

## Research Article

# A Solution for Universal Classification of Species Based on Genomic DNA

Mariko Kouduka,<sup>1</sup> Daisuke Sato,<sup>1</sup> Manabu Komori,<sup>1</sup> Motohiro Kikuchi,<sup>2</sup> Kiyoshi Miyamoto,<sup>3</sup> Akinori Kosaku,<sup>3</sup> Mohammed Naimuddin,<sup>1,4</sup> Atsushi Matsuoka,<sup>5</sup> and Koichi Nishigaki<sup>1,6</sup>

<sup>1</sup>Department of Functional Materials Science, Saitama University, Saitama, Japan

<sup>2</sup>Chitose Salmon Aquarium Chitose, Youth Educational Foundation, Chitose, Hokkaido, Japan

<sup>3</sup>Institute of Medical Science, Dokkyo Medical University, Tochigi, Japan

<sup>4</sup>Biol. Res. and Functions, National Inst. AIST, Tsukuba, Ibaraki, Japan

<sup>5</sup>Department of Geology, Niigata University, Niigata, Japan

<sup>6</sup>Rational Evolutionary Design of Advanced Biomolecules, Saitama Small Enterprise Promotion Corporation, SKIP City, Saitama, Japan

Received 22 July 2006; Revised 8 October 2006; Accepted 8 October 2006

Recommended by Cheng-Cang Wu

Traditionally, organisms have been classified on the basis of their phenotype. Recently, genotype-based classification has become possible through the development of sequencing technology. However, it is still difficult to apply sequencing approaches to the analysis of a large number of species due to the cost and labor. In most biological fields, the analysis of complex systems comprising various species has become an important theme, demanding an effective method for handling a vast number of species. In this paper, we have demonstrated, using plants, fish, and insects, that genome profiling, a compact technology for genome analysis, can classify organisms universally. Surprisingly, in all three of the domains of organisms tested, the phylogenetic trees generated from the phenotype topologically matched completely those generated from the genotype. Furthermore, a single probe was sufficient for the genome profiling, thereby demonstrating that this methodology is universal and compact.

Copyright © 2007 Mariko Kouduka et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Because the functional and morphological diversities of an organism represent the value of the organism itself, the traditional biological techniques used to characterize these properties provide indispensable information. Conventional biology techniques face difficulties, however, such as classifying characterless organisms like microbes [1] and analyzing communities composed of huge numbers of various organisms [2], owing to both the instability of phenotypes, which are easily affected by environmental factors [3], and an insufficient number of experts [4].

A potential solution to these problems has been to characterize an organism according to the sequence of the small subunit ribosomal RNA (16S/18S rRNA), an approach that has been applied to various organisms [5–7]. Similarly, cytochrome oxygenase subunit 1 (COD1), gyrase, and other genes have been used for this purpose [8]. The superiority of

these approaches is that they are based on popular and well-established sequencing technology and can provide the determinate result of nucleotide sequence, which can be further computer-analyzed and can fuel the activity of bioinformatics. Nevertheless, this approach cannot be said to be a readily usable method for classifying species because (i) it is rather costly and time-consuming for application to a large number of species (e.g., >100), especially for scientists in general all over the world, and (ii) it often results in an insufficient amount of information for identifying and classifying species [8]. The latter problem can be overcome by sequencing additional genes [8–10]; however, this makes the approach more complicated and less accessible.

Here, we present a solution for the universal classification of species together with a demonstration of its effectiveness, which has been tested by applying it to taxonomically well-established organisms such as plants, fish, and insects. Genome profiling (GP) has already been established as

a method for the identification of species [11], and has sometimes been applied to clustering of organisms [12]. Due to the nature of the samples used in GP (mostly, bacteria, fungi, and protozoa, which taxonomically can sometimes be subject to debate), it has not been possible to establish GP as a technique for classification up till now. However, here we show for the first time that GP [13, 14] is successfully applied to the purpose of classification. Owing to its convenience and its highly informative nature, this technique of classification by GP can be widely applied to biological research, including botanical research.

## 2. MATERIALS AND METHODS

### 2.1. Materials

The plant, insect, and fish species used in this study are listed in Table 1.

### 2.2. DNA preparation

All samples were well washed with distilled water prior to DNA extraction. In particular, the legs of insects were vigorously washed with SDS detergent to remove contaminating microbes. DNAs of plants were prepared according to the conventional alkaline extraction method [16], whereas those of fish and insects were extracted by the phenol-chloroform-proteinase K-method [17] using a tiny portion (~0.05 mg) of caudal fins or legs. For convenience, here we define genomic DNA as the whole set of DNAs contained in cell [13].

### 2.3. Genome profiling

GP is a well-established method [18] comprising the three following major steps: random PCR [19], temperature gradient gel electrophoresis (TGGE), [20] and data-normalization by computer-processing [18]. Random PCR has the ability to pick up, for example, the top ten DNA fragments that are generated by more stable hybridization of the primer DNA. The random PCR solution (50  $\mu$ l) contained 200  $\mu$ M dNTPs (N = G, A, T, C), 0.5  $\mu$ M primer, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.02 unit/ $\mu$ l Taq DNA polymerase (Takara Bio, Japan) and template DNA (arbitrary amount). Random PCR was carried out with 30 cycles of denaturation (94°C, 30 s), annealing (28°C, 2 min) and extension (47°C, 2 min) using a PTC-100TM PCR machine (MJ Research, Inc, USA). DNA samples, together with the internal reference DNA, were subjected to  $\mu$ TGGE (micro-temperature gradient gel electrophoresis) [21] (one inch square size) using a  $\mu$ -TG apparatus (Taitec, Japan) with a linear temperature gradient of 15°C to 60°C. The gel used was 6% acrylamide (acrylamide: bis = 19 : 1) containing 90 mM Tris-HCl, 90 mM boric acid, 2 mM EDTA (pH 8.0), and 8 M urea. Detection of DNA bands was carried out either by monitoring fluorescence with a fluorescence imager, Molecular Imager FX (BioRad, USA), or by staining with silver [12]. From the resulting band patterns, which were rather complicated, characteristic or “featuring” points (e.g.,

kinked points) were extracted and then processed with the aid of computer to generate *spiddos* (species identification dots) [11, 18]. Sets of *spiddos* are able to provide a sufficient amount of information for provisionally identifying species, which is usually done by calculating the pattern similarity score (PaSS) between two genomes [18]. Using *spiddos*, we can define PaSS of the genomes between two species as follows:

$$\text{PaSS} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{|\vec{P}_i - \vec{P}'_i|}{|\vec{P}_i| + |\vec{P}'_i|}, \quad 0 \leq \text{PaSS} \leq 1, \quad (1)$$

where  $\vec{P}_i$  and  $\vec{P}'_i$  correspond to the normalized positional vectors (composed of two elements, mobility, and temperature) for *spiddos*  $P_i$  and  $P'_i$  collected from two genome profiles (discriminated with or without a prime), respectively, and  $i$  denotes the serial number of *spiddos* [18].

GP needs to be carried out with the following specific precautions to get successful results. (i) During random PCR, contamination by other organisms should be carefully avoided. In particular, the entire random PCR solution without the template DNA should be UV-irradiated prior to the PCR reaction in order to inactivate any contaminating DNAs that could act as the template [12]. (ii) The random PCR reaction should be terminated before the primer molecules are consumed in order to attain a “double-strand stop,” which means that the major PCR products are in a double-stranded state and are suitable for TGGE analysis (i.e., the melting transition of double stranded DNA to single stranded one can be detected). (iii) The GP pattern should be strictly checked from the viewpoint of “quality score” in order to rule out false positives: the number of bands (usually more than eight are required) and the clearness of bands and background should be sufficient (note that random PCR generates DNA products at different molecular ratios (eventually, the sum of them is equivalent to that of the input primer) and, sometimes, overexpression of highly expressed DNAs (where the primer binds strongly for forward and reverse extension) suppresses the appearance of lower expressed ones, leading to less than eight observable bands). GP patterns that are sufficient in both the number and the clarity of bands are categorized “quality A” and used for the further analysis; those that are sufficient in only one of the two (but the other is still permissible) are categorized “quality B” and can be used with caution (note that quality B patterns were not used in this study); and the remaining patterns (“quality C”) should be discarded and the whole experiment should be re-tried.

### 2.4. Cluster analysis

We developed a clustering and displaying program termed “Free Lighter” on the basis of *Ward’s method* [22, 23], which is a type of nearest neighbor method with an objective function of minimizing the “error sum of squares.” We also tested other derivative methods such as the *group average method*,

TABLE 1: Taxonomy<sup>†</sup> of the species dealt with in this study.

No.	Species/conventional name	Family	Order	Calss	Phylum
A1	<i>Typha orientalis</i> /Bulrush sp.	Typhaceae	Typhales	Mono*	Anth*
A2	<i>Arundinaria argenteostriata</i> /Bamboo sp.	Poaceae	Cyperales	Mono*	Anth*
A3	<i>Tricyrtis hirta</i> /Lily sp.	Liliaceae	Liliales	Mono*	Anth*
A4	<i>Cosmos bipinnatus</i> /Cosmos sp.	Asteraceae	Asterales	Dico*	Anth*
A5	<i>Taraxacum officinale</i> /Dandelion sp.	Asteraceae	Asterales	Dico*	Anth*
A6	<i>Callicarpa dichotoma</i> /Beauty-berry sp.	Verbenaceae	Lamiales	Dico*	Anth*
A7	<i>Gardenia jasminoides</i> /Gardenia sp.	Rubiaceae	Rubiales	Dico*	Anth*
A8	<i>Papaver nudicaule</i> /Poppy sp.	Papaveraceae	Papaverales	Dico*	Anth*
A9	<i>Viola xwittrockiana</i> /Pansy sp.	Violaceae	Violales	Dico*	Anth*
A10	<i>Camellia sasanqua</i> /Camellia sp.	Theaceae	Theales	Dico*	Anth*
A11	<i>Davidia involucrata</i> /Dove tree sp.	Davidiaceae	Cornales	Dico*	Anth*
A12	<i>Hydrangea macrophylla</i> /Hydrangea sp.	Hydrangeaceae	Rosales	Dico*	Anth*
B1	<i>Chilocorus rubidus</i> /Beetle sp. 1	Coccinellidae	Coleoptera	Inse*	Arth*
B2	<i>Oxycetonia jucunda</i> /Beetle sp. 2	Scarabaeidae	Coleoptera	Inse*	Arth*
B3	<i>Bombylius major</i> /Horse fly sp.	Bombyliidae	Diptera	Inse*	Arth*
B4	<i>Camponotus japonicus</i> /Ant sp. 1	Formicidae	Hymenoptera	Inse*	Arth*
B5	<i>Formica japonica</i> /Ant sp. 2	Formicidae	Hymenoptera	Inse*	Arth*
B6	<i>Apis mellifera</i> /Bee sp.	Apidae	Hymenoptera	Inse*	Arth*
B7	<i>Limnitis camilla</i> /Butterfly sp. 1	Nymphalidae	Lepidoptera	Inse*	Arth*
B8	<i>Anthocharis scolymus</i> /Butterfly sp. 2	Pieridae	Lepidoptera	Inse*	Arth*
B9	<i>Pieris rapae crucivora</i> /Butterfly sp. 3	Pieridae	Lepidoptera	Inse*	Arth*
B10	<i>Eurema laeta</i> /Butterfly sp. 4	Pieridae	Lepidoptera	Inse*	Arth*
B11	<i>Gonolabis marginalis</i> /Earwig sp.	Anisolabididae	Dermaptera	Inse*	Arth*
B12	<i>Bothrogonia ferruginea</i> /Stinkbug sp.	Cicadellidae	Hemiptera	Inse*	Arth*
B13	<i>Blattella germanica</i> /Cockroach sp.	Blattellidae	Blattaria	Inse*	Arth*
B14	<i>Reticulitermes speratus</i> /Termite sp.	Rhinotermitidae	Isoptera	Inse*	Arth*
C1	<i>Oncorhynchus masou</i> /Salmon sp. 1	Salmonidae	Salmoniformes	Acti*	Chor*
C2	<i>Oncorhynchus tshawytscha</i> /Salmon sp. 2	Salmonidae	Salmoniformes	Acti*	Chor*
C3	<i>Oncorhynchus mykiss</i> /Rainbow trout	Salmonidae	Salmoniformes	Acti*	Chor*
C4	<i>Salmo trutta</i> /Brown trout	Salmonidae	Salmoniformes	Acti*	Chor*
C5	<i>Salvelinus malma malma</i> /Dolly Varden	Salmonidae	Salmoniformes	Acti*	Chor*
C6	<i>Salvelinus leucomaenis</i> /Whitespotted char	Salmonidae	Salmoniformes	Acti*	Chor*
C7	<i>Hucho perryi</i> /Japanese huchen	Salmonidae	Salmoniformes	Acti*	Chor*
C8	<i>Osmerus eperlanus mordax</i> /Rainbow smelt	Osmeridae	Salmoniformes	Acti*	Chor*
C9	<i>Cyprinus carpio</i> /Carp sp. 1	Cyprinidae	Cypriniformes	Acti*	Chor*
C10	<i>Phoxinus phoxinus</i> /Carp sp. 2	Cyprinidae	Cypriniformes	Acti*	Chor*
C11	<i>Misgurnus anguillicaudatus</i> /loach sp. 1	Cobitidae	Cypriniformes	Acti*	Chor*
C12	<i>Barbatula barbatula</i> /loach sp. 2	Balitoridae	Cypriniformes	Acti*	Chor*
C13	<i>Silurus asotus</i> /Amur cat fish sp.	Siluridae	Siluriformes	Acti*	Chor*
C14	<i>Cottus nozawae</i> /Bullhead sp.	Cottidae	Scorpaeniformes	Acti*	Chor*

<sup>†</sup>This table is built based on NCBI's Taxonomy (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=taxonomy>) and Iwanami Biology Encyclopedia, 4th edition [15].

\*Mono: Monocotyledonopsida, Dico: Dicotyledonopsida, Anth: Anthophyta, Inse: Insecta, Acti: Actinopterygii, Chor: Chordata.

median method, furthest neighbor method, as well as Ward's method, thereby confirming the well-known phenomenon of occasional minor changes in phylogenetic trees. These methods are based on the distance defined in (2) which implies that Clusters  $a$  and  $b$  are to be merged into  $c$ , and  $x$  is an arbitrary cluster:

$$d_c = \alpha_a d_{xa} + \alpha_b d_{xb} + \beta d_{ab} + \gamma |d_{xa} d_{xb}|, \quad (2)$$

where  $\alpha_a$ ,  $\alpha_b$ ,  $\beta$ , and  $\gamma$  are weighing parameters,  $d_c$ ,  $d_{xa}$ ,  $d_{xb}$ , and  $d_{ab}$  represent distances between relevant clusters such as Cluster  $x$  and Cluster  $a$  for  $d_{xa}$ .

The parametric differences among these methods are  $\alpha_a = n_a/n_c$ ,  $\alpha_b = n_b/n_c$ ,  $\beta = 0$ ,  $\gamma = 0$  for the group average method;  $\alpha_a = 0.5$ ,  $\alpha_b = 0.5$ ,  $\beta = 0.25$ ,  $\gamma = 0$  for the median method;  $\alpha_a = 0.5$ ,  $\alpha_b = 0.5$ ,  $\beta = 0$ ,  $\gamma = 0.5$  for the furthest neighbor method; and  $\alpha_a = (n_x + n_a)/(n_x + n_c)$ ,  $\alpha_b = (n_x + n_b)/(n_x + n_c)$ ,  $\beta = n_x/(n_x + n_c)$ ,  $\gamma = 0$  for Ward's method. In this experiment,  $d_{ab}$  was set to be 1-PaSS ( $a, b$ ). The clustering results were found to be rather robust against changes in the above parameters, although there was a slight change in the order of neighbor joining in several cases.

### 3. RESULTS AND DISCUSSION

#### 3.1. The method employed: genome profiling

The GP technology comprises three essential steps [13, 14, 18]. The first step is to extract DNA fragments specifically from the genomic DNA of an organism through random PCR [24], resulting in the specific reduction of the amount of DNA to be analyzed. The second step is  $\mu$ TGGE to obtain the sequence-related information (i.e., a property related to the melting temperature of DNA [25]). The final step is to computer-process the experimental raw data (i.e., the band patterns in the gel) to obtain a normalized digital pattern (*spiddos*) that can be used for further analyses such as species identification and clustering [12, 18].

Figure 1 outlines the whole procedure used in this study to classify species by GP. Random PCR is a process in which DNA fragments are sampled at random from genomic DNA through the hybridization of a mismatch-containing primer to the template DNA during PCR [24]. In other words, this process is equivalent to the statistical approach used in random-sampling in a public opinion poll, from which an unbiased image of the whole can be grasped.  $\mu$ TGGE [21] is used to get information related to the melting temperature ( $T_m$ ) of the sampled DNAs which is sequence-dependent [20]. Importantly, an internal reference DNA should be co-migrated on the gels in order to obtain experimental fluctuation-free (or normalized) data by the subtraction method [18]. Those featuring points that appear in electrophoretic band patterns (i.e., start of the melting transition; see Figure 2) are marked and converted to provide the coordinates of *spiddos* (species identification dots [18]).

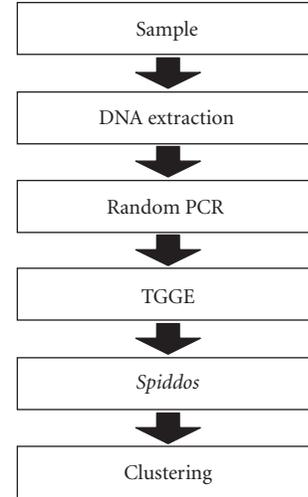


FIGURE 1: The procedure used to classify species by GP.

By calculating a pattern similarity score (PaSS) between two genomes using *spiddos* as defined in Materials and Methods, we can get the information on how closely the two relevant genomes (organisms) are related ( $0 \leq \text{PaSS} \leq 1$ ; PaSS = unity for a complete match). The PaSS value is known to be strongly correlated with the relatedness between two genomes although PaSS itself is based on a stochastic process [18]. This means that the random-sampling method cannot rule out the possibility of selecting a biased subpopulation; thus, the larger the sampling number of sampling becomes, the closer to the true image the sampled one will be. This is also the case with the GP approach; however, the sampling number of sampling can be increased by carrying out another random PCR using a different primer, resulting in the accumulation of information on a genome [13]. As far as the experiments that we have done are concerned (several hundreds of species and strains), a strong correlation between the PaSS value and the relatedness of two genomes has been observed [12, 18, 26], which can be theoretically rationalized [11] and is also experimentally verified in this paper (taking into consideration all of these facts, the PaSS value can be assumed to be semiquantitative as discussed later).

#### 3.2. Three domains of organisms tested

Here, to test our method, we collected and used three domains of organisms—namely, plants, fish, and insects—that are taxonomically well established (Table 1). Figures 3(a)–3(c) shows the species that we tested here and the PaSS values obtained among them, respectively. To illustrate the technique, some of the results of GP and *spiddos* representations for plant and fish are shown in Figure 2, where the pairs of Panels a/b (A1: *Typha orientalis*, A9: *Viola xwittrockiana*) and Panels e/f (C1: *Oncorhynchus masou*, C4: *Salmo trutta*) apparently represent closer relationships than do the pairs of Panels c/d (A11: *Davidia involucreta*, A12: *Hydrangea*

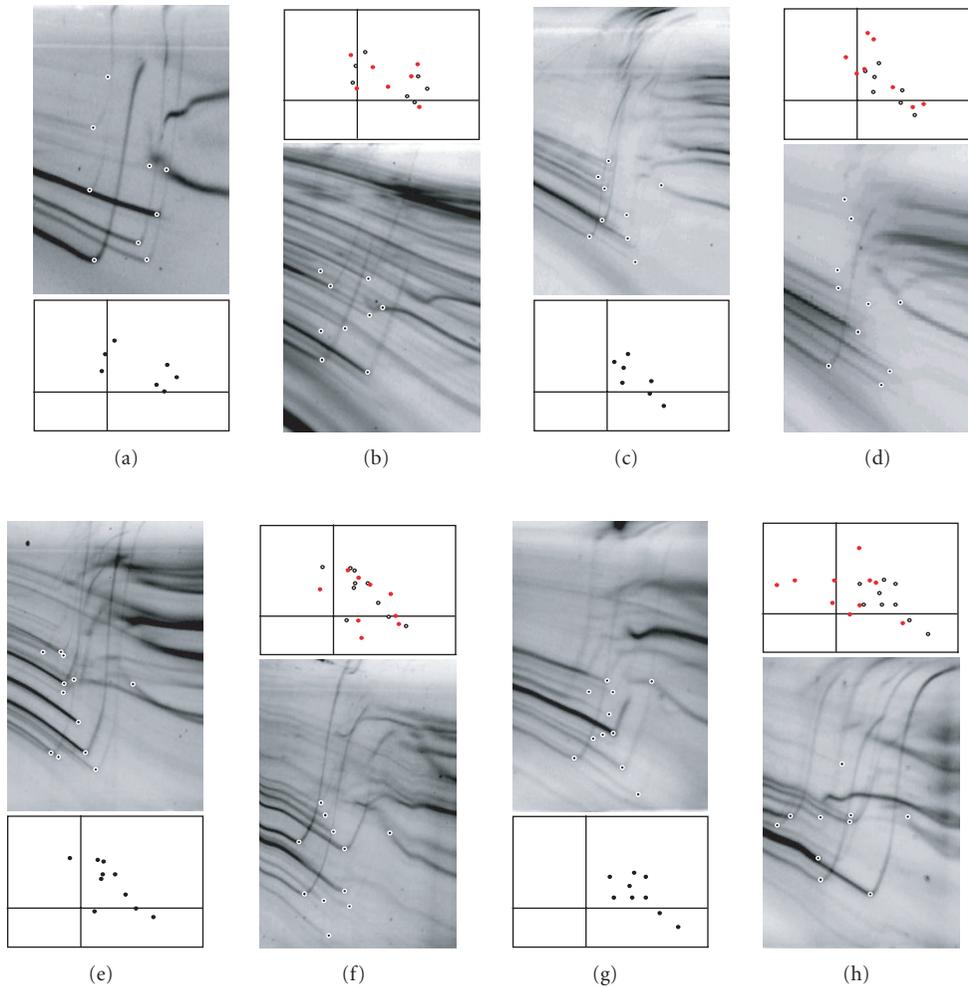


FIGURE 2: Genome profiles and *spiddos* patterns. Genome profiles of eight species of plant and fish are shown (top and bottom) together with their *spiddos* patterns (two Panels in the center): (a) Bulrush (A1); (b) Pansy (A9); (c) Dove tree (A11); (d) Hydrangea (A12); (e) Chinook salmon (C1); (f) Brown trout (C4); (g) European smelt (C8); (h) Stone loach (C12) (the same symbols as in Table 1 are used in the parenthesis). In the photographs, electrophoresis was performed from top to bottom with the temperature gradient running from left (low) to right (high temperature). The featuring points are plotted with a small black dot, whereas those corresponding to the internal reference DNA are plotted in white. *Spiddos* (i.e., the normalized coordinates of featuring points) are dark in color (middle-upper panels) and red (middle-lower panels). For the comparison, blank *spiddos* (corresponding to those in the panel immediately above) are superimposed in each lower panel. The pairs a/c and e/f are relatively close, while the pairs c/d and g/h are relatively distant.

*macrophylla*) and g/h (C8: *Osmerus eperlanus mordax*, C12: *Barbatula barbatula*), as expected. The results obtained here for plants, fish, and insects were clustered (nearest neighbor method [27]) to determine intra-domains. On the basis of taxonomical knowledge, which has been established according to phenotypic traits, and the clustering results, two phylogenetic trees were constructed as shown in Figure 4. All of the organisms examined were classified topologically to the same position in both the phenotype- and the genotype-based trees (note that, throughout this study, no arbitrary selection of data was made except for ruling out a few low-quality data samples, meaning that the correspondence between the phenotype-based and the genotype-based classifications was perfect so far as tested).

### 3.3. Unexpected findings

It is surprising that such a limited amount of information (GP obtained with only a single oligonucleotide probe) can provide such perfect results for all organisms tested. Although it had been believed that GP should be able to classify species [28], it was considered that it would be better to use three or more probes per genome. Of course, another surprise is the strong correspondence between the results obtained by two quite different approaches: the traditional phenotype-based taxonomy (which by its nature is based on the well-considered but rather arbitrarily selected traits) and a genome sequence-based one (which is not directly based on sequence information itself). Theoretically, some ranks

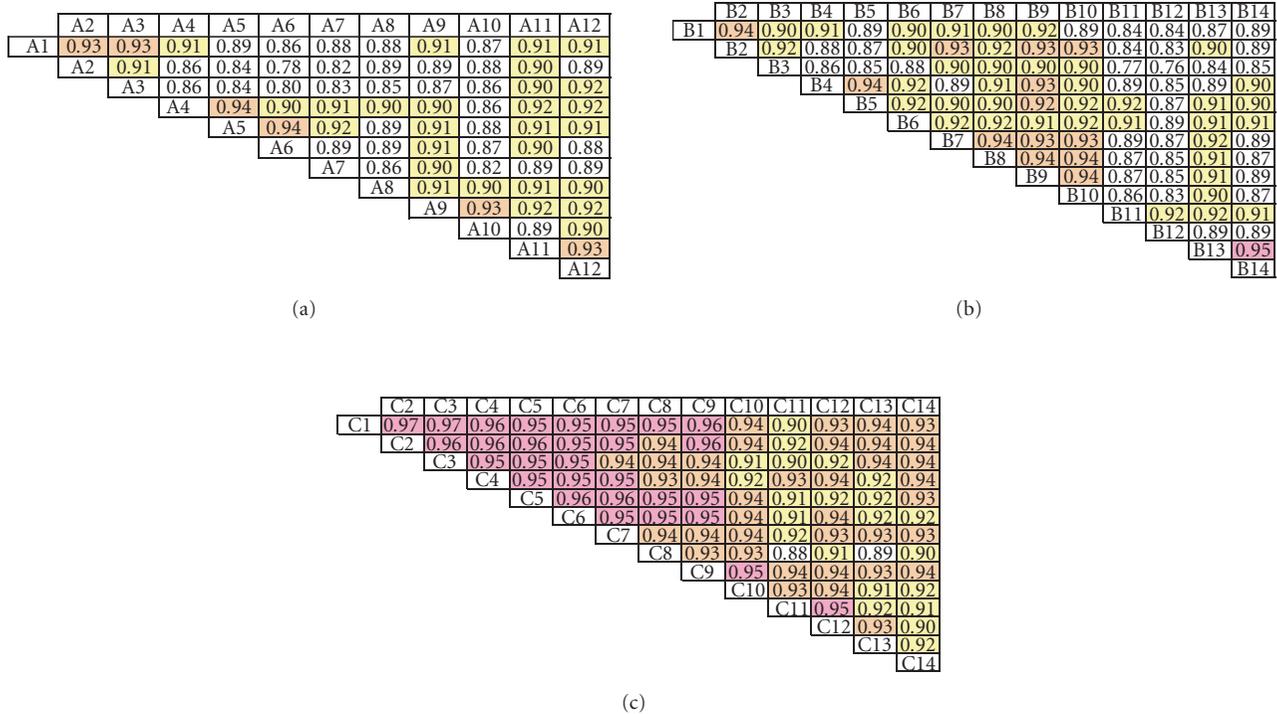


FIGURE 3: PaSS matrices for three domains of organisms. Colors represent the degree of PaSS values: Pink ( $\geq 0.95$ ), Orange ( $< 0.95$  and  $\geq 0.93$ ), yellow ( $\leq 0.92$  and  $\geq 0.90$ ), and white ( $< 0.90$ ).

of hierarchy must be interconvertible in phenotype-based taxonomy so that such a correspondence is not a matter of course.

### 3.4. Methodological benefits

Apart from the rRNA and the other sequencing approaches described above, this is the first report describing the procedure of making phylogenetic trees of mutually-distant organisms based on the same criterion. Because the rRNA approach needs different pairs of primers to analyze a wide range of organisms, our simple approach is advantageous and can be used to complement the rRNA one. Notably, the length of branches in the phenotype-based tree is arbitrary and mainly implies topological meaning, whereas those in the genotype-based tree have a quantitative meaning: the longer the tree is, the more distantly related the organisms are. The results obtained here indicate that the quantitative expression of PaSS is very effective to some extent, even though the accuracy of the measure given by PaSS is thought to be limited, a priori, due to its stochastic nature [11] (i.e., there are some steps that are stochastic in nature and can influence determination of the PaSS value: for example, random PCR may or may not select a DNA fragment containing mutations, and the degree of displacement caused by a point mutation depends on the type of mutation such as A to G or A to C substitution [19], among others; further consideration will lead to the conclusion that this is the case not only

with the GP approach but also with other approaches that depend on the comparison of a particular gene sequence [29]). In other words, even though the clustering of the organisms considered here was done, for the sake of simplicity, on the basis of a single experimental result obtained with a common oligonucleotide probe (dAGAACGCGCCTG, pfm12), the results were taxonomically consistent. At the current level of data (i.e., relatively small-scale sampling), we may be able to suggest conservatively that fish are widely classified on the basis of a more limited number of genes than are other organisms such as plants and insects, as can be seen in Figures 2 and 4, where relatively small differences in the genome sequence (measured by PaSS) are observed among species of fish, although the possibility of biased sampling cannot be completely ruled out.

### 3.5. Expandable amount of information

Evidently, by using more kinds of probes, taxonomically more reliable results can be obtained, as we have demonstrated experimentally for other organisms such as fungi and rice (to be published elsewhere). An excellent feature of this methodology is that the amount of information used for classification can be increased on demand [18] without limitation, simply by repeatedly performing an additional random PCR with a different oligonucleotide probe and thus obtaining additional *spiddos* (which can be expressed as “*information-scalable*”). Thus, this methodology has the

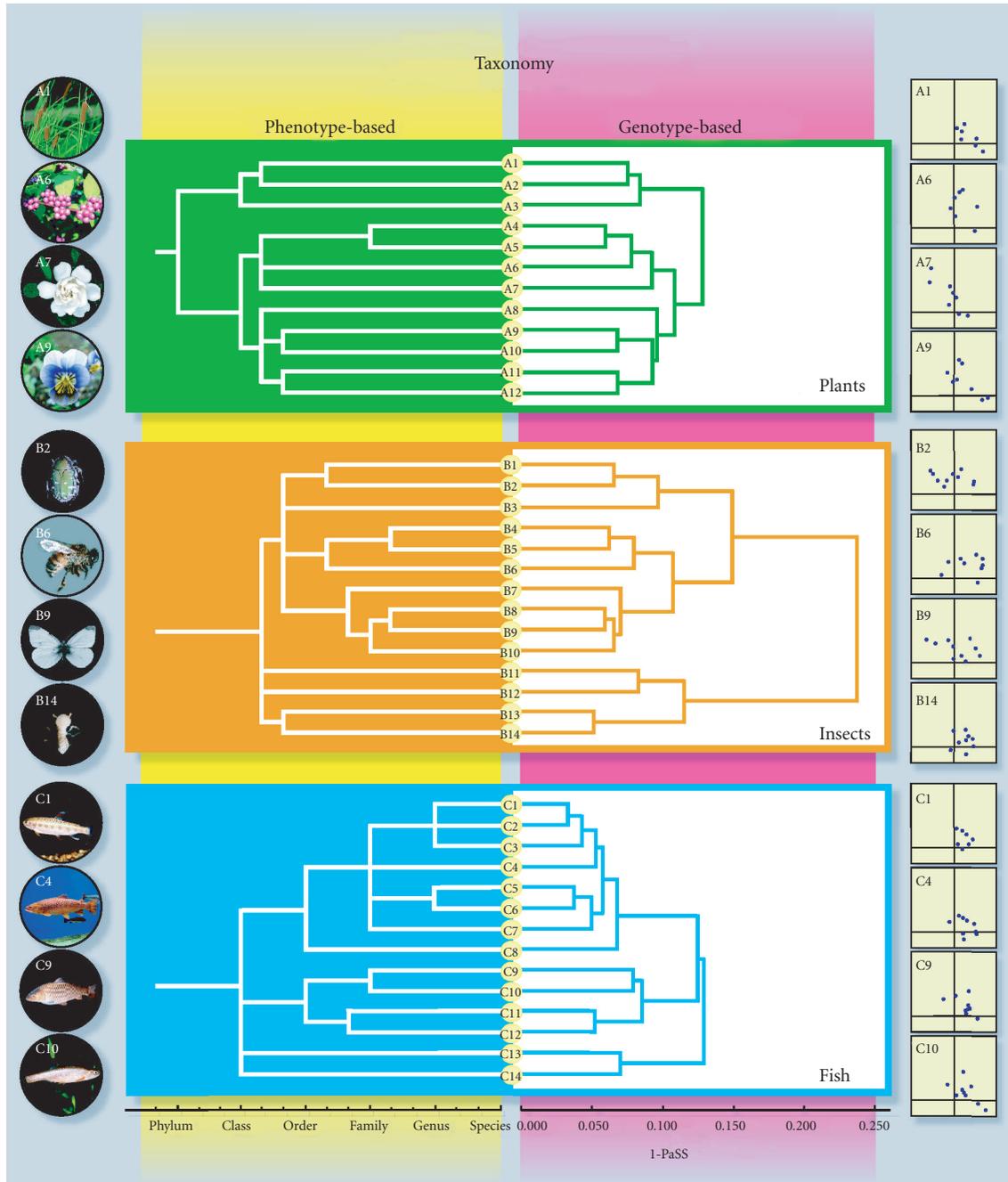


FIGURE 4: Phylogenetic trees of plants (A1 ~ A12), insects (B1 ~ B14), and fish (C1 ~ C14). Phenotypic (left) and genotypic (right) trees are drawn on the basis of taxonomic hierarchy or PaSS value, respectively. The same nomenclature as in Table 1 are used. Photographs (far left) and *spiddos* (far right) are included to illustrate the technique. Trees were drawn by the *group average method* (plants) or the *median method* (insects and fish).

potential to become a highly accurate classification tool, as well as a convenient, universal one. Moreover, it has been already demonstrated that DNAs that provide *spiddos* can be collected and sequenced if necessary [26, 30].

As mentioned above, GP has proved to be useful for identifying species [12, 14], in addition to classifying species as

shown here. It could therefore be said that the ability of GP to classify species was indirectly supported by the earlier studies, but it had not been explicitly demonstrated. We cannot but feel the splendor of the correspondence between the phenotypic world and the genotypic one, although its meaning needs to be considered more deeply hereafter.

## REFERENCES

- [1] R. I. Amann, W. Ludwig, and K.-H. Schleifer, "Phylogenetic identification and in situ detection of individual microbial cells without cultivation," *Microbiological Reviews*, vol. 59, no. 1, pp. 143–169, 1995.
- [2] J. L. Sebat, F. S. Colwell, and R. L. Crawford, "Metagenomic profiling: microarray analysis of an environmental genomic library," *Applied and Environmental Microbiology*, vol. 69, no. 8, pp. 4927–4934, 2003.
- [3] B. G. Miner, S. E. Sultan, S. G. Morgan, D. K. Padilla, and R. A. Relyea, "Ecological consequences of phenotypic plasticity," *Trends in Ecology and Evolution*, vol. 20, no. 12, pp. 685–692, 2005.
- [4] American Museum of Natural History, "The Global Taxonomy Initiative: using systematic inventories to meet country and regional needs," in *DIVERSITAS/Systematics Agenda 2000 International Workshop*, New York, NY, USA, 1999.
- [5] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4576–4579, 1990.
- [6] B. L. Maidak, J. R. Cole, C. T. Parker Jr., et al., "A new version of the RDP (Ribosomal Database Project)," *Nucleic Acids Research*, vol. 27, no. 1, pp. 171–173, 1999.
- [7] J. R. Cole, B. Chai, R. J. Farris, et al., "The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis," *Nucleic Acids Research*, vol. 33, pp. D294–D296, 2005.
- [8] R. Urwin and M. C. J. Maiden, "Multi-locus sequence typing: a tool for global epidemiology," *Trends in Microbiology*, vol. 11, no. 10, pp. 479–487, 2003.
- [9] M. C. J. Maiden, J. A. Bygraves, E. Feil, et al., "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, pp. 3140–3145, 1998.
- [10] A. Sorokin, B. Candelon, K. Guilloux, et al., "Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains," *Applied and Environmental Microbiology*, vol. 72, no. 2, pp. 1569–1578, 2006.
- [11] M. Kouduka, A. Matsuoka, and K. Nishigaki, "Acquisition of genome information from single-celled unculturable organisms (radiolaria) by exploiting genome profiling (GP)," *BMC Genomics*, vol. 7, p. 135, 2006.
- [12] T. Watanabe, A. Saito, Y. Takeuchi, M. Naimuddin, and K. Nishigaki, "A database for the provisional identification of species using only genotypes: web-based genome profiling," *Genome Biology*, vol. 3, no. 2, pp. 1–8, 2002, research0010.
- [13] K. Nishigaki, M. Naimuddin, and K. Hamano, "Genome profiling: a realistic solution for genotype-based identification of species," *Journal of Biochemistry*, vol. 128, no. 1, pp. 107–112, 2000.
- [14] M. Naimuddin and K. Nishigaki, "Genome analysis technologies: towards species identification by genotype," *Briefings in Functional Genomics and Proteomics*, vol. 1, no. 4, pp. 356–371, 2003.
- [15] R. Yasugi, O. Koseki, M. Hurutani, and T. Hidaka, *Iwanami Biology Encyclopedia*, Iwanami Press, Tokyo, Japan, 4th edition, 1996.
- [16] H. Wang, M. Qi, and A. J. Cutler, "A simple method of preparing plant samples for PCR," *Nucleic Acids Research*, vol. 21, no. 17, pp. 4153–4154, 1993.
- [17] A. Shinomiya, M. Matsuda, S. Hamaguchi, and M. Sakaizumi, "Identification of genetic sex of the medaka by PCR," *The Fish Biology Journal Medaka*, vol. 10, pp. 31–32, 1999.
- [18] M. Naimuddin, T. Kurazono, Y. Zhang, T. Watanabe, M. Yamaguchi, and K. Nishigaki, "Species-identification dots: a potent tool for developing genome microbiology," *Gene*, vol. 261, no. 2, pp. 243–250, 2000.
- [19] R. M. Myers, S. G. Fischer, L. S. Lerman, and T. Maniatis, "Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis," *Nucleic Acids Research*, vol. 13, no. 9, pp. 3131–3145, 1985.
- [20] K. Nishigaki, Y. Husimi, M. Masuda, K. Kaneko, and T. Tanaka, "Strand dissociation and cooperative melting of double-stranded DNAs detected by denaturant gradient gel electrophoresis," *Journal of Biochemistry*, vol. 95, no. 3, pp. 627–635, 1984.
- [21] M. Biyani and K. Nishigaki, "Hundredfold productivity of genome analysis by introduction of microtemperature-gradient gel electrophoresis," *Electrophoresis*, vol. 22, no. 1, pp. 23–28, 2001.
- [22] J. H. Ward Jr., "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.
- [23] J. D. Jobson, *Applied Multivariate Data Analysis*, vol. 2 of *Categorical and Multivariate Methods*, Springer, New York, NY, USA, 1992.
- [24] Y. Sakuma and K. Nishigaki, "Computer prediction of general PCR products based on dynamical solution structures of DNA," *Journal of Biochemistry*, vol. 116, no. 4, pp. 736–741, 1994.
- [25] D. Riesner, G. Steger, U. Wiese, M. Wulfert, M. Heibey, and K. Henco, "Temperature-gradient gel electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction," *Electrophoresis*, vol. 13, no. 9–10, pp. 632–636, 1992.
- [26] M. Naimuddin, T. Kurazono, and K. Nishigaki, "Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution," *Nucleic Acids Research*, vol. 30, no. 10, p. e42, 2002.
- [27] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [28] K. Nishigaki, N. Amano, and T. Takasawa, "DNA profiling. An approach of systemic characterization, classification and comparison of genomic DNAs," *Chemistry Letters*, vol. 20, no. 7, pp. 1097–1100, 1991.
- [29] H. Shimodaira, "An approximately unbiased test of phylogenetic tree selection," *Systematic Biology*, vol. 51, no. 3, pp. 492–508, 2002.
- [30] K. Nishigaki, A. Saito, T. Hasegawa, and M. Naimuddin, "Whole genome sequence-enabled prediction of sequences performed for random PCR products of *Escherichia coli*," *Nucleic Acids Research*, vol. 28, no. 9, pp. 1879–1884, 2000.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

