whole-genome multiple displacement amplification, Hum Mutat, 2006, 27, 496-503

[38] Taylor J.W., Jacobson D.J., Kroken S., Kasuga T., Geiser D.M., Hibbett D.S., et al., Phylogenetic species recognition and species concepts in fungi, Fungal Genet Biol, 2000, 31, 21-32

[39] Coleman A.W., Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide, Mol Phylogenet Evol, 2009, 50, 197-203

[40] Mayr E., Systematics and the Origin of Species, Columbia University Press, New York, 1942

[41] Ronaghi M., Uhlen M., Nyren P., A sequencing method based on real-time pyrophosphate, Science, 1998, 281, 363-365

[42] Trombetti G.A., Bonnal R.J.P., Rizzi E., De Bellis G., Milanesi L., Data handling strategies for high throughput pyrosequencers, Bmc Bioinformatics, 2007, 8, S1-22

[43] Glenn T.C., Field guide to next-generation DNA sequencers, Mol Ecol Resour, 2011, 11, 759-769

[44] Yoon H.S., Price D.C., Stepanauskas R., Rajah V.D., Sieraki M.E., Wilson W.H., et al., Single-Cell genomics reveals organismal interactions in uncultivated marine protists, Science, 2011, 332, 714-717

[45] Dupont C.L., Chappell D., Logares R., Vila-Costa M., A hitchhiker's guide to the new molecular toolbox for ecologists, In: P. Kemp (Ed.), Eco-DAS VIII Symposium Proceedings (11-16 October 2008, Hawaii, USA), ASLO, 2010, 17-29