Genome profiling (GP) as an effective tool for monitoring culture collections: A case study with *Trichosporon*

Keiichi Hamano^{a,b}, Sachika Ueno-Tsuji^{a,c}, Reiko Tanaka^d, Motofumi Suzuki^e,
Kazuko Nishimura^d, and Koichi Nishigaki^{c*}

^aRational Evolutionary Design of Advanced Biomolecules (REDS), Saitama
Small Enterprise Promotion Corporation, SKIP city, 3-12-18 Kamiaoki,
Kawaguchi, Saitama 333-0844, Japan

^bTAITEC Corporation 2693-1, Nishikata, Koshigaya, Saitama, 343-0822,
Japan

^cDepartment of Functional Materials Science, Saitama University, 255, Shimo-
Okubo, Sakura-ku, Saitama 338-8570, Japan

^dMedical Mycology Research Center, Chiba University, Chiba 260-8673,
Japan

^eJapan Collection of Microorganisms, RIKEN BioResource Center, Wako,
Saitama 351-0198, Japan

*Corresponding author

Email addresses:

KH: nyoihoju@w5.dion.ne.jp

SUT: sachikat@mail.saitama-u.ac.jp

RT: moru@faculty.chiba-u.jp

MS: msuzuki@jcm.riken.jp

KaN: nishik@faculty.chiba-u.jp

KoN: koichi@fms.saitama-u.ac.jp

Abstract

Species identification and classification of a large number of microbes are essential and heavy workloads in culture collections and relevant laboratories. The identification of species usually requires different methods depending on species. Therefore, the development of a method which is simple and applicable to any organisms will lessen the burdens, increase the reliability of databases and thus enhance the science on microbes. The genome profiling (GP) method, developed previously, was found effective in monitoring authenticities of all strains/species tested in culture collections and expectedly various species, which was shown by applying the GP and the conventional sequencing methods to identifying and classifying species/strains belonging to the genus *Trichosporon* (38 strains; 16 species). Small differences between strains (11 strains of *T. asahii* and 4 strains of *T. coremiiforme*) can be reliably discriminated by GP, which was unsuccessful in the conventional sequencing approach. Importantly, seven possible false-assignments contained in the database were all pointed out by the GP method with near-perfect correctness, showing the power of the GP method. GP was shown to be a potent tool for rapidly and correctly monitoring species and strains of fungi in culture collections owing to its simple and informative natures.

Keywords

Genome Profiling

Trichosporon

Genome Distance

Species Identification

TGGE

1. Introduction

The role of culture collections is to provide the substantial basis upon which all of the activities in biological fields are performed. It requires incessant collection, maintenance, and distribution of various species with certification. However, the process of confirmation of species identity is laborious, time-consuming and costly. The current method of IGS and/or ITS sequencing is often insufficient to discriminate between intraspecies strain differences, as can be seen in the case with the genus of *Trichosporon* ([Rodriguez-Tudela et al., 2005](#)). Those species and strains had been classified depending on their morphological and biochemical properties; the shape of segregating cells and the amount of xylose contained in a cell. Recently, the number of species belonging to the genus *Trichosporon* has increased from 25 up to 33 ([Fell and Scorzetti, 2004](#); [Middelhoven et al., 2000](#); [Middelhoven et al., 2004](#); [Molnar et al., 2004](#); [Sugita and Nakase, 1998](#); [Sugita et al., 2001](#)), requiring re-definition of species and thus making culture collection centers busy in re-labeling and re-confirming identities. Therefore, the performance of re-confirmation is, essentially, important and busy for culture collection centers, necessarily requiring effective methods. So far, several approaches have been developed based on the genomic DNA for this purpose: PFGE (pulsed-field gel electrophoresis) ([Schwartz and Cantor, 1984](#)), RAPD (random amplified polymorphic DNA) ([Sugita et al., 2001](#)), AFLP (amplified

fragment length polymorphism) (Wolf et al., 2001), 18S rDNA sequencing (Takashima and Nakase, 1999), and others (Naimuddin and Nishigaki, 2003). Among these, PFGE, RAPD, and AFLP utilize the size information of DNA fragments obtained, while the 18S rDNA sequencing method depends on the sequence information. Unfortunately, none of these methods can escape from the information shortage that leads to a failure in correct identification and classification of species. Although the sequencing itself is determinate and reliable, the amount of information provided by a single gene sequencing is limited and often insufficient. Thus, it is less favorable in cost-performance and not infallible in classifying species (Fitz-Gibbon and House, 1999; Koonin and Galperin, 1997; Snel et al., 1999). In the case of fungi, the D1/D2 domain of 26S rDNA and ITS (internal transcribed spacer) sequences have been used frequently for identification of species. It is known that the dissimilarity between species in these sequences is within 1% for most clinical fungi (Rodriguez-Tudela et al., 2005; Sugita et al., 1999). Naturally, the similarity between strains in a species is much higher and far more difficult to discriminate. Regarding this fact, Sugita *et al.* reported that the IGS1 (intergenic spacer) sequence of *Trichosporon* was a better indicator of the diversity of species than the ITS, dividing species of *Trichosporon asahii* into seven genotypes (Sugita et al., 2004). In this sense, the sequencing approach has a potential to discriminate strains although it requires analyses of species-dependent particular sequences.

In this context, a simple and universal method to identify and classify species has been developed: the genome profiling (GP) method (Ahmed et al., 2011; Hamano et al., 1996; Kouduka et al., 2007; Naimuddin et al., 2000; Nishigaki et al., 2000; Watanabe et al., 2002) which is mainly based on the statistical concept of random sampling and the rapid acquisition of sequence information. GP can be highly powerful and reproducible in species identification due to i) reproducible sampling of DNA fragments from the original genomic DNA by random PCR, which can employ a defined set of primers for any kind of organisms, ii) acquisition of the sequence-

derived information without sequencing by using μ TGGE (temperature gradient gel electrophoresis) to which random PCR products can be directly applied without purification of DNA fragments, and iii) normalization of the acquired data using internal references, which leads to obtaining species-specific data. The effectiveness of the GP method has been demonstrated by applying it to i) the ability to identify species (Hatakeyama et al., 2009; Naimuddin et al., 2000), ii) the ability to classify organisms (Kouduka et al., 2007), and iii) the ability to measure the extent of mutations (Futakami and Nishigaki, 2007; Futakami et al., 2007). However, it has not been clarified whether GP can solve the practical problem common to culture collection centers: the need to monitor the diversity of species and strains with sufficient accuracy, at an affordable cost and with limited man-power. We selected the *Trichosporon* species (38 strains out of 16 species) as test samples for demonstrating the effectiveness of the GP-based checking of microbial collections. The results were highly satisfactory and the reason for this success is discussed.

2. Materials and Methods

2.1. Strains of *Trichosporon* and *Candida*

The fungi examined in this study were 38 strains from 16 species of the genus *Trichosporon* and the strains *Candida albicans* JCM 1542^T and *Candida tropicalis* JCM 1541^T (Table 1). The strains *Trichosporon* and *Candida* were provided by the Medical Mycology Research Center, Chiba University (Chiba, Japan), and Japan Collection of Microorganisms, RIKEN BioResource Center (Wako, Japan), respectively.

2.2. DNA extraction

Genomic DNA was extracted at the collection center using a commercially available DNA extraction kit (for yeast), Dr. GenTLE™ (TaKaRa Bio Inc., Japan) following the vendor's protocol (*i.e.*, a pellet of fungal cells was resuspended in the GenTLE solution and incubated at 70°C for 10 min.)

2.3. Genome profiling (GP)

This method consists of three main steps; i) random PCR, ii) μ TGGE, and iii) computer-aided data processing (Fig. 1).

2.3.1. Random PCR

The random PCR can employ an arbitrary primer for the PCR reaction due to the relaxed mode of primer binding to the template DNA performed under sufficiently low temperatures (Nishigaki et al., 1990; Sakuma and Nishigaki, 1994). This nature enables us to do PCR with those samples using as templates which genome sequences are quite unknown. Thus, it enables us to collect random DNA fragments from various portions of genomic DNA depending on the sequence of the oligonucleotide used (Nishigaki and Saito et al., 2000). This is basically the same with those techniques termed as RAPD (Williams et al., 1990) and ap-PCR (Welsh and McClelland, 1990) devised concurrently and independently (Nishigaki et al., 1990). We used seven primers (probes) for this analysis after preliminary probe match experiments regarding 14 probes: pfM12 (5'-Cy3-labeled dAGAACGCGCCTG), pfM19 (5'-Cy3-dCAGGGCGCGTAC), Hunt (5'-Cy3-dTGCTGCTGCTGC), B81 (5'-Cy3-d-dGGCCGACTTGGC), B83 (5'-Cy3-dCAGGCCGAAGTC), B87 (5'-Cy3-dTATCCACCGCTC), and B89 (5'-Cy3-dACTAACCTGGAC). The PCR solution (30 μ l) contained 200 μ M dNTP, 0.5 μ M primer, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 1 unit of rTaq DNA polymerase (TaKaRa, Japan) and 4-10 ng of a template DNA. Random PCR was

performed using a thermal cycler, Mastercycler gradient (Eppendorf, Germany) with an initial denaturation at 95°C for 2 min, followed by 30 cycles of 15 s incubation at 94°C, 30 s at 40°C, and 30 s at 50°C, and a final treatment of 30 s at 72°C.

2.3.2. μ TGGE (temperature gradient gel electrophoresis) analysis

Temperature gradient gel electrophoresis (TGGE) (Thatcher and Hodson, 1981) enables us to measure the DNA sequence-inherent melting temperature (Nishigaki and Naimuddin et al., 2000), for which the device can be miniaturized to a size of square inches without losing accuracy (Biyani et al., 2001; Tominaga, 2006). In this study, a μ TGGE apparatus, μ -TG (TAITEC, Japan) was used with a specifically fabricated gel cassette. The gel used was a conventional one consisting of 6% (w/v) acrylamide and bis-acrylamide (19 : 1), 6.5 M urea dissolved in 1 $\times$ TBE buffer (0.1 M Tris, 0.09 M boric acid, 0.001 M EDTA [pH 8.4]). The μ TGGE electrophoresis was performed under the conditions of 100V, 9 min duration and a temperature gradient of 15 to 55°C in 5 $\times$ TBE buffer using 9 μ l of a random PCR solution containing internal references (IntRefs. 5 and 6 DNAs), and triple dyes (Nippon Gene, Japan). After electrophoresis, bands were detected by a fluoroimager, Typhoon 9400 (GE Healthcare Bio-Science, USA). Every TGGE experiment has been done twice, confirming the quality of experimental data (omitting those of an apparently low quality, i.e., gels of a scanty number of bands or distorted gel or so; the frequency of such failures is less than 5 % according to our experience). The internal reference DNAs (IntRefs. 5 and 6) were prepared by PCR from the plasmid pBR322 DNA (DNA regions: 3501-3700 and 401-1300, respectively). Fluorescently-labeled primers Ref5F (5'-Cy3- dAGTGGTCCTGCAACTTTATC) and Ref5R (dAACATGGGGGATCATGTAAC), Ref6F (5'-Cy3- dGCCGGCATCACCGGCGCCACAGGTGCGGTTG) and Ref6R (dTAGCGAGGTGCCGCGGCTTCCATTTCAGGTC) were used to generate IntRefs.

5 and 6, respectively. A conventional PCR program was used: initial denaturation at 95°C for 2 min, followed by 25 cycles of 15 s at 94°C, 30 s at 55°C, and 30 s at 72°C, and a final processing of 30 s at 72°C.

2.3.3. Computer-aided data processing

In order to obtain reproducible and objective data, the normalization process using internal references aided by a computer was important and effective. A set of featured points, corresponding to the initial DNA melting temperatures of each double-stranded DNA, were manually collected by assigning each genome profile shown on the computer display. These points were processed to obtain the normalized mobility and temperature characteristics, termed *spiddos* (species identification dots) (Naimuddin et al., 2000) (Fig. 1, s-Figure 1(Supplemental)). A measure of similarity between two genomes, i.e., the pattern similarity score (*PaSS*), was calculated as follows:

$$PaSS = 1 - \frac{1}{n} \sum_{i=1}^n \frac{|\bar{p}_i^{(1)} - \bar{p}_i^{(2)}|}{|\bar{p}_i^{(1)}| + |\bar{p}_i^{(2)}|} \quad (1)$$

where $\bar{p}$ of each *spiddos* (1 to n) designates its position vector and is a function of temperature and mobility (that is, $|\bar{p}| = P(T, m)$). The superscripts 1 and 2 in parentheses in Equation (1) represent genomes 1 and 2, respectively. *PaSS* will be unity for a complete match in two sets of *spiddos*. In general, $0 < PaSS \leq 1$. In this measurement, specific coordinate differences between the corresponding *spiddos* are accumulated and subtracted from unity. *Spiddos* are especially important since the coordinate of each is a unique and intrinsic representation of the corresponding sequence due to the normalization process adopting the internal references (Naimuddin et al., 2000). This process ensures elimination of experimental

fluctuations such as instrument noise, inconsistencies in gel formation, ambient temperature changes and so on. Each sequence has unique *spiddos* of a coordinate composed of two variables, mobility and temperature, which depend on the sequences of both the primer and the template. Therefore, a set of *spiddos* are regarded as species-intrinsic points and can be obtained reproducibly at any time at any place so far as the same species is dealt with. Accordingly, *PaSS* (pattern similarity score) defined as in Eq. (1) is a universal value obtained for two genomes, which is larger than zero and smaller than or equal to one.

The deviation of the *PaSS* values between repeated trials in a year or more of interval and by different examiners is known to be able to be confined within 0.8 % or less.

2.4. Phylogenetic analysis

Phylogenetic trees were generated using the program DendroUPGMA (Garcia-Valve et al., 1999) with Pearson coefficients and the unweighted pair group method with arithmetic mean algorithm (<http://genomes.urv.cat/UPGMA/>), employing *PaSS* ($0 < PaSS \leq 1$) matrix. Graphical trees were drawn using the TreeView software (Roderic, 1996).

3. Results

3.1. Trichosporon species identified by the GP analysis

To represent the mutual relationships of *Trichosporon* species and strains (38 strains of 16 species), a phylogenetic tree was drawn based on the genome distance (Kouduka et al., 2006; Ahmed and Nishigaki, 2008), which is equivalent to the value of $1 - PaSS$ (pattern similarity score), obtained from the GP experiments using 7 independent probes with respect to all of the strains (Fig. 2; raw data in Fig. 3). All of

the disputable species are shown in colour: T-20, 12, 14, 13, 15, 16, 17, 35, 19, 34, 11, 36, 33, 18 & 21 (from top to bottom). T-20 was assigned as *T. cutaneum*. On the other hand, T-19, T-18 and T-21 were labelled as the same species (*T. cutaneum*) but were positioned separately from each other. Similarly, T-35 and T-36, which were assigned as *T. montevidense*, and T-34 and T-33, assigned as *T. moniliiforme*, were distantly mapped in this phylogenetic tree. The remaining T-17 and T-16, which are also located within these disputable species needed to be confirmed. However, at this point we could not determine which was wrong (or even that both were wrong). Conveniently, GP provides more information than that used for obtaining PaSS (the pattern similarity score used for drawing the phylogenetic tree which is calculated using *spiddos* (See s-Figure 1(Supplemental)): *i.e.*, the DNA mobility transition profile corresponding to its melting process induced by denaturant or temperature (Nishigaki et al., 1984). As shown in Fig. 4-A, the questionable T-20, which had been assigned as *T. cutaneum*, is obviously more similar to T-6 (*T. asahii*) than T-18 (*T. cutaneum*) especially in the DNA of bands a, b, and c.

Quite similar situations exist for T-35 (Fig. 4-B), T-19 (Fig. 4-C), and T-33 (Fig. 4-D), where all of the disputable species (especially, T-19 and T-33) have remarkably similar bands to those of their putative species (*e.g.*, bands d, e and f for T-17, T-16, T-35 and T-19; bands g, h and i for T-23 & T-33). Empirically, those bands that are highly similar in their pattern among different species were revealed as ccgf (commonly conserved genetic fragments originated from orthologs, paralogs, or like) in the preceding paper (Naimuddin et al., 2002), allowing us to provide those questionable species with the correct identification. As shown in Fig. 4-E, T-11 has a distinctly different band pattern from those of T-1 (*T. asahii*) and T-23 (*T. faecale*), supporting its naming (*T. asahii* var. *faecale*; a variant strain derived from *T. asahii* and has features of *T. faecale*). The ccgf band 'j' appeared in the genome profiles of T-13 and T-15 when a different probe (primer B89) that closely connects these organisms was used (Fig. 4-F). In this analysis, the genome distances (d_G) used are

the averages over those obtained with seven different probes (primers) (see Methods). The reliability of this phylogenetic tree was confirmed by examining the effect of adding outgroup species (*Candida* species) to the species of interest here (Fig. 2, s-Figure 2(Supplemental)). The available ccgf analysis could completely confirm the inferences derived from the phylogenetic tree as listed in Table 2. In order to confirm these results further, 26S rDNAs (D1/D2 region) of all 38 strains were sequenced. The species name was assigned based on the 26S rDNA (D1/D2) sequence by BLAST-searching over the sequence database of DDBJ (Table 2). By comparing the GP analysis with the sequence one, those species for which the authenticity of identification was in doubt (6 strains (T-11, -19, -20, -23, -33, and -35) and a group of 4 strains (T-12, -13, -14, and -15)) were all confirmed to be mis-assigned (except the pair of T-12 and T-14) (Table 2). On the other hand, the remaining 28 strains were all confirmed to be authentic by both the GP and the 26S rDNA sequencing analyses, meaning that the GP analysis is an excellent 'rapid-screener' of species identity.

3.2. Another information provided by the GP analysis

Importantly, a triangle distance map of Fig. 3 (*i.e.*, table of genome distance d_g), which is a different presentation of the data used above, can provide additional information as well as serving as confirmation. As shown in Fig. 3, the pairs of closest distance but separately located in the phylogenetic tree are mapped off the diagonal (white boxes such as pairs of T-20/T-6, T-20/T-8, T-33/T-23, T-35/T-17, and T-35/T-19). These results led to the inference that T-6, T-8, and T-20 belong to the same species (which was correct and finally assigned to *T. asahii* (see Table 2)). Similarly, the pair of T-23 and T-33 (*T. faecales*) and the trio of T-17, T-19, and T-35 (*T. coremiiforme*) did so also. These are accordant to the conclusion attained in Table 2. A novel fact is that T-4, T-13, T-18, and T-29 all appear rather extraneous in

this figure as underscored with a lining of blue line, assumingly indicating that these strains might have undergone frequent mutations and been driven away from the other strains of *Trichosporon* genus. In the same context, T-35 (finally assigned to be *T. coremiiforme*) seems to enter the central place among *Trichosporon* genus (at least with respect to the species investigated here, see Fig. 2) since it is rather close in the genome distance d_g to all of these species. Such interpretations deduced from the triangle distance map are hard to be obtained by other analyses and may be useful for managing and analyzing culture collections. Thus, such a triangle map provides us with the in-depth information as to the mutual closeness for all pairs of species. It is true that the effectiveness of the interpretations offered here needs to be verified by additional studies. What is evident is that the GP method can provide rich information on species identification ([Nishigaki and Naimuddin et al., 2000](#)).

4. Discussion

The genome profiling (GP)-based clustering of fungal species (*Trichosporon* sp.) could reveal incorrect labelling of those species in a culture collection with almost perfect accuracy (no false negatives or false positives if we regard that the 'grey judgment' on the pairs of T-13/T-15 and T-12/T-14, which was correctly inferred, also falls in this category), which was confirmed by 26S rDNA sequencing (Table 2). GP could provide not only spiddos (a set of normalized coordinates of featuring points) but also ccgf (commonly conserved genetic fragments)-type information ([Naimuddin et al., 2000](#); [Naimuddin et al., 2002](#)), which is useful for close comparison between closely related species (Fig. 4). Conveniently, the ccgf DNA fragments are good samples for the comparison of sequences due to the readiness in DNA preparation (PCR using the same primer adopted for the random PCR and a small portion of gel which contains the ccgf band. Conveniently, all of the random PCR products contain the same sequence with that of the primer used at both ends of a DNA fragment),

which can circumvent the efforts of designing and preparing PCR primers (Nishigaki and Naimuddin et al., 2000). Although we employed the conventional approach of sequencing D1/D2 26S rDNA from the constraint of demonstration paper, the ccgff sequencing approach will be much convenient for the routine work in the culture collection if the detailed analysis (*i.e.*, sequencing) is required regarding particular organisms.

In this experiment, seven different probes, which were empirically selected, were used to obtain genome distances. As a result, the selection of those probes was not only effective but also necessary for this particular genus. The number of probes (primers) used in this experiment is rather arbitrary. It is already theoretically and empirically shown that the larger the number of probes the more accurately we can understand the genome by the GP method. Therefore, we have tried to employ probes as many as possible. However, a single probe experiment (usually, the probe pfM12 is used but in this particular genus the primer named Hunt was best) was sufficient for consistently classifying species (Kouduka et al., 2007; Watanabe et al., 2002) (See s-Figure 2(Supplemental)), meaning that the informational content provided by a single experiment is already sufficiently high (Nishigaki and Naimuddin et al., 2000). Needless to say, the additional data provided by another probe experiment improves the classification result to some extent. These are statistically reasonable since the amount of information is nearly proportional to the reliability of experimental results. It is noteworthy that the effort of the experiments (7 experiments for each sample) is still smaller than that required for the sequencing experiment (See s-Table 1(Supplemental) and s-Table 2(Supplemental)). The appropriate number and species of probes to be used must be determined depending on the individual situations. Empirically, in most cases including the application to fungal species, a single- or two-probe experiment must be effective for identifying species correctly, thus saving significant effort.

A problem posed in large scale-culture collections and even small scale-ones like conventional microbiology laboratories is to perform routine checking experiments with sufficient reliability at an affordable cost and manpower. In a sense, cost-effectiveness is the top priority in these facilities. A tentative estimation was done regarding the efforts and costs of the GP and the sequencing approaches (See s-Table 1(Supplemental) and s-Table 2(Supplemental)), revealing that GP is around 5 times more cost-effective and roughly 1.7 fold more time-saving than the sequencing approach though such estimation depends on experimental environments. The reason why the GP approach is so cost-effective and rapid is simple: i) it does not require the laborious preparation of DNA to a sufficient level of purity. Instead, it only requires extracting crude DNA from cells, which can be directly used for random PCR and the products thus obtained can be subjected to the succeeding final experimental step of the temperature gradient gel electrophoresis (TGGE) without purification. ii) The TGGE analysis provides us with a much greater amount of information in a single experiment than that obtained in the conventional gel electrophoresis (Nishigaki and Naimuddin et al., 2000). The inclusion of the internal reference method (Naimuddin et al., 2000), which eliminates experimental fluctuations, was responsible for the generation of highly reliable and reproducible data.

In addition, another merit of the GP approach should be underscored: it can offer, effectively, a limitless amount of information just by changing the probe used for random PCR (the possible number of different probes (primers) made of dodeca-oligonucleotide is over 1.6×10^7). In this experiment, only seven probes out of such a huge number were used. If necessary, independent (orthogonal) information can be obtained up to a satisfactory level by employing the same approach with different probes. Therefore, it is unnecessary to sequence different genes such as cytochrome, oxidase, gyrase, different parts of rRNA genes by designing and synthesizing various primers. Besides, those DNA fragments obtained by GP can easily be sequenced by

using the same primer used for the random PCR, naturally, of which success is proven. These characteristics of GP must be beneficial especially for large scale-culture collection centers where frequent and large scale analyses are required.

As is the case with all methodologies, the GP method has its own limitation: it contains a stochastic property in the calculation of *PaSS* values (also their derivative genome distances) ([Kouduka et al., 2006](#); [Ahmed and Nishigaki, 2008](#)), which becomes more significant when the relationship of two species is distant although such a stochastic effect is negligible for sufficiently close species as shown in this study ([Watanabe et al., 2002](#)). Empirically, we have already clarified that the range of effective applications is unexpectedly wide; e.g., the genera of insects (Odonata (*i.e.*, dragonfly), Orthoptera (*i.e.*, grasshopper), Hemiptera (*i.e.*, cicada), Lepidoptera (*i.e.*, butterfly), Coleoptera (*i.e.*, beetle), and other related taxa) could be classified congruently with that provided by the classical phenotype-based method ([Ahmed et al., 2011](#)). Through the success of this study, fungal species of *Trichosporon* were shown to be genomically closer than the application limit of the GP approach. So far as the theoretical treatment of such a limitation becomes available, we have to apply this convenient method (GP) with special care to clustering of distantly related species. Fortunately, the results accumulated up to now provided us with very optimistic expectation for applying the GP to a wider range of organisms including all fungi and bacteria. Naturally, there is no problem theoretically in applying the GP to closely related species where the whole process is governed in essence by the deterministic, not stochastic, processes (detailed discussion partly appearing in [Futakami et al., 2007](#)).

The classification of *Trichosporon* based on GP and D1/D2 26S rDNA sequencing gave similar but not identical results (members of Cluster A made of *T. asahii* also closely clustered in Cluster I of the D1/D2 26S rDNA sequence-based tree) (Figs. 2 and 5), reflecting the difference in the criteria of the two approaches. This kind of disagreement is quite common in this field ([Sugita et al., 1999](#)). Notably, the

members of Cluster I in Fig. 5, resulting from 26S rDNA sequencing, have at most a single point mutation while the corresponding species are more clearly separated in Fig. 2 obtained by the GP analysis (though the significance of this difference remains to be solved). Up to now, the interspecies differences of the genus *Trichosporon* had been investigated by sequence analysis of ITS ([Sugita et al., 1999](#)), Cyt b ([Biswas et al., 2005](#)), 18S rDNA ([Takashima et al., 1999](#)), D1/D2 26S rDNA ([Fell et al., 2000](#)), and IGS ([Rodriguez-Tudela et al., 2005](#); [Sugita et al., 2004](#); [Sugita et al., 2002](#)), while RAPD ([Sugita et al., 2001](#)), AFLP ([Wolf et al., 2001](#)), and MLST ([Juan et al., 2004](#)) analyses have been employed for the investigation of intraspecific differences. These analyses were adopted for the specific purpose of discriminating a particular group of species and strains that was not possible by standard methods. Intriguingly, our results demonstrated the successful discrimination of both of inter- and intra-specific differences by a single GP experiment (Fig. 2).

Finally, the best way to take advantage of the GP method in culture collections must be categorized into two phases: Phase I) to use it as a daily, preliminary assay applied to all kinds of organisms and Phase II) to use it for in-depth analyses including ccgf-sequencing together with the multi-probed approaches. Evidently, the data obtained by the GP analysis is consistent with and complementary to the data obtained by the other methods as shown in this study (Table 2; GP and 26S rDNA sequencing approaches). Considering the readiness and universality of GP ([Kouduka et al., 2007](#)), it is most suitable as a standard approach for most of culture collections. This is essentially important for running reliable culture collections since they must be usually operated with a limited amount of manpower and fund against an increasingly large number of collections, each of which requires identification experiments with highly trained and multi-step skills (which can be considered to be a possible cause for the misidentification revealed in this study).

In conclusion, the GP method, which is composed of i) random sampling of DNA fragments from the genomic DNA (*i.e.*, random PCR), ii) acquisition of the sequence-

based information from the DNA fragments (*i.e.*, TGGE) without conventional sequencing, and iii) computer processing of raw data for normalization, could identify all of the *Trichosporon* species and strains in a culture collection correctly, which was confirmed by sequencing of the corresponding 26S rDNA. At the same time, all of the GP-based inferences for possible false-assignments contained in the *Trichosporon* collections were verified to be correct, indicating the power of the GP analysis. In addition, the usability of the GP method was shown in the triangle distance map, which could reveal unnoticed closeness of seemingly distant species, and in the ccgf-sequencing, which enables us to compare species readily at the sequence level. From the success of this study, the GP method has become evident to be applied safely to a wider range of organisms including fungi.

5. Acknowledgements

This work was performed as part of the Rational Evolutionary Design of Advanced Biomolecules (REDS) Project, Saitama Prefecture Collaboration of Regional Entities for the Advancement of Technological Excellence, supported by JST during 2003-2007.

References

- Ahmed, S., Komori, M., Tsuji-Ueno, S., Suzuki, M., Kosaku A., Miyamoto, K., Nishigaki, K., 2011. Genome profiling (GP) method based classification of insects: congruence with that of classical phenotype-based one. PLoS One (accepted).

- Ahmed, S., Nishigaki, K., 2008. Error-Robust nature of genome profiling applied for clustering of species demonstrated by computer simulation. *International Journal of Biological and Medical Sciences* 3, 82-88.
- Biswas, S.K., Wang, L., Yokoyama, K., Nishimura, K., 2005. Molecular phylogenetics of the genus *Trichosporon* inferred from mitochondrial cytochrome b gene sequences. *J. Clin. Microbiol.* 43, 5171-5178.
- Biyani, M., Nishizawa, H., Miyatani, Y., Kanaumi, E., Eli, P., Kinoshita, Y., Nishigaki, K., 2001. Micronized gel electrophoresis: construction and its versatile applications. *Chem. Lett.* 30, 138-139.
- Fell, J.W., Boekhout, T., Fonseca, A., Scorzetti, G., Statzell-Tallman, A., 2000. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *Int. J. Syst. Bacteriol.* 50, 1351-1371.
- Fell, J.W., Scorzetti, G., 2004. Reassignment of the basidiomycetous yeasts *Trichosporon pullulans* to *Guehomyces pullulans* gen. nov., comb. nov. and *Hyalodendron lignicola* to *Trichosporon lignicola* comb. nov. *Int. J. Syst. Evol. Microbiol.* 54, 995-998.
- Fitz-Gibbon, S.T., House, C.H., 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* 27, 4218-4222.
- Futakami, M., Nishigaki, K., 2007. Measurement of DNA mutations caused by seconds-period UV-irradiation. *Chem. Lett.* 36, 358-359.
- Futakami, M., Salimullah, M., Miura, T., Tokita, S., Nishigaki, K., 2007. Novel mutation assay with high sensitivity based on direct measurement of genomic DNA alterations: comparable results to the Ames test. *J. Biochem.* 141, 675-686.
- Garcia-Vallve, S., Palau, J., Romeu, A., 1999. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli*. and *Bacillus subtilis*. *Mol. Biol. Evol.* 9, 1125-1134.

- Hamano, K., Takasawa, T., Kurazono, T., Okuyama, Y., Nishigaki, K., 1996. Genome profiling: establishment and practical evaluation of its methodology. Nikkashi. 1996(1), 54-61.
- Hatakeyama, Y., Hamano, K., Iwano, H., 2009. An ultimate method for detection of infected pathogenic microorganism from silkworm using HDGP. J. Seric. Sci. Jpn. 77, 1-7.
- Juan, C.R., Koreen, L., Park, S., Perlin, D.S., 2004. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. J. Clin. Microbiol., 42, 2480-2488.
- Koonin, E.V., Galperin, M.Y., 1997. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. Curr. Opin. Genet. Dev. 1997, 7: 757-763.
- Kouduka, M., Matsuoka, A., Nishigaki, K., 2006. Acquisition of genome information from single-celled unculturable organism (radiolaria) by exploiting genome profiling (GP). BMC Genomics 7, 135.
- Kouduka, M., Sato, D., Komori, M., Kikuchi, M., Miyamoto, K., Kosaku, A., Naimuddin, M., Matsuoka, A., Nishigaki, K., 2007. A solution for universal classification of species based on genomic DNA. Int. J. Plant. Genomics. article ID: 27894, 8 pages.
- Middelhoven, W.J., Scorzetti, G., Fell, J.W., 2000. *Trichosporon veenhuisii* sp. nov., an alkane-assimilating anamorphic basidiomycetous yeast. Int. J. Syst. Evol. Microbiol. 50, 381-387.
- Middelhoven, W.J., Scorzetti, G., Fell, J.W., 2004. Systematics of the anamorphic basidiomycetous yeast genus *Trichosporon* Behrend with the description of five novel species: *Trichosporon vadense*, *T. smithiae*, *T. dehoogii*, *T. Scarabaeorum* and *T. gamsii*. Int. J. Syst. Evol Microbiol. 54, 975-986.
- Molnar, O., Schatzmayr, G., Fuchs, E., Prillinger, H., 2004. *Trichosporon mycotoxinivorans* sp. nov., a new yeast species useful in biological detoxification of various mycotoxins. Syst. Appl. Microbiol. 27, 661-671.

- Naimuddin, M., Kurazono, T., Nishigaki, K., 2002. Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution. *Nucleic Acids Res.* 30, e42.
- Naimuddin, M., Kurazono, T., Zhang, Y., Watanabe, T., Yamaguchi, M., Nishigaki, K., 2000. Species-identification dots: a potent tool for developing genome microbiology. *Gene.* 261, 243-250.
- Naimuddin, M., Nishigaki, K., 2003. Genome analysis technologies toward species identification by genotype. *Brief Func Genomics Proteomics* 1, 356-71.
- Nishigaki, K., Amano, N., Takasawa, T., Kinoshita, Y., Husimi, Y., 1990. DNA profiling: method and principle. *Seibutu-buturi* 30, S230 [in Japanese].
- 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.
- Nishigaki, K., Naimuddin, M., Hamano, K., 2000. Genome profiling: A realistic solution for genotype-based identification of species. *J. Biochem.* 128, 107-112.
- Nishigaki, K., Saito, A., Hasegawa, T., Naimuddin, M., 2000. Whole genome sequence-enabled prediction of sequences performed for random PCR products of *Escherichia coli*. *Nucleic Acids Res.* 28, 1879-1884.
- Roderic, D.M.P., 1996. Tree View: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12, 357-358.
- Rodriguez-Tudela, J.L., Diaz-Guerra, T.M., Mellado, E., Cano, V., Tapia, C., Parkins, A., Gomez-Lopez, A., Rodero, L., Cuenca-Estrella, M., 2005. Susceptibility patterns and molecular identification of *Trichosporon* species. *Antimicrob. Agents Chemother* 49, 4026-4034.
- Sakuma, Y., Nishigaki, K., 1994. Computer prediction of general PCR products based on dynamical solution structures of DNA. *J. Biochem.* 116, 736-741.
- Schwartz, D.C., Cantor, C.R., 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell.* 37, 67-75.

- Snel, B., Bork, P., Huynen, M.A., 1999. Genome phylogeny based on gene content. *Nat. Genet.* 21, 108-110.
- Sugita, T., Ichikawa, T., Matsukura, M., Sueda, M., Takashima, M., Ikeda, R., Nishikawa, A., Shinoda, T., 2001. Genetic diversity and biochemical characteristics of *Trichosporon asahii* isolated from clinical specimens, houses of patients with summer-type-hypersensitivity pneumonitis, and environmental materials. *J. Clin. Microbiol.* 39, 2405-2411.
- Sugita, T., Ikeda, R., Nishikawa, A., 2004. Analysis of *Trichosporon* isolates obtained from the houses of patients with summer-type hypersensitivity pneumonitis. *J. Clin. Microbiol.* 42, 5467-5471.
- Sugita, T., Nakajima, M., Ikeda, R., Matsushima, T., Shinoda, T., 2002. Sequence analysis of the ribosomal DNA intergenic spacer 1 regions of *Trichosporon* species. *J. Clin. Microbiol.* 40, 1826-1830.
- Sugita, T., Nakase, T., 1998. *Trichosporon japonicum* sp. nov. isolated from the air. *Int. J. Syst. Bacteriol.* 48, 1425-1429.
- Sugita, T., Nishikawa, A., Ikeda, R., 1999. Identification of medically relevant *Trichosporon* species based on sequences of internal transcribed space regions and construction of database for *Trichosporon* identification. *J. Clin. Microbiol.* 37, 1985-1993.
- Sugita, T., Takashima, M., Nakase, T., Ichikawa, T., Ikeda, R., Shinoda, T., 2001. Two new yeasts, *Trichosporon debeurmannianum* sp. nov. and *Trichosporon dermatitis* sp. nov., transferred from the *Cryptococcus humicola* complex. *Int. J. Syst. Evol. Microbiol.* 51, 1221-1228.
- Takashima, M., Nakase, T., 1999. Molecular phylogeny of the genus *Cryptococcus* and related species based on the sequences of 18S rDNA and internal transcribed spacer region. *Microbiol. Cult. Coll.* 15, 35-47.
- Thatcher, D., Hodson, B., 1981. Denaturation of proteins and nucleic acids by thermal-gradient electrophoresis. *Biochem. J.* 197, 105-109.

- Tominaga, T., 2006. Rapid discrimination of *Listeria monocytogenes* strains by microtemperature gradient gel electrophoresis. J. Clin. Microbiol. 44, 357-358.
- Watanabe, T., Saito, A., Takeuchi, Y., Naimuddin, M., Nishigaki, K., 2002. A database for the provisional identification of species using only genotypes: web-based genome profiling. Genome Biology 3, research0010.1-0010.8.
- Welsh, J., McClelland, M., 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213-7218.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18, 6531-6535.
- Wolf, D.G., Falk, R., Hacham, M., Theelen, B., Boekhout, T., Scorzetti, G., Shapiro, M., Block, C., Salkin, I.F., Polacheck, I., 2001. Multidrug-resistant *Trichosporon asahii* infection of nongranulocytopenic patients in three intensive care units. J. Clin. Microbiol. 39, 4420–4425.

6. Figure Caption

Fig. 1

Main steps of genome profiling (GP). GP consists of DNA extraction, random PCR, μ TGGE and computer-aided data processing (normalization). Random PCR collects double stranded (ds) DNA fragments from genomic DNA by mis-match-allowing hybridization of a primer. It has a meaning of ‘random sampling (in Statistics)’ which extracts partial information from a massive one. Though the random PCR ([Nishigaki et al., 1990](#)) is identical to the concurrently devised RAPD (random amplified polymorphic DNA) ([Williams et al., 1990](#)), the succeeding treatment of the products differs in GP from that of RAPD: TGGE is adopted instead of PAGE as TGGE can exploit the sequence information of DNA through the phenomenon of double-

stranded DNA melting (Nishigaki et al., 1984). RAPD has a risk of band-misassignment due to the temperature fluctuation as shown in the inset where the same samples display quite different band patterns depending on the temperature (This virtual image of RAPD is constructed from the corresponding results obtained by two operations of TGGE (main panel), each of which contains electrophoretic patterns at T_1 and T_2 . In addition, TGGE can discriminate single stranded (ss) DNA which has different band pattern (usually monotonous line) in TGGE from the corresponding dsDNA and often erroneously assigned in RAPD).

The key step of the GP method is the computer aided normalization of featured points (shown in circles in the panel of TGGE), which generates fluctuation-free and reproducible (thus, species specific) data termed *spiddos* (Naimuddin et al., 2000).

Fig. 2

Phylogenetic tree drawn based on the genome distance. The genome distance ($=1-PaSS$) adopted here is the average value over the values obtained with the seven different probes. UPGMA algorithm was used to make this tree. Those species that are disputably assigned are shown in bold and color. T-serial numbers are the same used in Tables 1 and 2. The distinct clusters are designated A – D.

Fig. 3

Triangle distance map (tabular presentation of genome distances). This matrix is the average of d_G ($=1-PaSS$) obtained from the GP experiments using 7 different probes with respect to 38 strains of *Trichosporon*. In this map, genome distances for closely related species (strains) appear near the diagonal in a pale colored box as strains are arranged in the alphabetical ordered of their academic name whereas the pale color off the diagonal represents the closeness between species which are nominally distant (*i.e.*, possibly indicating mis-naming). Strains T-4, T-13, T-18, and

T-29-relevant boxes are highlighted with a bold blue line: lines L1 – L4. It seems that these strains might have undergone frequent mutations and been driven away from the other strains of *Trichosporon* genus, making the relevant boxes dark.

Fig. 4

Detailed genome profile analysis (ccgf analysis). Throughout panels A to E, the genome profiles were obtained using the probe B87 (5' TATCCACCGCTC 3') while in panel F the probe B89 (5' ACTAACCTGGAC 3') was used. In each panel, the species of interest is located at the center. Featuring points of the internal reference DNAs are marked with white dots. Apparent ccgfs (*i.e.*, similar characteristic bands between two species possibly originated from a common gene or so ([Naimuddin et al., 2002](#))) are designated with letters a – j. The number of spiddos for each species is 11 (T-18), 5 (T-20), 6 (T-6), 9 (T-36), 9 (T-35), 7 (T-17), 11 (T-21), 13 (T-19), 10 (T-16), 11 (T-34), 10 (T-33), 10 (T-23), 7 (T-1), 6 (T-11), 10 (T-23), 6 (T-12), 3 (T-14), 2 (T-13), and 3 (T-15).

Fig. 5

Phylogenetic tree based on D1/D2 26S rDNA sequences using the UPGMA algorithm. Each sequence was assigned based on BLAST searches. Those species that led to the decision of misassignment and re-named are indicated in bold. The distinct Clusters I – III are indicated at the right side with Cluster names obtained from the GP-based tree (Fig. 2) where Clusters B and C are divided into subclusters (shown in parenthesis).

7. Tables

Table 1 Fungal strains used in this study

Table 2 Species identification by GP and Sequencing

Table 1: Fungal strains used in this study

Serial No.	Strain number ^a	Species name (registered) ^b	Source ^a
T-1	IFM 48429	<i>Trichosporon asahii</i>	human (=CBS 2479 ^T)
T-2	IFM 48575	<i>Trichosporon asahii</i>	human (=CBS 4829)
T-3	IFM 50223	<i>Trichosporon asahii</i>	human
T-4	IFM 51250	<i>Trichosporon asahii</i>	human
T-5	IFM 51598	<i>Trichosporon asahii</i>	human
T-6	IFM 52660	<i>Trichosporon asahii</i>	human
T-7	IFM 52822	<i>Trichosporon asahii</i>	human
T-8	IFM 53544	<i>Trichosporon asahii</i>	human
T-9	IFM 53746	<i>Trichosporon asahii</i>	human
T-10	IFM 53858	<i>Trichosporon asahii</i>	horse
T-11	IFM 51597	<i>Trichosporon asahii</i> var. <i>faecale</i>	human
T-12	IFM 48559	<i>Trichosporon asteroides</i>	human (=CBS 2481, JCM 2937 ^T)
T-13	IFM 48608	<i>Trichosporon asteroides</i>	human (=CBS 7623)
T-14	IFM 48960	<i>Trichosporon asteroides</i>	cow (=CBS 6183)
T-15	IFM 48961	<i>Trichosporon asteroides</i>	human (=CBS 7624)
T-16	IFM 48963	<i>Trichosporon coremiiforme</i>	human (CBS 2482)
T-17	IFM 53879	<i>Trichosporon coremiiforme</i>	human
T-18	IFM 40140	<i>Trichosporon cutaneum</i>	human (=IFO 1198)
T-19	IFM 41129	<i>Trichosporon cutaneum</i>	human
T-20	IFM 41606	<i>Trichosporon cutaneum</i>	human
T-21	IFM 48620	<i>Trichosporon cutaneum</i>	activated sludge (=CBS 2545, JCM 2940)
T-22	IFM 48565	<i>Trichosporon domesticum</i>	house (=CBS 8280, JCM 9580 ^T)
T-23	IFM 48560	<i>Trichosporon faecale</i>	human (=CBS 4828, JCM 2941)
T-24	IFM 48564	<i>Trichosporon gracile</i>	deer (=CBS 6861, JCM 3940)
T-25	IFM 48609	<i>Trichosporon gracile</i>	sour milk (=CBS 8189 ^T)
T-26	IFM 48551	<i>Trichosporon inkin</i>	human (=CBS 5585, IFO 10131)
T-27	IFM 50224	<i>Trichosporon inkin</i>	human
T-28	IFM 48577	<i>Trichosporon jirovecii</i>	human (=CBS 6864 ^T)
T-29	IFM 48578	<i>Trichosporon jirovecii</i>	Nile crocodile (=CBS 6950)
T-30	IFM 48579	<i>Trichosporon laibachii</i>	white rat (=CBS 2495)
T-31	IFM 48563	<i>Trichosporon loubieri</i>	cow (=CBS 7065, JCM 3939)
T-32	IFM 53856	<i>Trichosporon loubieri</i>	cow
T-33	IFM 48558	<i>Trichosporon moniliiforme</i>	curding milk (=CBS 2467, JCM 2389 ^T)
T-34	IFM 48581	<i>Trichosporon moniliiforme</i>	water (=CBS 5959)
T-35	CBS 8261	<i>Trichosporon montevidense</i>	human
T-36	IFM 48582	<i>Trichosporon montevidense</i>	water purification tank (=CBS 6721 ^T)
T-37	IFM 48611	<i>Trichosporon mucoides</i>	human (=CBS 7625 ^T)
T-38	IFM 49226	<i>Trichosporon sporotrichoides</i>	moist humus (=CBS 8245, JCM 9941 ^T)
X-1	JCM 1542 ^T	<i>Candida albicans</i>	skin disorder (=CBS 562, IFO 1385)
X-2	JCM 1541 ^T	<i>Candida tropicalis</i>	bronchitic patient (=CBS 94, IFO 1400)

^aSymbols and abbreviations: IFM = Medical Mycology Research Center, Chiba University, Chiba, Japan; CBS = Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; JCM = Japan Collection of Microorganisms, RIKEN BioResource Center, Wako, Japan; IFO = Institute for Fermentation, Osaka, Japan; T = type strain.

^bPhenotypically identified based on morphological and biophysical traits.

Species identification by GP and Sequencing

Table 2: Species identification by GP and Sequencing

Serial No. ^a / Labeled name	Closest species ^b (PaSS score)	ccgf ^c	Inference by GP ^d	Assignment by sequencing ^e (DDBJ Accession No.)	GP's Score ^f
T-1 / <i>T. asahii</i> IFM 48429			+	+	(AB492277) ○
T-2 / <i>T. asahii</i> IFM 48575			+	+	(AB492241) ○
T-3 / <i>T. asahii</i> IFM 50223			+	+	(AB492243) ○
T-4 / <i>T. asahii</i> IFM 51250			+	+	(AB492244) ○
T-5 / <i>T. asahii</i> IFM 51598			+	+	(AB492245) ○
T-6 / <i>T. asahii</i> IFM 52660			+	+	(AB492246) ○
T-7 / <i>T. asahii</i> IFM 52822			+	+	(AB492247) ○
T-8 / <i>T. asahii</i> IFM 53544			+	+	(AB492248) ○
T-9 / <i>T. asahii</i> IFM 53746			+	+	(AB492249) ○
T-10 / <i>T. asahii</i> IFM 53858			+	+	(AB492250) ○
T-11 / <i>T. asahii</i> var. <i>faecale</i> IFM 51597	T-24 (0.932) / T-25 (0.920)		− / +	− [<i>T. insectorum</i>] ^h	(AB492267) ○!
T-12 / <i>T. asteroides</i> IFM 48559	T-14 (0.943) ^g	*	− / + [=T-14]	+	(AB492252) △
T-13 / <i>T. asteroides</i> IFM 48608	T-15 (0.952) ^g	*	− / + [=T-15]	− [<i>T. japonicum</i>] ^h	(AB492268) ○!
T-14 / <i>T. asteroides</i> IFM 48960	T-12 (0.943) ^g	*	− / + [=T-12]	+	(AB492251) △
T-15 / <i>T. asteroides</i> IFM 48961	T-13 (0.952) ^g	*	− / + [=T-13]	− [<i>T. japonicum</i>] ^h	(AB492269) ○!
T-16 / <i>T. coremiiforme</i> IFM 48963			+	+	(AB492253) ○
T-17 / <i>T. coremiiforme</i> IFM 53879			+	+	(AB492256) ○
T-18 / <i>T. cutaneum</i> IFM 40140			+	+	(AB492257) ○
T-19 / <i>T. cutaneum</i> IFM 41129	T-17 (0.962) / T-35 (0.961)	*	− [<i>T. coremiiforme</i>]	− [<i>T. coremiiforme</i>]	(AB492255) ○!
T-20 / <i>T. cutaneum</i> IFM 41606	T-6 (0.961)	*	− [<i>T. asahii</i>]	− [<i>T. asahii</i>]	(AB492242) ○!
T-21 / <i>T. cutaneum</i> IFM 48620			+	+	(AB492258) ○
T-22 / <i>T. domesticum</i> IFM 48565			+	+	(AB492259) ○
T-23 / <i>T. faecale</i> IFM 48560	T-33 (0.966)	*	− / +	+	(AB492261) △
T-24 / <i>T. gracile</i> IFM 48564			+	+	(AB492264) ○
T-25 / <i>T. gracile</i> IFM 48609			+	+	(AB492263) ○
T-26 / <i>T. inkin</i> IFM 48551			+	+	(AB492265) ○
T-27 / <i>T. inkin</i> IFM 50224			+	+	(AB492266) ○
T-28 / <i>T. jirovecii</i> IFM 48577			+	+	(AB492270) ○
T-29 / <i>T. jirovecii</i> IFM 48578			+	+	(AB492271) ○
T-30 / <i>T. laibachii</i> IFM 48579			+	+	(AB492272) ○
T-31 / <i>T. loubieri</i> IFM 48563			+	+	(AB492273) ○
T-32 / <i>T. loubieri</i> IFM 53856			+	+	(AB186489) ○
T-33 / <i>T. moniliiforme</i> IFM 48558	T-23 (0.966)	*	− [<i>T. faecale</i> ?]	− [<i>T. faecale</i>]	(AB492260) ○!
T-34 / <i>T. moniliiforme</i> IFM 48581			+	+	(AB492274) ○
T-35 / <i>T. montevidense</i> CBS 8261	T-17 (0.965)	*	− [<i>T. coremiiforme</i>]	− [<i>T. coremiiforme</i>]	(AB492254) ○!
T-36 / <i>T. montevidense</i> IFM 48582			+	+	(AB492275) ○
T-37 / <i>T. mucoides</i> IFM 48611			+	+	(AB492276) ○
T-38 / <i>T. sporotrichoides</i> IFM 49226			+	+	(AB492262) ○

^a The serial number is given in the alphabetical order of the species nomenclature. ^b Inferred from PaSS scores (written in parentheses) for potentially misidentified cases.

^c Assigned to be positive and marked with an asterisk (*) if there are evident ccgf bands that support the PaSS-based assignment (see Naimuddin et al., 2002).

^d If a species is adequately located in the GP-based phylogenetic tree, then '+', if the doubt raised by GP analysis and also consistent with the ccgf analysis, then '−' else '−/+'. The inferred species is shown in brace.

^e D1/D2 region of 26S rDNA was sequenced and BLAST hit (best scored species) is shown. If the result agreed with the labeled name, then +, else −. Note that the accession number automatically allotted sometimes falls in the same sequence with a different accession number.

^f If both assignment results, GP- and sequencing-based are consistent, then '○' (perfect agreement) or '△' (not perfect but consistent), else '×'. An exclamation mark (!) is added when the final assignment led to a revision of the stock label.

^g These four strains of *T. asteroides* can be separated into two pairs (i.e., T-12 / T-14 and T-13 / T-15) since these two pairs have significantly higher genome distance (d_G=0.073) between pairs than the average d_G of the other *Trichosporon* species (0.049), indicating that these pairs belong to different species.

^h Species not included explicitly before this analysis, which cannot be named at '*Inference by GP*'.

Main steps of genome profiling (GP)

Fig. 1

Phylogenetic tree drawn based on the genome distance

Fig. 2

Triangle distance map

Fig. 3

Detailed genome profile analysis (ccgf)

Fig. 4

Fig. 5