Quantitative comparison of fungi genomes performed by μTGGE genome profiling

Naoji Tamura, Hideo Fujino, Mariko Kouduka and Koichi Nishigaki

埼玉大学紀要　工学部　第37号　2004
抜　刷

Quantitative comparison of fungi genomes performed by μTGGE genome profiling

Naoji Tamura[1], Hideo Fujino[2], Mariko Kouduka[3] and Koichi Nishigaki[3*]

Owing to diversities of phenotypes, fungi are one of the most intimately studied organisms and very popular among all. Micro-TGGE genome profiling, a newly introduced approach for genome-based identification of species which random-samples portions of genome DNA and extracts sequence-derived information by μTGGE was applied to provisional identification of fungi and shown to be convenient, effective and reproducible. Quantitative analysis of the similarity between species in terms of PaSS (pattern similarity score) revealed that fungi were, unexpectedly, closely related with each other when viewed from the genome sequence. This must be the first result, for all organisms, obtained by quantitatively comparing genomes among more than 10 species at the order taxon level. It was discussed that since genome profiling does not require an immediate intervention of experts, this approach will allow general scientists as well as well-trained experts contribute to the progress of science through accumulating novel findings on fungi and other organisms with an aid of genome profiles.

Key words: Fungi, genome profiling, μ-TGGE, species-identification, genotype

1. Introduction

Fungi have been familiar with mankind for long and have been well-utilized for both nutritional and medicinal purposes. The taxonomy on fungi has most developed among all of species based on phenotypes owing to their readiness in collecting, storing, transferring and comparing. However, discrimination between fungi of similar appearances seems not to be an easy task even for experts in this field. Therefore, genotype-based identification of species, though in a limited scale, has been adopted on the analysis of 16S rRNA as had been done with bacteria and archaebacteria[1]. This approach has been developed to sustain a database of more than 70,000 species[2]. As it is based on the sequence information, it inevitably requires such experimental steps as cloning of 16S rDNA and sequencing of it, which are, by no means, easy to perform even in this mass sequencing era. In addition, the amount of information provided by sequencing of 16S rRNA is often not sufficient. In this context, genome profiling (GP) has been developed as an effective measure[3]. GP has a favorable property of providing a sufficient amount of information with ease. It enables us to identify species and to measure the distance between species[4]. The latter feature is very useful for finding a possible taxonomical position of species and provides us with a tool to review the taxonomical system established based on phenotype

[1] Kuki Technical High-School, Noguki 474, Kuki City, Saitama 346-0002, Japan

[2] Saitama Kinoko Association

[3] Department of Functional Materials Science, Saitama University. 255 Shimo-Okubo Sakura-Ku Saitama , Saitama 338-8570, Japan

only. We applied this novel, convenient and intriguing approach, for the first time, to identification and classification of fungi and obtained an affirmative result. Fungi were shown to be the species that can be easily analyzed by GP, suggesting wide applicability to the microbial world.

2. Materials and Methods

2.1 Fungi DNA preparation

DNA was prepared by alkaline extraction method, which has been shown to be, in principle, applicable to all kinds of organisms[5,6]. The procedures adopted here are as follow: i) A portion (10 mg) of the cap of a fungus was taken, washed in distilled water and then transferred into an Eppendorf tube. ii) After adding 500 μl of distilled water to it, the sample was kept boiling for 5 min. iii) The water was replaced with 50 μl of 0.5 M NaOH and then, the sample in alkaline was transferred to a microhomogenizer vessel and smashed. iv) The suspension was spun at 15,000 rpm for 1min, and 30 μl of the resultant supernatant was transferred to another fresh tube, added with 370 μl of 200 mM Tris-HCl (pH 8.0), and then mixed well. All fungi used for experiments were collected in forests and skirts of mountains in Kantou and Touhoku districts of Japan and morphologically identified by a well-trained scientist consulting fungi encyclopedia[7].

2.2 Micronized genome profiling

Micronized genome profiling is composed of two elementary procedures: random PCR and μTGGE[8-10]. Random PCR was performed using a single primer of dodeca-nucleotides (pfM12, dAGAACGCGCCTG) with 5' -end FITC-labeled(Fig.1 (a)). This sequence of the primer is strongly recommended for wide use by the authors developed this methodology based on so-far-performed experimental data[11]. The PCR reaction buffer of 100 μl contained template DNAs of around 10^{-17} M and 200 mM dNTP (N=G, A, T, C), 0.5 mM primer (pfM12), 2 unit of Taq DNA polymerase (Biotech International, Ltd., Perth, WA), 25mM $MgCl_2$, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1 % Triton X-100. PCR was carried out with cycles of denaturation (94℃, 30 s), annealing (25℃, 2 min), and extension (47℃, 2 min) using a PCR machine such as PTC-100TM (MJ Research, Inc.,

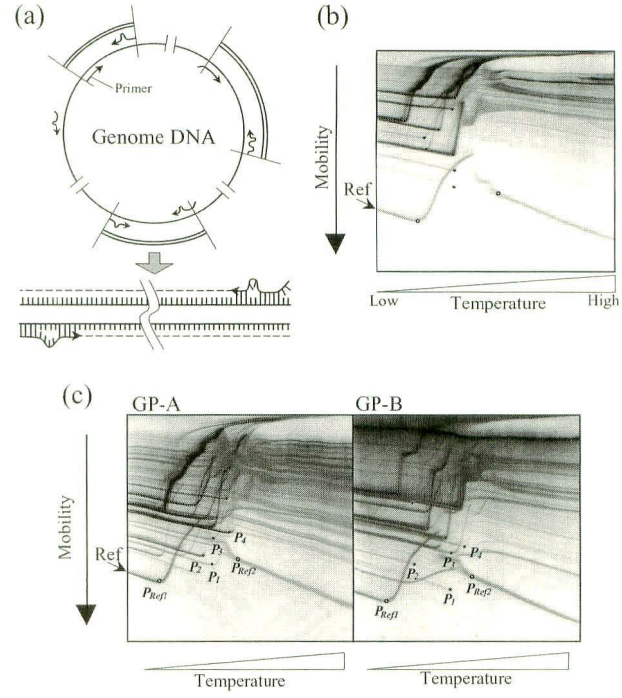

Fig. 1. Genome profiling. Genome profiling is composed of random PCR (a) and temperature gradient gel electrophoresis (TGGE) (b). (a) Mismatch-containing binding of primers leads to generation of DNA fragments from various sites of a template DNA. All of the genome DNAs such as chromosomes are shown circularly aligned. (b) DNA fragments charged at the top in line migrate down in the gel set with a linear temperature gradient (perpendicularly to the migration direction). Internal reference (Ref) and spiddos (small circles) are indicated. (c) Two genome profiles (GP-A and B) are put together to show corresponding spiddos used for calculation of PaSS. When species are closely related, corresponding spiddos are easy to be assigned as can be seen here. For less related species, mathematical pairing-optimization is adopted, which provides us with a unique solution for the pairing problem. The pictures used are taken from Ref 20 with permission.

Massachusetts). Random PCR products were mixed with 1 ml of the internal reference DNA and 0.5 ml of a buffer solution containing 0.3 % phenol red and 0.3 % XcFF and subjected to micro-temperature gradient gel electrophoresis (μTGGE) (30℃ — 70℃) (Fig.1 (b)). Micro-TGGE[8] was performed using μ-TG (Taitec Co., Ltd., Saitama). Gels were composed of 6 % acrylamide (acrylamide:bis=19:1) containing TBE buffer (90 mM Tris-HCl, 90 mM boric acid, 2 mM EDTA) and 8 M urea. The gel size was $24 \times 16 \times 1$ mm^3 and usually precast and stored in refrigerator for weeks. DNA bands were detected by fluorescence using Molecular Imager FX (Biorad, Hercules) or by silver staining[12,13].

2.3 Computation of PaSS

Genome profiles obtained by GP (Fig.1 (c)) are rather less manageable due to their complexity. However, it was explored that the featuring points expressed there, designated as *spiddos* (species identification dots), can be used as representative ones for genome profiles[4] through normalization realized by adopting an internal reference. It provides us with a unique, reproducible, and sufficient amount of information for identifying species[10]. Briefly, we describe below how to measure a similarity between genome profiles by terms of PaSS (Pattern Similarity Score). Those featuring points (termed spiddos) which appear on a genome profile are known to correspond to structural transitions of double-stranded DNA such as partial melting of DNA[14,15] and can be extracted unambiguously. Using Formula 1, we can calculate PaSS values.

$$PaSS = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\overrightarrow{Pi} - \overrightarrow{Pi'}\right|}{\left|\overrightarrow{Pi}\right| + \left|\overrightarrow{Pi'}\right|} \quad (0 \leqq PaSS \leqq 1) \qquad (1)$$

Where $\overrightarrow{Pi}$ and $\overrightarrow{Pi'}$ correspond to the normalized positional vectors (composed of two elements, mobility and melting temperature) for spiddos Pi and Pi' collected from two genome profiles (distinguished with or without a prime), respectively, and i denotes the serial number of a spiddo. A database site for performing necessary calculations automatically has been constructed on Internet (On-web GP, http://gp.fms.saitama-u.ac.jp/)[16].

3. Results and Discussion

3.1 Species-identification and reproducibility

Fungi have been identified based on their phenotype. Owing to their characteristic shapes and colors, species-identification of fungi must have been relatively easy in comparison to that of the other organisms. However, morphology-based identification inevitably requires the intervention of experts in the relevant field, inhibits a rapid progress of Science by excluding the contribution of non-experts and poses the experts with a series of laborious procedures for definite final identification. Our approach applied to this field may be useful for reducing such procedures and increasing contributive players. Therefore, it was necessary to examine whether genome profiling can be applied to fungi at large and can generate a sufficient amount of information at a high cost-performance with ease, which is by no means a mater of fact.

The wild fungi collected in mountainous regions of Japan were subjected to genome profiling (GP) following the method written in Materials and Methods. Micronized genome profiles obtained from PCR products with a primer (pfM12) are shown in Fig.2. The reproducibility of this methodology can be evidently seen in Panels *7a/7b* or *12a/12b* in Fig.2. As a reference, genome profiles of relatively distant species (yeast) from the fungi dealt here are also shown (Panels *12a* and *12b*). Each species was examined twice independently, beginning with the first step of DNA preparation using another portion of the cap of each fungus. The genome profiles thus obtained were processed to more quantitative data, *spiddos* (dots observed in Fig.2) and PaSS (pattern similarity score) as described in Materials and Methods. These results are

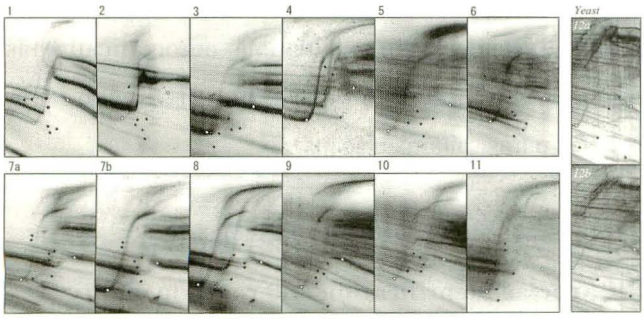

Fig. 2. Genome profiles obtained for the fungi tested. Panels 1-11: 1, *Cortinarius pseudosalor* J.Lange; 2, *Cortinarius armillatus* Fr.; 3, *Amanita vaginata* var. fulva Gill.; 4, new species; 5, *Hygrophorus virgineus* Fr.; 6, *Hygrophorus russula* (Fr.) Quel.; 7a and 7b, *Suillus cavipes* (Opat.) Smith (The same speacies were tested for reproducibility); 8, *Hericium erinaceum*(Fr.)Pres.; 9, *Lactarius chrysorheus* Fr.; 10, *Polyporus brumalis* Fr.; 11, Boletopsis leucomelas (Fr.)Fayod. Panels 12a and 12b are genome profiles for yeast, which is a kind of fungi but is distant from those species shown in Panels 1－11. Spiddos are shown with dots for all genome profiles.

schematically summarized in Fig. 3. Evidently, each self-PaSS (*i.e.*, the similarity between two GP trials of the same individual, shown in a blank box on diagonal) is significantly higher than any other PaSS values which were obtained by pairing of different species. The average value of the self-PaSS was 0.979 ± 0.013, indicating that this experimental system was well-controlled within a permissible error range (empirically, it must be higher than 0.95 for the identical species[4], meaning that more than 5% difference in PaSS from the complete match (PaSS=1) proves to be non-identity.). All the PaSS between different entries were lower than the relevant self-PaSS, ranging from 0.806 to 0.920 (yet they are sufficiently higher than the value expected for a pair of no relation, 0.667, with a standard deviation of 0.091)[4].

Therefore, the genome profiling method can be adopted to provisionally identify species belonging to fungi and to cluster and classify fungi based on the PaSS value.

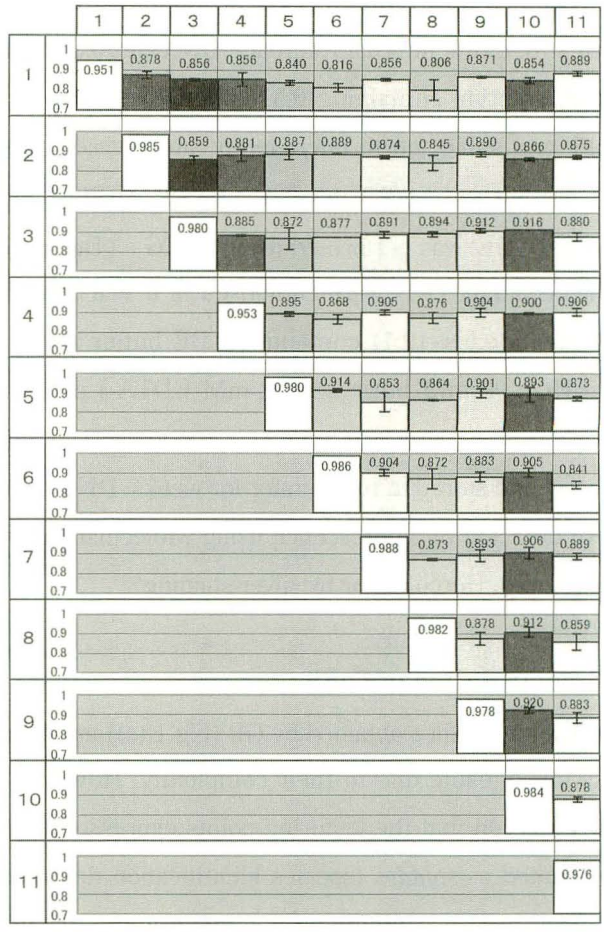

Fig. 3. Schematic representation of PaSS values between pairs of fungi. The same numbering is used as in Fig. 2. The PaSS value is shown by the height of a bar together with the numerical value. Each species was examined twice, resulting in relatively high reproducibility (see error bars). Self-PaSS is shown on diagonal.

3.2 Taxonomy and PaSS

In Fig. 4, the conventional classification of fungi is drawn following the nomenclature adopted by Kirk et al.[17] together with photos of each sample (collected in this study) and PaSS values for a pair of neighboring identities at each level of taxonomy (*order*, *family*, *genus* and *species*). Unexpectedly, there were very little differences among the PaSS values covering several levels of taxonomy; *e.g.*, 0.881 for the pair of *Agaricales* (F.1 in the figure) and *Boletades* (F.2), and 0.878 for *Cortinarius pseudosalor* J. Lange (F.1.1.1.a) and *Cortinarius armillatus*, Fr. (F.1.1.1.b). This fact is very

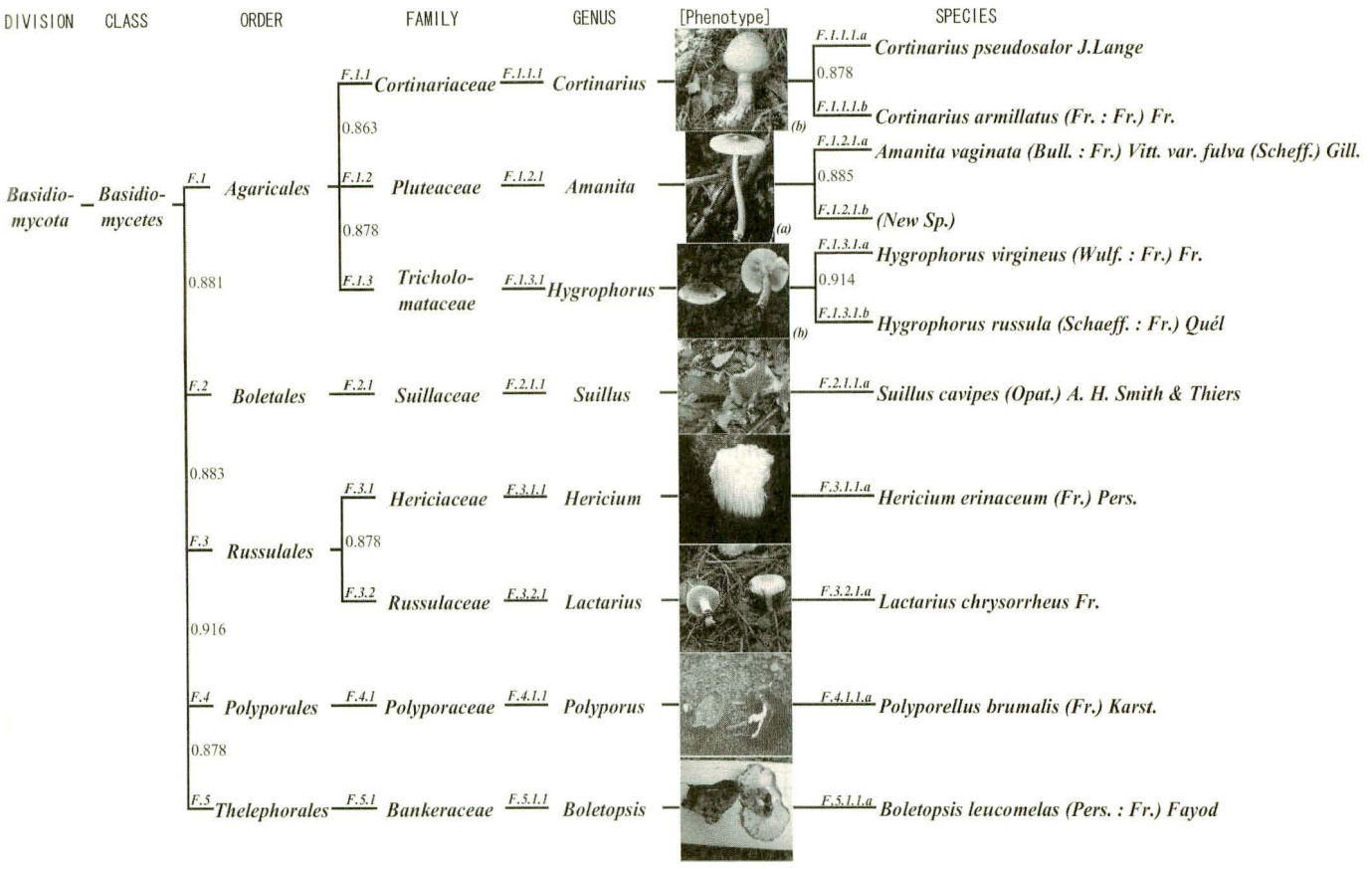

Fig. 4. A taxonomical relationship of fungi examined. Fungi examined here are arranged following the traditional, phenotype-based taxonomy. The pictures of fungi used are also shown. The number beside each branch designates PaSS value for two individuals separated at the corresponding level of taxon.

interesting since from a taxonomical viewpoint it suggests that each species is *equivalently distant* from each other at the genome level, irrespectively of phenotypic classification. So far as we know about genome profiling, this is a unique nature of fungi as in the other domains of organisms such as bacteria, plants and animals a pair of species belonging to different '*orders*' are more distant (lower in PaSS value) than those belonging to the same genus (to be published elsewhere). One possible rationale for this fact may be to suppose that there have been occurring frequently horizontal transfers of genes or recombination among the species of fungi. If so, the genetic constituents of fungi will be mingled and become more homologous (similar but different). Naturally, this hypothesis can be verified by the whole genome sequencing, which will be done in the future. Hereupon, GP method has potential to do the equivalent. GP can extract various sets of DNA fragments from genome DNAs by altering the *probe* sequence (*i.e.*, the primer sequence used for random PCR), although in this study, a single standard primer (pfM12) was used, and still sufficed the current aim. Theoretically, it can be shown that infinitely increasing the number of probes for random PCR leads to extracting the whole sequence of the genome DNA of interest. Therefore, if it occurs that the fact found here needs to be confirmed, one easy way must be to advance GP analysis by changing probes.

Another possible interpretation for the closeness in PaSS values among a wide range of species of fungi may come from the realization that there is no quantitative measure involved in the hierarchy of taxonomy.

Division and clustering might have been rather arbitrarily made, though consistent among peers, based on phenotypic traits without a quantitative measure. Naturally, the density of taxonomical divisions can not be constant throughout all the organisms, denoting that a group of organisms may be divided into too many taxa irrespective of genome distances among those organisms (though unknowable even now). Now we can define genome differences, though not easy to calculate, based on genome sequences by employing a similar concept with Hamming distance (*i.e.*, difference in letters along the sequence). In other words, this apparent discrepancy results from the fact that the phenotypic approach is qualitative while genome-based one is quantitative. It is noteworthy that the quantitative nature is endowed at the cost of loss of meanings. This type of discussion must have been done first of all with this group of fungi.

Since among microbial species the horizontal transfer of genes is so frequent[18,19], the boundary of species is very ambiguous. Therefore, a mapping technology-free GP must be useful for microbial studies.

3.3 Methodological features

Genome profiling has been developed for more than ten years since its invention[9,10]. The feature of genome profiling (GP) has recently been argued in comparison with the current technologies for genome analysis such as DNA micro-array, AFLP, 16S/18S rDNA sequencing and whole genome sequencing[3]. Evidently, 16S/18S rDNA approach has been most popular up to now and has accumulated a large amount of data on the platform of public domain (more than 70,000 species[2]. However, as discussed in the review[3], 16S/18S rDNA approach is not always universal nor effective for genotype-based identification of species. Obviously, it is more laborious in collecting data from various unknown species than GP since it has to specifically amplify 16S rDNA prior to the knowledge on the individual sequence, making it difficult to design appropriate primers (Sequencing itself requires more costly apparatus and more time-consuming processes). The most essential advantage of GP over 16S/18S rDNA approach is in the fact that the amount of information is not limited to a single sequence as in 16S rDNA approach and that the information on the construct of genomes (i.e., of which genes a genome consists) is also obtainable (Note that random PCR collects DNA sequences from various parts of genome DNA). The other methods listed above are yet to be used for genome-based identification of species. In this sense, GP is currently the most promising method to identify species, which has been preliminarily demonstrated with fungi in this paper.

Application of GP to fungi is rather easier than to any other organisms due to the nature of fungi that its fruit body can be readily picked, stocked and used for preparation of its DNA. The most excellent property of GP approach is that one can register an unidentified species by terms of genome profiles without losing novel phenotypes observed. A species deposited on a database will be automatically linked via GP with the species identified by an expert later. This is because the computer can recognize the identity of species by means of PaSS value. Then, what is important, the properties deposited will be revived and utilized even if registered without identification of species. Therefore, GP will help ordinary scientists to contribute to taxonomy.

It is, of-course, neither to say that fungi can be definitely identified without the knowledge on phenotypes nor, to neglect the importance of exploring phenotypic traits of organisms. Since a phenotype may be changed depending on a single gene introduced by plasmids, viruses or other movable elements, GP can not always be competent in discriminating such a difference. Edible-or-not should be decided based on individual phenotypic properties, not on GP. With all this fact, GP is, to some extent, beneficial in telling

poisonous mushrooms if the probe for random PCR is carefully selected to pick up the DNA region responsible for the poison.

Versatile natures of fungi are very attractive. Owing to these natures, taxonomy on fungi has been well developed. Now, it is the time when genome information is utilized for identifying species, which is especially the case with microbes.

4. Acknowledgements

This study was performed during the long-term leave of one of the authors (N.T.) provided by the Committee for High School Education of Saitama Prefecture.

References

1. Gutell, R.R., Larsen, N. and Woese, C.R. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. 58: 10-26.
2. Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker CT Jr, Saxman PR, Farris RJ, Garrity GM, Olsen, G.J., Schmidt, T.M., Tiedje, J.M. 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29: 173-174.
3. Naimuddin, M. and Nishigaki, K. 2002. Genome analysis technologies towards species identification by genotype. Briefings in Functional Genomics & Proteomics 1: 1-16.
4. Naimuddin, M., Kurazono, T., Zhang, Y., Watanabe, Y., Ymaguchi, M. and Nishigaki, K. 2000. Species identification dots: A potent tool for developing Genome Microbiology. Gene 261: 243-250.
5. Hamano, K., Takasawa, T., Kurazono, T., Okuyama, Y. and Nishigaki, K. 1996. Genome Profiling-Establishment and practical evaluation of its methodology. Nikkashi. 1996: 54-61 (In Japanese).
6. Wang, H., Qin, M., Cutler, A.J. 1993. A simple method of preparing plant samples for PCR. Nucleic Acids Res. 21: 4153-4154.
7. Imazeki, R., Hongo, T. 1986. Genshoku-Nippon-Shinkinrui-Zukan (Neo-encyclopedia of Japan-homed Fungi, Hoikusha, Osaka, Japan (In Japanese).
8. Biyani, M., Nishigaki, K. 2000. Hundredfold productivity of genome analysis by introduction of microtemperature-gradient gel electrophoresis. Electrophoresis 22: 23-28.
9. Nishigaki, K., Amano, N., Takasawa, T. 1991. DNA Profiling: An approach of systematic characterization, classification, and comparison of genomic DNAs. Chemi. Lett. 1991: 1097-1100.
10. Nishigaki, K., Amano, N., Takasawa, T., Kinoshita, Y. and Husimi, Y. 1990. DNA profiling- method and principle, Seibutu-buturi. 30: S230 (In Japanese).
11. Nishigaki, K., Saito. A. 2002. Genome structures embossed by oligonucleotide-stickiness. Bioinformatics 18: 1153-1161.
12. Boulikas T. Hancock R. 1981. A highly sensitive technique for staining DNA and RNA in polyacrylamide gels using silver. J. Biochem. Biophys Methods 5: 219-228.
13. Ohsawa, K., Ebata, N. 1983. Silver stain for detecting 10-femtogram quantities of protein after polyacrylamide gel electrophoresis. Anal Biochem. 135: 409-15.
14. Nishigaki, K., Husimi, Y., Masuda, M., Kaneko, K., Tanaka, T. 1984. Strand dissociation and cooperative melting of double-stranded DNAs detected by denaturant gradient gel electrophoresis. J. Biochem. 95: 627-635.
15. Wada, A., Yabuki, S., Husimi, Y. 1980. Fine structure in the thermal denaturation of DNA: high temperature-resolution spectrophotometric studies. CRC Crit. Rev. Biochem. 9: 87-144.
16. Watanabe, T., Saito, A., Takeuchi, Y., Naimuddin, M. and Nishigaki, K. 2002. A database for the provisional identification of species using only genotypes: web-based genome profiling. Genome Biology 3: Research 0010.1.

17. Kirk, P.M., Cannon, P.F., David, J.C. and Stalpers, J.A. 2001. Ainsworth & Bisby' s Dictionary of the Fungi, CAB International, CABI Bioscience, New York, USA.
18. Hensel, R., Zwickl, P., Fabry, S., Lang, J., Palm, P. 1989. Sequence comparison of glyceraldehyde-3-phosphate dehydrogenases from the three urkingdoms: evolutionary implication. Can. J. Microbiol. 35: 81-85.
19. Lake, J.A., Jain, R., Rivera, M.C. 1999. Mix and match in the tree of life. Science 283: 2027-2028.
20. Naimuddin, M., Kurazono, T. and Nishigaki, K. 2002. Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution. Nucleic Acids Res. 30:e42.