Polar Biosci., **20**, 103–112, 2006

A novel bioinformatics tool for phylogenetic classification of genomic sequence fragments derived from mixed genomes of uncultured environmental microbes

Takashi Abe[1*], Hideaki Sugawara[1], Shigehiko Kanaya[2] and Toshimichi Ikemura[3,4*]

[1]*Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, and The Graduate University for Advanced Studies (Sokendai), Mishima, Shizuoka 411-8540*
[2]*Department of Bioinformatics and Genomes, Graduate School of Information Science, Nara Institute of Science and Technology, Takayama, Ikoma, Nara 630-0101*
[3]*The Graduate University for Advanced Studies (Sokendai), Hayama Center for Advanced Research, Hayama-cho, Kanagawa 240-0193*
[4]*Nagahama Institute of Bio-Science and Technology, Nagahama, Shiga 526-0829*
[*]*Corresponding author. E-mail: takaabe@genes.nig.ac.jp or ikemura_toshimichi@soken.ac.jp*

(Received March 30, 2006; Accepted June 19, 2006)

Abstract: A Self-Organizing Map (SOM) is an effective tool for clustering and visualizing high-dimensional complex data on a two-dimensional map. We modified the conventional SOM to genome informatics, making the learning process and resulting map independent of the order of data input, and developed a novel bioinformatics tool for phylogenetic classification of sequence fragments obtained from pooled genome samples of microorganisms in environmental samples allowing visualization of microbial diversity and the relative abundance of microorganisms on a map. First we constructed SOMs of tri- and tetranucleotide frequencies from a total of 3.3-Gb of sequences derived using 113 prokaryotic and 13 eukaryotic genomes, for which complete genome sequences are available. SOMs classified the 330000 10-kb sequences from these genomes mainly according to species without information on the species. Importantly, classification was possible without orthologous sequence sets and thus was useful for studies of novel sequences from poorly characterized species such as those living only under extreme conditions and which have attracted wide scientific and industrial attention. Using the SOM method, sequences that were derived from a single genome but cloned independently in a metagenome library could be reassociated *in silico*. The usefulness of SOMs in metagenome studies was also discussed.

key words: Self-Organizing Map, uncultured microorganism, phylogenetic classification, environmental microorganism, metagenome

Introduction

Environmental microorganisms, especially those living under extreme conditions, cannot be cultured easily under laboratory conditions. Genomes of uncultured organisms have remained mostly uncharacterized and are thought to contain a wide range of novel genes of scientific and industrial interest. Metagenomic approaches, which are analyses of mixed populations of uncultured microbes, have been developed to identify novel and industrially useful genes and to study microbial diversity in a wide variety of environments

(Amann *et al*., 1995; Hugenholtz and Pace, 1996; DeLong, 2002). In the metagenomic approach, genomic DNAs are extracted directly from an environmental sample that contains multiple organisms, and the DNA fragments are cloned and sequenced. This is a powerful strategy for comprehensive analysis of biodiversity in an ecosystem including polar regions. However, with a simple collection of many sequence fragments, it is difficult to predict from what phylotypes individual sequences are derived. This is because the conventional phylogenetic classification of genomic sequences is based on sequence homology searches, which require orthologous sequence sets; and therefore, this strategy cannot be applied to poorly characterized or novel gene sequences. We developed a new phylogenetic classification method on the basis of Kohonen's Self-Organizing Map (SOM) (Kohonen, 1982, 1990; Kohonen *et al*., 1996), by modifying the SOM for genome informatics (Kanaya *et al*., 1998, 2001; Abe *et al*., 2002, 2003). In the present study, the SOM method was optimized for phylogenetic classification of genomic sequences from environmental samples.

Methods

Nucleotide sequences

Nucleotide sequences were obtained from http://www.ddbj.nig.ac.jp/anoftp-e.html. When the number of undetermined nucleotides (*Ns*) in a sequence exceeded 10% of the window size, the sequence was omitted from the analysis. When the number of *Ns* was less than 10%, the oligonucleotide frequencies were normalized to the length without *Ns* and included in the analysis.

SOM methods

A SOM implements a nonlinear projection of multi-dimensional data onto a two-dimensional array of weight vectors, and this effectively preserves the topology of the high-dimensional data space (Kohonen, 1982, 1990; Kohonen *et al*., 1996). We modified the conventional SOM for genome informatics, on the basis of batch-learning SOM, to make the learning process and resulting map independent of the order of data input (Kanaya *et al*., 2001; Abe *et al*., 2002, 2003, 2005). The batch learning SOM (BL-SOM) is suitable for actualizing high-performance parallel-computing and thus for a large scale computation using the Earth Simulator of Japan Agency for Marine-Earth Science and Technology. The batch-learning SOM program (BL-SOM) can be obtained from G-inforBIO (http://wdcm.nig.ac.jp/inforbio/).

Results and discussion

SOMs for oligonucleotide frequencies in 126 genomes

To test the classification power of the SOM for a wide range of genome sequences, we analyzed frequencies of short oligonucleotides in the 126 genomes for which complete sequences are available: 0.3 Gb for 113 prokaryotes and 3.0 Gb for 13 eukaryotes (see Fig. 1 legend). SOMs were constructed with tri- and tetranucleotide frequencies for 330000 nonoverlapping 10-kb sequences and overlapping 100-kb sequences with a sliding step size of 10 kb derived from the total 3.3-Gb sequence. To set the initial weight

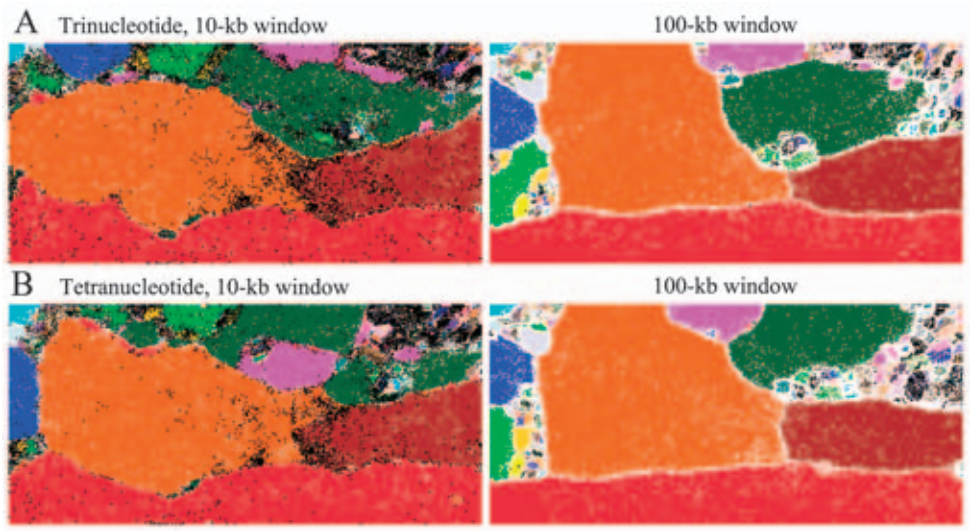

Fig. 1. SOMs for nonoverlapping 10-kb and overlapping 100-kb sequences of 126 genomes. 10-kb and 100-kb tri- (A) and tetranucleotide (B) SOMs. Lattice points that contain sequences from more than one species are indicated in black those including no sequences are indicated in white and those including sequences from a single species are indicated in colors shown in Table 1.

vectors, frequencies for the 330000 sequences were analyzed by PCA (Kanaya *et al*., 1998, 2001; Abe *et al*., 2003). After 100 learning cycles, the sequences of most species were separated (self-organized) according to species (Fig. 1A, B, Table 1). Lattice points that contained sequences from a single species are indicated in color, those including sequences from more than one species are indicated in black, and those with no sequences are indicated in white. Even in the 10-kb SOMs, most eukaryotic sequences were classified into the species-specific territories. For example, 98 and 99% of human sequences were classified into human territories (■ in Fig. 1A, B) of the tri- and tetranucleotide SOMs (Tri- and Tetra-SOMs), respectively. In the 100-kb SOMs, the species-specific separations became more evident, and many prokaryotes also occupied clear species-specific territories. The species territories were surrounded often with contiguous white lattice points into which no genomic sequences were classified. Therefore, the species borders could be drawn automatically on the basis of the contiguous white lattices. Collectively, the SOM recognized the species-specific characteristics (a key combination of oligonucleotide frequencies) that are the representative signature of each genome.

The frequencies of each tetranucleotide in each lattice point in the tetranucleotide 100-kb SOM were calculated and represented as different levels of red and blue (Fig. 2). Transitions between the red and blue levels often coincided with the species borders (Fig. 2). The clearest example was CATG, which was overrepresented in human (H), *Arabidopsis* (A), Zebra fish (Z) and rice (R), underrepresented in *Drosophila* (D), and moderately represented in Fugu (F). SOMs utilized complex combinations of multiple oligonucleotides for sequence separations resulting in effective classification according to species.

Table 1. Species names of 126 genomes (13 eukaryotes, 15 archaea and 98 bacteria) analyzed in Fig. 1.

Eukaryotes		
Arabidopsis thaliana	*Caenorhabditis elegans*	*Dictyostelium discoideum*
Drosophila melanogaster	*Entamoeba histolytica*	*Medicago truncatula*
Homo sapiens	Puffer fish *Fugu rubripes*	*Plasmodium falciparum*
Saccharomyces cerevisiae	*Schizosaccharomyces pombe*	rice *Oryza sativa*
zebrafish *Danio rerio*		
Archaea		
Aeropyrum pernix	*Archaeoglobus fulgidus*	*Halobacterium* sp.
Methanobacterium thermoautotrophicum	*Methanococcus jannaschii*	*Methanopyrus kandleri*
Methanosarcina acetivorans	*Methanosarcina mazei*	*Pyrobaculum aerophilum*
Pyrococcus abyssi	*Pyrococcus furiosus*	*Pyrococcus horikoshii*
Sulfolobus solfataricus	*Sulfolobus tokodaii*	*Thermoplasma acidophilum*
Bacteria		
Agrobacterium tumefaciens	*Anabaena* sp.	*Aquifex aeolicus*
Bacillus anthracis	*Bacillus halodurans*	*Bacillus subtilis*
Bifidobacterium longum	*Borrelia burgdorferi*	*Bradyrhizobium japonicum*
Brucella melitensis	*Brucella suis*	*Buchnera aphidicola*
Buchnera sp.	*Campylobacter jejuni*	*Caulobacter crescentus*
Chlamydia muridarum	*Chlamydia trachomatis*	*Chlamydophila pneumoniae*
Chlorobium tepidum	*Clostridium acetobutylicum*	*Clostridium perfringens*
Corynebacterium efficiens	*Corynebacterium glutamicum*	*Deinococcus radiodurans*
Escherichia coli	*Fusobacterium nucleatum*	*Haemophilus influenzae*
Helicobacter pylori	*Lactobacillus plantarum*	*Lactococcus lactis*
Listeria innocua	*Listeria monocytogenes*	*Mesorhizobium loti*
Mycobacterium leprae	*Mycobacterium tuberculosis*	*Mycoplasma genitalium*
Mycoplasma pneumoniae	*Mycoplasma pulmonis*	*Neisseria meningitidis*
Oceanobacillus iheyensis	*Pasteurella multocida*	*Pseudomonas aeruginosa*
Pseudomonas putida	*Pseudomonas syringae*	*Ralstonia solanacearum*
Rickettsia conorii	*Rickettsia prowazekii*	*Salmonella enterica*
Salmonella typhimurium	*Shewanella oneidensis*	*Shigella flexneri*
Sinorhizobium meliloti	*Staphylococcus aureus*	*Streptomyces coelicolor*
Staphylococcus epidermidis	*Streptococcus agalactiae*	*Streptococcus pneumoniae*
Streptococcus pyogenes	*Synechocystis* sp.	*Thermosynechococcus elongatus*
Thermotoga maritime	*Treponema pallidum*	*Tropheryma whipplei*
Thermoanaerobacter tengcongensis	*Thermoplasma volcanium*	*Ureaplasma urealyticum*
Vibrio cholerae	*Vibrio parahaemolyticus*	*Vibrio vulnificus*
Xanthomonas axonopodis	*Xanthomonas campestris*	*Xylella fastidiosa*
Yersinia pestis		

Application to metagenomic studies

In the metagenomic study of environmental microorganisms, DNA is extracted directly from mixed genomes in an environmental sample without cultivation, and the DNA fragments are cloned into plasmid vectors, then sequenced. For example, Tyson *et al.* (2004) applied the metagenome shotgun-sequencing to mixed genomes collected from an acidophilic biofilm growing in acid mine drainage (AMD), which is a worldwide environmental problem. They focused on this biofilm to reconstruct dominant genomes with shotgun sequencing because of low-complexity of the mixed genomes in the biofilm and deposited approximately 2455 sequence fragments in DDBJ/EMBL/GenBank. SOMs should be useful for reassociating sequence fragments from such environmental samples *in silico* according to individual genomes, because genomic sequences with similar oligonucleotide frequencies were self-organized. Such SOM method was developed as fol-

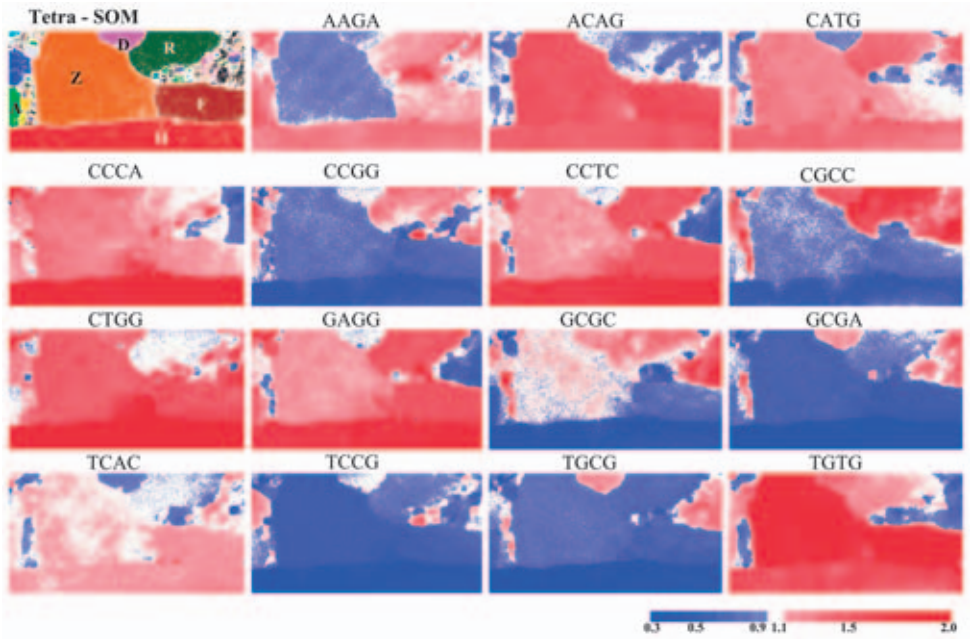

Fig. 2. Level of each tetranucleotide in the 100-kb tetranucleotide SOM. Diagnostic examples of species separations are presented. The level of each tetranucleotide for each lattice point in the 100-kb Tetra-SOM was calculated and normalized with the level expected from the mononucleotide composition of the lattice point. The observed/expected ratio is indicated in colors at the bottom of the figure. The 100-kb tetranucleotide SOM in Fig. 1 is presented in the first panel; *Arabidopsis* (A), rice (R), *Drosophila* (D), Fugu (F), Zebra fish (Z), and human (H).

lows (Abe *et al*., 2005).

In the DNA databases, only one strand of a pair of complementary sequences is registered. Our previous analyses revealed that sequence fragments from a single prokaryotic genome are often split into two territories that reflect the transcriptional polarities of the genes present in the fragment (Abe *et al*., 2003). For phylogenetic classification of sequences from uncultured microbes, it is not necessary to know the transcriptional polarity of the sequence, and the split into two territories complicates assignment to species. Therefore, we tested another type of SOM in which frequencies of a pair of complementary oligonucleotides (*e.g*., AACC and GGTT) were summed (Abe *et al*., 2005, 2006), and the SOM for the degenerate sets of tetranucleotides was designated DegeTetra-SOM. We constructed a DegeTera-SOM with all available sequences of known species compiled in the DNA databases rather than only those from completely sequenced genomes. This is because environmental samples of interest presumably contain poorly characterized and novel microorganisms. In Fig. 3, the DegeTetra-SOM constructed with 210000 non-overlapping 5-kb sequences (a total of 1.05 Gb) from 1502 species of known prokaryotes for which at least 10 kb of sequence has been deposited in DDBJ/EMBL/GenBank, is shown. These 1502 prokaryotes were then classified into 25 phylotypes with reference to the NCBI Taxonomy Database (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.

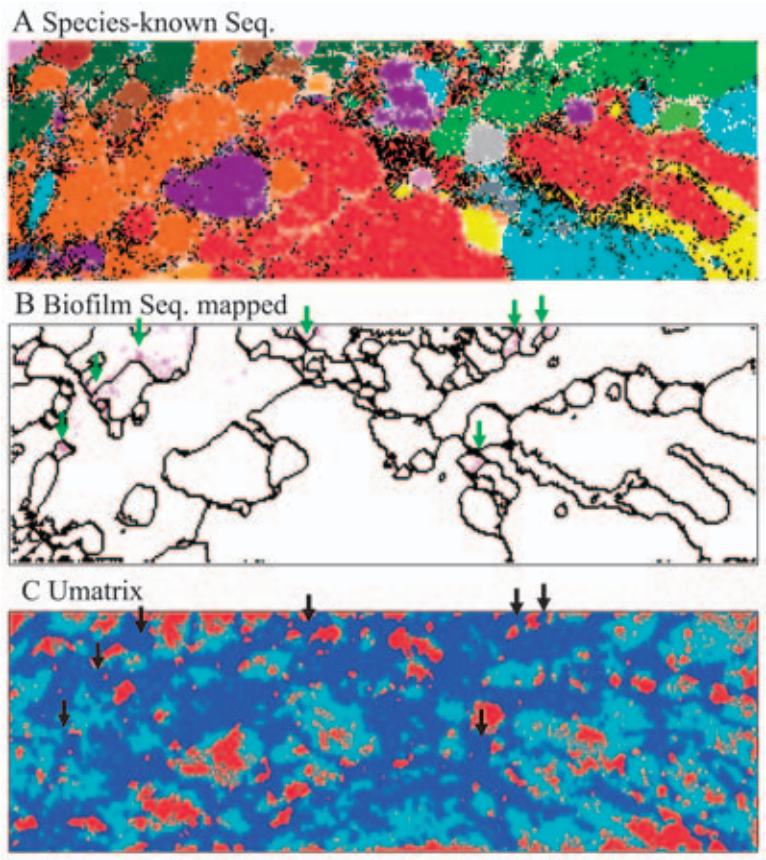

Fig. 3. SOM for phylogenetic classification of sequences from an environmental sample. (A) Species-known Seq.; DegeTetra-SOM of 5-kb sequences of species-known prokaryotes. The genomic sequences from 1502 prokaryotes were used. Prokaryotic sequences were classified into 25 phylotypes. Lattice points that include sequences from more than one phylotype are indicated in black, and those that contain sequences from a single phylotype are indicated in color shown in Table 2. (B) The biofilm sequences were mapped on the DegeTetra-SOM. The lattice points, on which biofilm sequences were mapped, are indicated in pink, and typical zones clustered with the mapped biofilm sequences are indicated with arrows. (C) Umatrix representing homogeneity (red) and heterogeneity levels of blue color in oligonucleotide frequencies among neighboring lattice points. Arrows showing biofilm sequences in (B) are redrawn.

html/). The classification according to phylotype was apparent (Species-known Seq. in Fig. 3A, Table 2); as a separate study, the classification using the 1-kb SOM was found to be significantly poorer than that with the 5-kb SOM (Abe *et al*., 2005). Then, the biofilm sequences were mapped on the 5-kb DegeTetra-SOM (Biofilm sequences are shown in Fig. 3B). Most of the biofilm sequences, which were derived from the low-complexity metagenome library, were located mostly in restricted but distinct territories, indicating that most sequence fragments derived from a single genome but cloned inde-

Table 2. Phylotypes analyzed in Fig. 3A.

α-proteobacteria (■)	β-proteobacteria (■)	γ-proteobacteria (■)
δ-proteobacteria (■)	ε-proteobacteria (■)	Actinobacteria (■)
Aquificae (■)	Bacteroidetes (■)	Chlamydiae (■)
Chlorobi (■)	Chloroflexi (■)	Crenarchaeota (■)
Cyanobacteria (■)	Deinococcus-Thermus (■)	Dictyoglomi (■)
Euryarchaeota (■)	Fibrobacteres (■)	Firmicutes (■)
Fusobacteria (■)	Nitrospirae (■)	Planctomycetes (■)
Spirochaetales (■)	Thermodesulfobacteriales (■)	Thermotogales (■)
Verrucomicrobiae (■)		

pendently could be reassociated *in silico*.

With the SOM method, distances of weight vectors between neighboring lattice points can be visualized as color levels using a Umatrix method (Kraaijveld *et al*., 1992; Ultsch, 1993), and this provides valuable information regarding levels of sequence homogeneity and heterogeneity in the respective zone (Fig. 3C). The zone with high homology levels (red), which is composed of lattice points with similar oligonucleotide frequencies, most likely represents a group of sequences derived from the same or closely related genomes; *e.g*., a large number of sequences completely derived from a single genome sequenced. Zones marked with light and dark blue represent correspondences to lattice points with moderate and low homology of the oligonucleotide pattern, respectively. Therefore, dark blue zones composed of the most heterogonous lattice points should represent sequences derived from many poorly characterized genomes, for which only a very limited portions of relevant genome sequences have been registered in current databases. Most of the biofilm sequences were mapped to the dark blue zones (refer to arrows in Fig. 3B and C), showing that these sequences are novel in the current database.

Separation between prokaryotic and eukaryotic sequences

When we consider phylogenetic classification of DNA sequences obtained from environmental samples including those from the poles, it is desirable to construct SOMs in advance with all available sequences from species-known prokaryotes and eukaryotes that have been compiled in DNA databases. This is because various eukaryotic DNAs may be present in the samples. Furthermore, when microorganisms that are symbiotic/parasitic with a higher eukaryote are analyzed, sequences from this eukaryote are included inevitably in the sequence collection. To examine the SOM separation of prokaryotic sequences from a wider range of eukaryotic sequences, 5-kb sequences from 147 prokaryotes were analyzed simultaneously with 5-kb sequences from 13 eukaryotes. To avoid excess representation of eukaryotic sequences derived from large genomes and to analyze an equivalent number of prokaryotic and eukaryotic sequences, 5-kb eukaryotic sequences were selected randomly from each eukaryote genome up to 25 Mb and the DegeTetra-SOM was constructed with these 5-kb prokaryotic and eukaryotic sequences (Fig. 4, Table 3). The power of the SOM to separate prokaryotic from eukaryotic sequences was very high; 99.5% of prokaryotic sequences were classified into prokaryotic territories, and 0.2% and 0.1% were classified into yeast *S. pombe* and *S. cerevisiae* territories, respectively. Separation among eukaryotic genomes was again apparent on the 5-kb SOM (Fig. 4B).

Next, after normalization of the sequence length we again mapped the biofilm se-

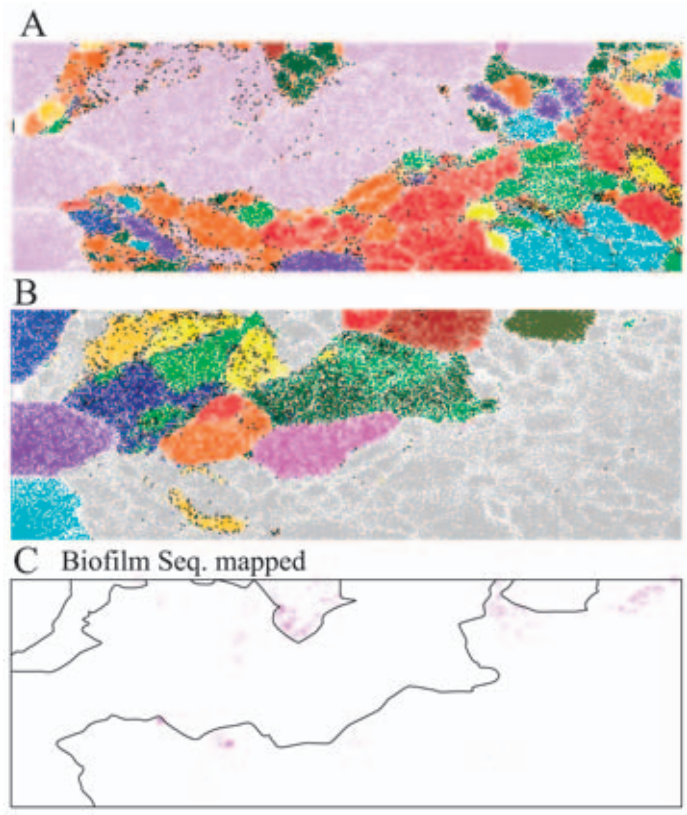

Fig. 4. DegeTetra-SOM of 5-kb sequences from 147 prokaryotes and 13 eukaryotes. (A) Prokaryotic sequences were classified into 12 phylotypes. Lattice points that contain prokaryotic sequences from more than one phylotype are indicated in black, and those that contain sequences from a single group are indicated in color shown in Table 3. Lattice points that contain sequences from both prokaryotic and eukaryotic sequences are also indicated in black and those that contain sequences only from eukaryotic genomes are indicated in color (■). (B) Lattice points that contain sequences only from prokaryotic genomes are indicated in color (■). Lattice points that contain sequences from a single eukaryotic species are indicated in color shown in 13 Eukaryotes of Table 1. Lattice points that contain sequences from more than one eukaryotic species or from both eukaryotic and prokaryotic species are indicated in black. (C) The biofilm sequences were mapped on the DegeTetra-SOM. The lattice points, on which biofilm sequences were mapped, are indicated in pink.

Table 3. Phylotypes analyzed in Fig. 4A.

α-proteobacteria (■)	β-proteobacteria (■)	γ-proteobacteria (■)
δ-proteobacteria (■)	Archaea (■)	Actinobacteria (■)
Chlamydia (■)	Cyanobacteria (■)	Firmicutes (■)
Fusobacteria (■)	Thermotogae (■)	Others (■)

quences on this SOM (Biofilm Seq. in Fig. 4C). It was apparent that a major portion of the biofilm sequences were classified into specific prokaryote territories. Because of this clear separation between prokaryotic and eukaryotic sequences, this method should be useful, not only for environmental samples, but also for clinical samples (*e.g.*, feces, sputum and snivel), which may be contaminated with eukaryotic DNAs (Hayashi *et al.*, 2005). Because no species information is required in advance, the method may be useful for identification of novel pathogenic microorganisms, including viruses that cause unidentified infectious diseases.

Concluding remarks

Novel tools are needed to enable comprehensive studies of the massive amounts of genomic sequences currently available. An unsupervised neural network algorithm, the Self-Organizing Map (SOM), is an effective tool for clustering and visualizing high-dimensional complex data on a single map. Our SOM recognized species-specific characteristics (key combinations of oligonucleotide frequencies) in sequence fragments permitting species-specific classification (self-organization) of the sequences without any information regarding the species. Because species-specific clustering on SOMs is very clear, SOMs are a powerful tool for phylotype classification of genomic sequences, especially sequence fragments obtained from mixed genomes of uncultured environmental microorganisms (Hayashi *et al.*, 2005; Uchiyama *et al.*, 2005; Abe *et al.*, 2005). For considering phylogenetic classification of environmental microorganisms, it is worthwhile to construct SOMs with all available sequences from all known species to extract sequence characteristics in a wider rage of genomes. In the case of sequences from totally novel organisms, sequences even from related species might not be represented on the SOM. Importantly, such sequences can be identified as novel ones by calculating the distance between the vector of the respective sequence data and that of the sequence-mapped lattice point (i.e., the lattice with the minimum distance from the sequence in the multidimensional space) and by comparing the distance with those for sequences belonging to one known phylotype.

Acknowledgments

This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) and for Grant-in-Aid for Scientific Research on Priority Areas “Applied Genomics” and by a grant from the Advanced and Innovational Research Program in Life Sciences, from the Ministry of Education, Culture, Sports, Science and Technology of Japan. A part of the present computation was done with the Earth Simulator of Japan Agency for Marine-Earth Science and Technology.

References

Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. and Ikemura, T. (2002): A novel bioinformatic strategy for unveiling hidden genome signatures of eukaryotes: Self-organizing map of oligonucleotide frequency. Genome Inform. Ser. Workshop Genome Inform., **13**, 12–20.

Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. and Ikemura, T. (2003): Informatics for unveiling hidden genome signatures. Genome Res., **13**, 693–702.

Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S. and Ikemura, T. (2005): Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. DNA Res., **12**, 281–290.

Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S. and Ikemura, T. (2006): A large-scale Self-Organizing Map (SOM) unveils sequence characteristics of a wide range of eukaryote genomes. Gene, **365**, 27–34.

Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995): Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev., **59**, 143–169.

DeLong, E.F. (2002): Microbial population genomics and ecology. Curr. Opin. Microbiol., **5**, 520–524.

Hayashi, H., Abe, T., Sakamoto, M., Ohara, H., Ikemura, T., Sakka K. and Benno, Y. (2005): Direct cloning of genes encoding novel xylanases from human gut. Can. J. Microbiol., **51**, 251–259.

Hugenholtz, P. and Pace, N.R. (1996): Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol., **14**, 190–197.

Kanaya, S., Kudo, Y., Abe, T., Okazaki, T., Carlos, D.C. and Ikemura, T. (1998): Gene classification by self-organization mapping of codon usage in bacteria with completely sequenced genome. Genome Inform. Ser. Workshop Genome Inform., **9**, 369–371.

Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H. and Ikemura, T. (2001): Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. Gene, **276**, 89–99.

Kohonen, T. (1982): Self-organized formation of topologically correct feature maps. Biol. Cybern., **43**, 59–69.

Kohonen, T. (1990): The self-organizing map. Proc. IEEE, **78**, 1464–1480.

Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J. (1996): Engineering applications of the self-organizing map. Proc. IEEE, **84**, 1358–1384.

Kraaijveld, M.A., Mao, J. and Jain, A.K. (1992): A non-linear projection method based on Kohonen's topology preserving maps. Proceedings of the 11th International Conference on Pattern Recognition, 44.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004): Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature, **428**, 37–43.

Uchiyama, T., Abe, T., Ikemura, T. and Watanabe, K. (2005): Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nature Biotechnol., **23**, 88–93.

Ultsch, A. (1993): Self organized feature maps for monitoring and knowledge aquisition of a chemical process. Proceedings of the International Conference on Artificial Neural Networks, 864.