

カルチャーコレクションにおける菌株  
管理ツールに向けた GP 法の応用研究

2014年3月

埼玉大学大学院理工学研究科  
理工学専攻（主指導教員 教授 西垣功一）

濱 野 圭 一

## 目次

要約	2
序論	5
1章 GP法開拓の背景と理論	12
1節 背景	12
2節 GP法の関数的特性とその理論	14
3節 内部標準試料のダブル化による精度向上の検討	23
2章 カルチャーコレクションに向けてのGP法の応用	26
1節 緒言	26
2節 実験材料および実験方法	30
3節 結果	36
4節 考察	45
3章 総括	51
謝辞	54
引用文献	55
図表	65

## 要約

カルチャーコレクション(Culture Collection: CC)は研究および産業上における微生物の公的な保存・提供の場として、微生物学の基礎および応用の両面で重要な役割を果たしてきた。近年では、CC は微生物の他、組織、細胞、DNA などを含め広い意味での生物遺伝資源という観点から、「バイオリソースセンター(BRC)」と呼称されることも一般的になっている。ライフサイエンス分野においては、しばしば「リソースなくしては研究なし」といわれるように、バイオリソースは欠くことができない極めて重要な資源である。

CC の売上の半分以上を占め、もっとも大きいサービス業務であるカルチャーの分譲は、品質を保証して提供されなければならない。しかしながら菌株には変異が起こってしまうのが実状である。一方、本来あってはならないことであるが、ヒューマンエラーは避け難く、細心の注意を払っていても菌株のアンプルの取り違いによる単純なミスラベルが起こりうる。また最新の学名更新作業から漏れ、古い学名のままの状態で保存されている菌株も稀に生じる。以上の理由から菌株の品質管理のためにチェック作業を定期的に行なう必要がある。そのための実際の作業には、同定・分類の過程が含まれる。微生物の同定・分類には、表現形質の違いに基づく表現型と、遺伝子領域およびゲノムの塩基配列を直接あるいは間接的に観る遺伝子型による手法がある。同定であれば複数の手法を使用して段階的に絞込みをして行い、分類であれば対象菌株の属・種に応じて使用する手法を選択して行なっており、ダイレクトかつ統一的な同定・分類法がないのが現状である。

1990年に多情報なゲノム解析技術として開発された GP 法(原法)は、ラン

ダム PCR でゲノム中から複数の遺伝子領域をランダムサンプリングし、得られた複数の DNA フラグメントを温度勾配ゲル電気泳動法 (TGGE) により分離展開することで、結果的にフラグメントの塩基配列情報を反映させている。今日の GP 法は原法にはなかった「内部参照試料による規格化处理」の追加により、GP データの客観普遍化に成功したものである。

これまでに GP 法を用いて大腸菌や類縁の腸内細菌に関する研究はあったが、専門的機関が管理する多種の微生物サンプルを対象にした体系的な同定・分類は試みられなかった。本学位論文研究は体系的に多数 (真菌: 24 種 73 株, 乳酸菌: 3 種 24 株、全 27 種 97 株) の微生物サンプルを対象として、カルチャーコレクションの菌株管理ツールとしての GP 法の有効性を検証し、菌株データベースの課題解消を目指したものである。

今回、*Trichosporon* および *Candida* 株を使った解析では、GP 法から得られたゲノム距離を使って、種レベルと株レベルを一括してクラスタリングした系統樹を作成することができ、その系統樹からラベル菌種名の誤りを指摘した。併せてゲノムプロファイル (GP) に検出された共通保存遺伝子断片 (*ccgf*) からラベル菌種名の誤りを指摘した。追跡調査として D1/D2 26S rDNA 解析を行なった結果、GP 法による指摘が正しいことを確認した。その後、菌株を保有するそれぞれの CC において正式にラベル菌種名の変更が実施された。

また、今回ゲノム距離マトリックスを一定の距離範囲ごとに色分けを施した三角距離図を作出することで、一見距離のあると思われた種が以外にも近い関係にあることを示す新たな知見を浮上させ、GP 法の解析方法としての用途幅の広さを示すことができた。

真菌および細菌を含めた全 27 種 97 株を対象とした GP 法で得られたゲノム距離を元にして作成した系統樹では、今後の更なるデータの蓄積による解析の必要はあるが、株の差異を表しつつ属毎・種毎にクラスタを形成し、統一した手法でグローバルな(下は株レベルから上はドメイン(界)レベルまで)生物を対象に GP 法による一括分類の可能性を示すことができた。これは CC の使命としてすべての菌株に共通して適用可能な同定・分類法が求められている現在、単に直接的に塩基配列を決める手法の利用に限らずに、幅広くかつ高効率・高性能な技術 (GP 法) の導入を現実的にしたという意味をもっている。

## 序論

地球上に現存すると推定される全生物種の 3000 万種のうち科学的に存在が確認されているのは約 175 万種と言われており、全体の約 6 パーセントに過ぎない<sup>(1)</sup>。未知の生物種の大半は微生物で占められている。全微生物種に対して既知の微生物種の数 は 1 パーセントにも満たないと推測されている。そして 99%の未知微生物種の中には医薬・健康・食品・環境浄化の視点から有用微生物が存在しているに違いないという思いから探索が進められている。

カルチャーコレクション(Culture Collection: CC)は研究および産業上における微生物の公的な保存・提供の場として、微生物学の基礎および応用の両面で重要な役割を果たしてきた。例えば日本薬局方や日本工業規格 (JIS) 等の、防菌防黴試験などの検定・品質管理用には CC の特定の株が指定されている。

近年では、CC は微生物の他、組織、細胞、DNA などを含め広い意味での生物遺伝資源という観点から、「バイオリソースセンター(BRC)」と呼称されることも一般的になっている。ライフサイエンス分野においては、しばしば「リソースなくしては研究なし」といわれるように、バイオリソースは欠くことができない極めて重要な資源である。

CC の機能には、1) 菌株の収集・寄託受付、2) 培養・保存、3) 分譲、4) 交換、5) 品質管理、6) 情報提供、6) 各 CC との連携、7) 同定やコンサルタンシーサービス がある<sup>(2)</sup>。

学問的に分類学においては、新種発表の論文に述べた菌株が、確かに存在する証拠として著者は CC に寄託することが義務づけられており、特許に用いる

菌株についてはブタペスト条約に基づき CC に寄託することが定められている  
(3)。

CC の使命の 1 つとして、寄託や収集された菌株を生物資源として生きた状態で永久的に保存することがある。現在菌株の保存方法として、凍結乾燥・L-乾燥、-80℃ディープフリーザー、-196℃液体窒素タンク等の増殖を伴わない長期保存法が主流となっている。

CC の売上の半分以上を占め、もっとも大きいサービス業務であるカルチャーの分譲は、品質を保証して提供されなければならない。提供されたカルチャーが寄託・収集されたときの性状とは変わっていたり、アンプルの表示ラベルと中身が異なっているようなことはあってはならないのである。

そのために継代培養時の培養条件のコントロールは重要であるし、液体窒素凍結法などの技術革新による安定な保存法の探求も重要である。しかし、それにもかかわらず変異は起こってしまうのが実際のところである。

一方、本来あってはならないことであるが、ヒューマンエラーは避け難く、細心の注意を払っていても菌株のアンプルの取り違いによる単純なミスラベルが起こりうる。また最新の学名更新作業から漏れ、古い学名のままの状態で保存されている菌株も稀に生じる(4)。

以上の理由から菌株の品質管理のためにチェック作業を定期的に行なう必要がある。そのための実際の作業には、同定・分類の過程が含まれる。菌株の同定・分類は、対象となる菌株が分類学的にどの属・種に帰属するのか、あるいはアンプルのラベル種名と中身が一致しているかを判断するために必要不可欠な作業であり、CC の信頼性を確保するために重要な位置を占めている。

現在、微生物の同定・分類に使用されている手法には、大きく分けて表現型によるものと遺伝子型によるものの2つ手法がある。表現型による手法には、その微生物のもつ菌体やコロニー、胞子などの形態的特徴に基づく方法をはじめ、糖の資化性、成育温度や酵素・代謝物などの生理・生化学的性状を調べる方法や、菌体の脂肪酸組成や細胞壁組成、キノン類の組成など化学的性状を調べる方法がある<sup>(5)</sup>。これらの各指標を単独で用いて同定あるいは分類を行うことはなく、得られた各指標を総合的に判断して同定・分類が行われる。そのため試験項目数は必然的に多くなり、操作は煩雑となる。当然これらの技術を習得するには長期のトレーニングを要する。にもかかわらず、微生物は元々その表現型の特徴が乏しい上、表現形質はさまざまな環境要因によって影響を受けやすいため、ベテランの研究者やキュレーターにしても識別がつかないケースが発生しているのが実情である。

一方、遺伝子型による分類・同定は、1970年代に入って、制限酵素が発見されたことで実用化されたり、DNA塩基配列解析技術が開発されるにしたがって、現実的になってきた。1980年代に入って Woese らは細菌の 16SrDNA 配列に基づく系統解析が有効であることを示した。中でもメタン生成細菌などが通常の細菌より進化的に古い細菌群と考え、Archaeobacteria(古細菌)に属するという概念を提案した<sup>(6)</sup>。その後、遺伝子領域が細菌の分類・同定指標として広く用いられるようになった。細菌における 16S リボゾーム RNA 遺伝子<sup>(6)</sup>は、真菌(菌類)においては 18S rDNA に対応する<sup>(7)</sup>。真核細胞ではその他に大サブユニット由来の 26S・28S rDNA<sup>(8)</sup>が主な指標として用いられている。しかしながらこれらの rRNA 遺伝子配列は種レベルの同定情報としては十分な差



異がなく、分解能が低いことが指摘されている。例えば、*Enterococcus durans*, *E. faecium*, *E. hirae*, *E. mundtii* の 4 種は 16S rDNA 配列の相同性が 98.7～99.7%の間にある<sup>(9)</sup>ことや、*E. seriolicida* の type strain (ATCC 49156)と *Lactococcus garviae* の 16S rDNA 配列の相同性が 100%であることが Collins MD らによって既に示されている<sup>(10)</sup>。そのため、より分解能の高い領域として、rRNA 遺伝子のスペーサー領域配列 ITS [Internal Transcribed Spacer]<sup>(11)</sup>や IGS [Intergenic Spacer region]<sup>(12)</sup> を加味して識別判断するほか、特定遺伝子として *gyr B*<sup>(13,14)</sup>、*rpoD*<sup>(14)</sup>、*Cyt b*<sup>(15)</sup> などのような House keeping 遺伝子配列を指標とすることが盛んに試みられている。

株レベルの識別には現在のところ、RAPD[Random Amplified Polymorphism DNA]<sup>(16,17)</sup>、AFLP[Amplified Fragment Length Polymorphism]<sup>(18)</sup>、PFGE[Pulse Field Gel Electrophoresis]<sup>(19)</sup>が多用されている。これらの手法は PCR や制限酵素反応により得られた DNA 断片群をゲル電気泳動法により解析する点で似ている。これらは間接的にゲノムシーケンスの差異を比較する技術である。その他、近年では MLST[Multi Locus Sequence Typing]による株識別の報告が増えている<sup>(20)</sup>。この技術はターゲットとする複数(7 種類程度)の House keeping 遺伝子セットの塩基配列を比較するものであり、共通する遺伝子セットが比較対象となる生物種間に存在することが前提となっている。そのため種内変異(株)解析に利用されているケースが多い。

未知の微生物検体の菌種を同定する場合、現状の同定作業のフローは次の通りである。まず、rRNA 遺伝子(細菌の場合 16S、真菌・菌類の場合 18S または 26S/28S)のシーケンスを解析する。得られたシーケンスを

DDBJ/EMBL/GenBank 等のデータベースにクエリーすることでそれと相同性の高い順のシーケンスが回答され、そのシーケンスに付随している生物種名が得られる。この時点では相同性の最も高い生物種が、即その未知微生物の菌種名と結論することはできない。現在のデータベースは真偽問わずに投げられたシーケンスを全て受入れているために、候補菌種という扱いに止めなければならない。つまり、“当たりを付けた”に過ぎない。次のステップとして候補菌種に関して表現型性状が既知であれば、未知株の性状を分析比較して絞り込んだり、あるいは候補菌種の標準株の PFGE、RFLP や RAPD、MLST データとの比較により種の同定を進めるという複数の手法を組み合わせた作業フローとなり、煩雑な操作が必要となる。

近年、次世代シーケンサーの出現によりシーケンシングのスピードアップが顕著であり、全ゲノムシーケンスによる菌種同定も視野に入ってきた感がある。しかし、導入コストおよびランニングコストの面で、CC が保有する全菌株に対してゲノムシーケンスデータを取得することは現実的ではないし、そもそも MLST 法で 7 つ程度の House keeping 遺伝子シーケンス情報で株の識別が可能である事実が示すように、菌種同定のためだけに全ゲノムシーケンス情報の取得は過剰作業である。

いずれにしても種レベルの識別には単一の特定遺伝子領域配列が指標となり、より詳細な株レベルの識別には複数の遺伝子領域あるいはゲノムワイドな解析手法が用いられている。識別レベルに応じて手法を使い分け、得られた指標結果を総合して識別を行っており、一元的に種および株レベルの識別可能な技術がないのが現状である。ひとつの手法でダイレクトに同定可能な手法が望ま

れている。

1990年に西垣らによって多情報なゲノム解析技術として開発された GP 法 [Genome Profiling] (原法) は、その要素技術としてランダム PCR を含んでいた (生物物理, 1990(8月) (21, 22))。その意味では 2 ヶ月後にランダム PCR の部分だけに関して発表された Williams らの RAPD 法<sup>(16,17)</sup>に先行して独立に開発された技術である。ランダム PCR は RAPD と同義であるが、GP 法(原法) はさらに温度勾配ゲル電気泳動法 (TGGE) との組合せにより DNA フラグメントを塩基配列情報を反映させて分離展開することで、結果的に得られたパターンはゲノムシーケンス情報を反映させている<sup>(22)</sup>。これまでに著者は GP 法 (原法) を大腸菌や類縁の腸内細菌 (4 菌種)、酵母 (1 菌種)、植物 (15 種)、動物 (3 種) を対象にその適用を試み、GP 法が普遍性の高いゲノム解析手法になりうることを示した。しかしながら当時はゲノムプロファイルパターンの差異を内部参照試料法によって定量化する方式は開発されていたが<sup>(23)</sup>、操作に手間のかかるものであり、通常は目視だけで比較する定性的な解析が用いられていた<sup>(24)</sup>。

その後 GP 法は本格的に「内部参照試料として DNA を用いた規格化処理」が導入され、それまでの移動度情報の規格化だけでなく、温度に関する規格化を特徴点を有する二つの DNA の導入により実現した。すなわち、それらの内部参照との比較から実験毎にそれぞれの DNA について再現性の高い特徴点を抽出し *Spiddos* (種同定点) に変換した後、それを用いて相同性を評価するのに *PaSS* (パターン・シミラリティ・スコア) を用いた。これにより GP データは客観化・普遍化された<sup>(25)</sup>。更に、Biyani らが TGGE の超小型化に成功し、

従来の TGGE に要した時間が 180 分から 10 分程度に大幅に短縮された<sup>(26)</sup>。

このような GP 法の発展により、簡便・迅速にゲノム情報の取得および定量解析が可能になり、幸塚らは GP 法による植物、昆虫、魚の各属レベルの分類結果が、従来の表現型による分類結果と一致することを実証し、GP 法のユニバーサルな生物分類法としての有効性を示した<sup>(27)</sup>。また二上らは変異原性物質を被曝させた大腸菌の継代培養を行い、そのゲノムの変化を GP 法で測定する方式で、変異原検出において Ames 試験の 100 倍高感度(10ppb 検出)を実現した<sup>(28)</sup>。

上記のように昆虫や魚など属レベルの分類が GP 法で可能であることは示されていたが、多数の体系的な微生物株を対象にした GP 法による同定・分類は試みられていなかった。本研究は体系的に多数の微生物サンプル（真菌：24 種 73 株，乳酸菌：3 種 24 株）を対象として、GP 法による同定・分類を試み、カルチャーコレクションにおける菌株管理ツールとしての有効性を検証し、そのあり方を考察した。

# 1 章 GP 法開拓の背景と理論

## 1 節 背景

1986 年、西垣は TGGE における温度と移動度の座標平面に現れた DNA の変性曲線には、その曲線の形状を規定する塩基配列情報量をもつことを示した<sup>(29)</sup>。また、簡易な利用としては、プロファイル上で顕著な転移を与える幾つかの点はその時の温度と移動度を代表する特徴点となりうることを示唆し、TGGE が DNA の塩基配列情報を引き出す解析手法になることを明確にした<sup>(23)</sup>。

1990 年、西垣は巨大ゲノムから再現的に遺伝子断片群をサンプリングする方法として、ランダム PCR を考案した<sup>(21)</sup>。ほぼ同時期に Williams らや McClelland らがランダム PCR と同義の RAPD を発表している<sup>(16)(17)</sup>。西垣らは *E. coli* の 3 株、すなわち S26, X4-4, LE392 のゲノムをテンプレートにランダム PCR を行い、得られたランダム PCR 産物を TGGE に供することで、ゲノムの塩基配列情報を取得可能であることを示した<sup>(22)</sup>。その際に、ランダム PCR は TGGE に対して、有限個の適度な DNA 断片数を産生し、かつ断片サイズも  $10^2 \sim 10^3$  bp で適度であり、当時のゲノム情報解析技術の RFLP と比較し、得られる情報量と簡便性で優位であることを示していた。

その後、GP 法の原理に関する議論として、関数的特性と GP 法によってゲノムから引出される情報の理論的背景について議論がなされ、著者はその実証を広範囲な生物種、すなわち、腸内細菌 9 株 (*E. coli* 6 株, *Shigella*, *Salmonella*, *Yersinia*)、植物 13 種、動物 3 種を用いて行なった。その結果、関数的特性については、操作条件が一定であれば、GP 法によって得られる情報はゲノムテ

ンプレートとプライマーの関数で表される事象であることを、i) いずれの生物種（株）に対しても、同一の操作で種（株）ごとに固有のゲノムプロファイルが得られ、ii) プライマーを換えることで、それに伴って同一ゲノムから異なるプロファイルが得られたことで、実証することができた<sup>(30)</sup>。さらに著者らは種同定に必要な情報量について議論し、GP 法がその情報量を満足しうることを理論的に証明した<sup>(31)</sup>。このことにより、GP 法による広範な生物・細胞のゲノムを一元的に比較・同定する方法論が確立された。なお、GP 法の関数的特性と理論的検討については、次節に述べる。

2000 年以降、バイオテクノロジーのハイスループット化が盛んに進んだ。GP 法に関してもその例に漏れず、ハイスループット化が進められた。GP 法の構成技術の TGGE では、従来は 20cm×20cm のポリアクリルアミドゲルを使用して、1 回の泳動に 3 時間を費やしていた。2001 年 Biyani らが 2.5cm×2.5cm ミニゲルを使用した TGGE で得られた DNA のバンドパターンが、従来の TGGE のバンドパターンと遜色のないことを実証した。また泳動時間も 10 分前後にまで大幅に短縮され、試料の微量化とハイスループット化を可能にした<sup>(26)</sup>。この実証を受けて、ミニゲル用の TGGE 装置の開発プロジェクトが産学協同開発として西垣らとタイテック株式会社とで始まった。著者はこのプロジェクトに企業側(タイテック)のスタッフとして参画し、主にミニゲル作製用ディスプレイサブカセットの開発とその評価および TGGE 装置（名称； $\mu$  TG）の泳動評価に従事した経緯がある。

## 2 節 GP 法の関数的特性とその理論

### 2-1 GP 法<sup>(30)</sup>の関数的特性

ランダム PCR 産物を  $P$ , 鋳型 DNA を  $T$ , プライマープローブを  $p_r$ , 反応条件を  $C$  と置くと、ランダム PCR には

$$P = f(T, p_r, C) \quad (1)$$

の関係がある。ここで  $P$  は多種の DNA 断片  $f_1, f_2, \dots, f_n$  からなり、それぞれ固有の濃度  $[f_1], [f_2], \dots, [f_n]$  に増幅されているもの全体とする。鋳型 DNA,  $T$ , およびプライマー  $p_r$  は共に塩基配列と濃度の関数であり、反応条件  $C$  は次式で表せるような媒介変数から成り立っている。

$$C = C(T_1, T_2, T_3, t_1, t_2, t_3, B) \quad (2)$$

この式において、 $T_1, T_2, T_3, t_1, t_2, t_3$  は PCR サイクルにおける 1. 熱変性, 2. アニーリング, 3. 鎖伸長のそれぞれの過程での温度 ( $T$ ) および反応時間 ( $t$ ) を表し、 $n$  は PCR のサイクル数、 $B$  は反応緩衝液組成条件を表している。一般に  $C$  は固定され、 $p_r$  は任意に設定しうるから、結局、ランダム PCR はあるゲノム DNA  $T$  に対してその断片配列集合  $P$  を与える操作  $f$  になっている。

一方、TGGE は  $P$  全体を泳動パターン情報  $I$  に変換する操作  $g$  になっている。すなわち、

$$\begin{aligned} I &= g(P) \\ &= g(f(T, p_r)) = g \cdot f(T, p_r) \end{aligned} \quad (3)$$

この(3)式における連結操作  $g \cdot f$  がゲノムプロファイリングそのものである。このような数理的表現をとる重要な意味は、実験的操作によって得られる DNA 断片を、(もしゲノムの塩基配列全体がわかれば) 理論的変換操作によって求

めることができる事実を明示することにある。このことは既に部分的には、ラムダファージのゲノム DNA(48502 塩基長)を用いた実験系で任意のランダム PCR 産物を理論的に予測することを通して実証されている<sup>(32)</sup>。その後、大腸菌ゲノム全体に対してランダム PCR を行い、その産物を予測することでこの理論の正当性を実証的に証明した<sup>(33)</sup>。また、式(3)は同一のゲノム DNA  $T$ であっても、プライマー  $p_r$  を換えることで異なるゲノム断面像、すなわちプロファイル  $I$  を与えることを表現している。

## 2-2 理論

### 2-2-1 ランダム PCR

通常の PCR では、Fig. 1(1)に示すように、鋳型となる DNA の特定部位に完全相補で特異的に対合する 2 種類のプライマーによって、複製される領域が決まる。しかし、ランダム PCR では Fig. 1(2)のように一般に完全相補でない対合構造で互いに向き合った状態にあるときに、その鋳型 DNA (+/-) とプライマーとの結合構造の安定性に応じて第一世代の娘鎖 (-/+ ) が形成される。次にそれを鋳型としてもう一度、プライマーの不完全相補鎖が生じ第二世代の娘鎖 (+/-) ができる (この場合も対合構造の安定性はその生産速度を支配している)。このようにして出来た第二世代娘鎖に含まれるプライマー結合部位は既にプライマーと完全相補な塩基配列となっており、次段以降の PCR サイクルでは通常の (特異的) PCR と同様な複製効率になる。以上のことを式で表せばランダム PCR で生じる DNA 断片の回収量  $Y$  は、親鎖および第一世代娘鎖に対してプライマーが結合したときに現れる不完全相補結合の安定性を表す



結合定数  $K_0$ ,  $K_1$  およびそれらの自由エネルギー変化表現  $\Delta G_0$ ,  $\Delta G_1$  を用いて

$$\text{Log} Y \propto \log K_0 \cdot K_1 = (\Delta G_0 + \Delta G_1) / RT \quad (4)$$

(ただし、 $R$  は気体定数、 $T$  は反応温度)

となる。この式はランダム PCR 産物が鋳型とプライマーの不完全相補複合体の存在割合に比例して生じること、および DNA 断片 (産物) を生じるためには結果的に生じる断片の両端で同時に複合形成がなされる (積事象である) 必要のあることを表している。このことは塩基配列既存のラムダフェージや大腸菌を鋳型 DNA としたランダム PCR 実験の解析から、理論的、実験的に支持されている<sup>(32, 33)</sup>。別の観点から表現すれば、このようにして同じプライマーを用いたランダム PCR によって増幅された DNA 断片は、その断片の両端に相似した塩基配列を持つものとして共通している。もし断片長もほぼ同じであれば、偶然一致の出現確率はそれだけ低くなり、加えて塩基配列まで類似している場合には、事実上、偶発的な一致の確率は超極小となる。そのような場合には相同する遺伝子から由来している確率が高くなることを意味している。

## 2-2-2 TGGE による塩基配列情報の抽出

DNA はその塩基配列固有の融解現象を呈する<sup>(34)</sup>。このことに基づいて発展した温度勾配ゲル電気泳動法 (TGGE) は、簡便かつ多情報に配列情報を与える<sup>(23, 35, 36)</sup>。二本鎖 DNA の TGGE における融解曲線のイメージ図を Fig. 1(3) に示す。図において、泳動時に協同的融解に対応する部分変性が生じているところ (すなわち部分的にまとまって二本鎖 DNA の塩基対合が崩れているところ) であり、一般に (A+T) 含量が高いほど低い温度で融解するが、厳密には同じ

(A+T)含量であってもその配列によって熱的安定性が異なることが知られている<sup>(34)</sup>では、移動度に顕著な変化が認められる。この協同的融解の起こる DNA 上の部位とその時の温度が塩基配列固有で (DNA によっては 2 段にも 3 段にも、協同的融解が起こり、その度ごとに移動度の変化が起きる)、その際の分子形状がそれ特有の移動度を与える。その結果、TGGE において DNA 塩基配列固有のプロファイルを描くことになる。すなわち TGGE では、この温度-移動度平面における曲線を規定するに足りるだけの情報量が産生されている<sup>(29)</sup>。

得られたゲノムプロファイル(GP)は生物種(株)固有のバンドパターンであり、慣れればこれから生物種を判定することも可能である。高度なパターン認識過程をより簡単な方式に置き換えるものとして、特徴点抽出法が開発された<sup>(25)</sup>。Fig. 2(b) に示すように、GP はそれぞれの DNA の二本鎖融解過程が現れたものであり、そこには融解開始点( $P_{ini}$ )、最小移動度点( $P_{min}$ )、融解終了点( $P_{end}$ )などの特徴点が現れている。とりわけ、融解開始点は再現性が高く、泳動条件の変更に対しても変動が少ないことが知られている。したがってこれらの点を抽出し登録することで、特定のゲノムを代表したり、同定に使用したりすることが可能となる。これらの点は *spiddo* (species identification dot; 種同定点) と定義している<sup>(25)</sup>。

Fig. 2(c) に示すように、*Spiddo*  $\vec{P}$  の座標は温度と移動度によって規定され、 $\vec{P}_i = P(T, m)$  と表される。移動度や融解温度のわかっている標準試料 (内部参照試料) を共泳動させておき、それを用いて規格化处理することで、*spiddo* の座標  $\vec{P}_i = P(T, m)$  はゲノム固有の物性値を与える。

通常、GP で現れるバンドの内、明瞭なもの 8~10 本から *spiddo* を得る。2 種のゲノム  $G^{(1)}$  と  $G^{(2)}$  の近さを議論するときには、式(5)で定義される *PaSS* (Pattern Similarity Score; パターン類似度) を算出する。なお、実際の操作としては、(5)式をプログラミングしたソフトウェアを使用して、コンピュータ画面上で GP 画像中の特徴点を打点する操作で、一括して 2 つのゲノム  $G^{(1)}$  ,  $G^{(2)}$  間の *PaSS* をマトリックス形式で得ることが機械的にできる状況となっている。

$$PaSS = 1 - \frac{1}{n} \sum_{i=1}^n \frac{|\vec{P}_i^{(1)} - \vec{P}_i^{(2)}|}{|\vec{P}_i^{(1)}| + |\vec{P}_i^{(2)}|} \quad 0 \leq PaSS \leq 1 \quad (5)$$

(5) 式は温度と移動度で規定されたベクトル量  $\vec{P}_i = P(T, m)$  をもつ *spiddos* を使用して、 $0 \leq PaSS \leq 1$  のスカラー量でゲノム塩基配列に基づいた生物種間の類似度を簡便な尺度で比較可能なことを表している。

ゲノムの類似度を生物種間の近似的なゲノム距離、つまりは「ゲノム準距離」に変換するならば、ゲノム準距離  $d_g$  は *PaSS* を使って(6)式のように表現できる。

$$d_g = 1 - PaSS \quad (6)$$

ゲノム準距離を用いることで、これまでに比較することができなかった例えば大腸菌と象といった明らかに遠縁にある種間の距離を一義的に知ることができるようになり、全生物を対象に同一の操作で種間の距離について議論が可能となる。このことは少なくともこれまでの GP 研究において実践的に支持されている。

### 2-2-3 生物種同定における情報量の理論的見積

種同定を考える際、ゲノム塩基配列の完全一致は疑いもなく同一と判定できるが、そこまで必要なく、ゲノム中の一部領域の一致でも同定に十分である場合がある。例えば、16SrDNA 配列 (1400–1500 塩基) ではその類似度が 97% 以上で同種と判定されているが、100%の類似度でも菌種を特定できないケースも報告されている<sup>(37)</sup>。では、ゲノム塩基配列に基づいて同定を行う場合に、必要な情報量とは一体どのくらいで十分なのであろうか？

ここで、種同定に必要な情報量を  $I_s$  とすると、 $I_s$  は次のように見積もることができる。全生物種の全ゲノムに亘り、 $s$  塩基長の配列である生物種のゲノムがユニークに特定されるならば、 $I_s$  は  $s$  塩基長の最小値にできる。全生物種数とそれらの平均ゲノムサイズをそれぞれ  $m$  と  $\langle n \rangle$  とすると、期待値理論に基づいて次式が成り立つ。

$$m \cdot \langle n \rangle \cdot p^s \leq 1 \quad (7)$$

ここで、 $p$  はある領域でまさにそのヌクレオチド G, A, T, C になっている期待値である。 $m$  は多くても  $m \leq 10^{30}$  と見積もられる。これは地球の面積 ( $5 \times 10^{18} \text{cm}^2$  と見積られる) における地表  $1 \text{cm}^2$  範囲の径  $1 \text{cm}$  の円柱空間中に  $2 \times 10^{11}$  種の生物種が存在しているという仮定 ( $2 \times 10^{11} \times 5 \times 10^{18} = 1 \times 10^{30}$ ) に基づいて得られている。これはちょうど、大腸菌培養時の飽和状態に相当する  $2 \times 10^9 \text{cells/ml}$  が  $1 \text{m}$  高さに積み上げられた状態であるが、過剰に見積もられた数字である。一方、ゲノム平均サイズ  $\langle n \rangle$  は、巨大ゲノムの一つとして知られる熱帯地域のシダ類 *Psilotum nudum* のゲノムサイズが  $1.3 \times 10^{11}$  もあることから、大きくみて  $10^{12}$  を使うことにする。

期待値  $p$  については、ゲノム配列がランダム配列ではないため、必ずしも  $1/4$  (ランダム配列の場合) ではないことから、議論の余地はあるところではあるが、保守的に考えて  $1/2$  とした。したがって塩基長  $s$  の最小値は(7)式において、 $m=10^{30}$ ,  $\langle n \rangle = 10^{12}$ ,  $p = 1/2$  とすると、 $s = 140$  となる。これはほんの  $140$  塩基長あるいは  $140$  ビットの情報量 (各文字は平均で  $1\text{bit}$  ( $= -\log_2 p = -\log_2 1/2$ ) の情報量をもっている) でゲノムの特定、すなわち全生物種からその種の特定が可能であることを示している。2 生物種が  $s$  塩基長の同一配列をもつリスク発生率  $P_R$  は次式で表される。

$$P_R = 1 - \{ {}_v C_0 q^0 (1-q)^v + {}_v C_1 q^1 (1-q)^{v-1} \} = \{(v-1)q\}^2 = (vq)^2 \quad (8)$$

ここで、 $q = p^s$ ,  $v = m \cdot \langle n \rangle$  であり、 $s = 200$  と仮定した場合、つまり  $q = p^s = (1/2)^{200} = 10^{-60}$ 、 $v = 10^{42}$  となり、リスク発生率は  $10^{-36}$  となる。このリスクを考慮すると種同定に十分な情報量  $I_s$  はわずか  $200$  文字 ( $200$  ビット) の配列で得られるということになる。このように正確な  $I_s$  の値が算出されたにも関わらず、 $200$  ビット以下の情報量で生物種が同定されている事実は、ゲノムサイエンスにおいて極めて重要である。すなわち、これは全ゲノム中のわずかな配列で各ゲノムを同定しうることを意味していることである。さらに、連続したものあるいは断続的なものに関係なく、ゲノム中のいずれの配列も同定に利用できることである<sup>(31)</sup>。

#### 2-2-4 GP 法で得られる情報量の理論的見積

GP 法で得られる情報量について考察する。まず、2-2-1 でも述べた、単一のプライマーを用いたランダム PCR において、プライマーはミスマッチを含

みながらもテンプレート DNA に不完全相補対合を形成するが、最終的には増幅されたフラグメントの両端は同一配列になっている。本研究で使用した 12mer のプライマーの場合、3-9 塩基が相補対合を形成すると見込まれ、情報量としては 6-18 ビットになる。したがって、各ランダム PCR 産物は両端がプライマー配列となっているため、12 (=2×6) ビット以上の情報を有していることになる。

次に、各特徴点の位置座標は DNA の塩基配列に対応していることが知られている<sup>(36)</sup>。したがって、特徴点には配列情報が含まれている。その含まれている全情報量を決定することは困難であるが、次のようにおよその見積りは可能である。例えば、それぞれの特徴点がゲルの 2 次元平面上の 100 (例えば 10×10) ブロックから 1 つだけに特定すると仮定した場合、1 つの特徴点がこれらのブロックの 1 つに位置する確率は、多くて 7 ビット (=  $-\log_2 P = -\log_2 1/100$ ) に相等する確率で各特徴点がアサインされることになる。ゲノムプロファイル 1 枚から 8~10 本のバンドが得られるのが好ましいことから、少なくとも 8 本とした場合、各バンドの二本鎖融解開始点を特徴点として抽出すると、1 回の GP 操作で得られる情報量の総和  $I_{GP}$  は次式のようにになる。

$$I_{GP} \geq \{(6 \times 2) + 7\} \times 8 = 152 \quad (9)$$

情報量の総和  $I_{GP}$  は、前節で述べた同定に必要な情報量の上限  $I_s$  (=200 ビット) の半分を優に超えている。この見積りは低くみているため、実際は 1 回の GP 法の操作でゲノムの同定に十分であることが確認されている。たとえ不十分な場合でも、プライマーを換えた GP 法の操作で、ゲノムの異なる断面をみることが可能であり、十分である。結果的に GP 法は種同定の要求に容易に応えうる

目的十分な手法となっている<sup>(31)</sup>。

### 3 節 内部標準試料のダブル化による精度向上の検討

#### 3-1 ダブルレファレンスの開発

GP 法において、ランダム PCR 産物を TGGE 展開して得られたゲノムプロファイルデータを比較可能とするためには標準規格化が必要である。標準規格化のために TGGE において、ランダム PCR 産物と既知 DNA を内部参照試料として共泳動させる。以前、内部参照試料はファージ fd DNA 中の 200bp の DNA 断片 1 種類を用い、その二本鎖融解開始点と二本鎖解離終了点を対角に取って行なっていた。しかし、完全二本鎖解離終了点は不安定な動きを示し泳動毎にゆらぐため、データ精度の影響が懸念されていた。そのような理由から 2 種類の内部参照試料の二本鎖融解開始点を用いることで、その懸念を解消し、データ精度の維持を目的に内部参照試料のダブル化を試み、ダブルレファレンスの開発を行なった<sup>(38)</sup>。

ダブルレファレンスの条件として、TGGE ゲル上にて 2 つの二本鎖融解開始点が適切な位置にくること、GP のバンドパターン中において内部参照試料のバンドパターンが識別しやすい形状であることを前提として、そのような DNA 断片の探索を行った。試行の結果、プラスミド pBR322 DNA 中の 3501-3700 と 401-1300 の部位に由来するそれぞれのサイズが、IntRef. 5: 200bp と IntRef. 6: 900bp の dsDNA 断片が適切であると判断した。IntRef. 5 および IntRef. 6 の作製方法は、2 章 2-3 **ゲノムプロファイリング法** Genome profiling (GP) で示したとおりである。PCR で増幅された IntRef. 5 および IntRef. 6 の TGGE におけるバンドパターンとそれぞれの塩基配列を Fig. 3-1 に示した。



### 3-2 ダブルおよびシングルレファレンスの座標補正による安定性の比較試験

開発したダブルレファレンスと従来のシングルレファレンスの安定性の比較は、同一ランダム PCR 産物の共泳動にダブルレファレンス (IntRef. 5, IntRef. 6 混合) とシングルレファレンス (IntRef. 1) をそれぞれ用いて TGGE を行い、その後、解析ソフトを使用してランダム PCR 産物の明瞭な特徴点を打点し、座標補正をダブルレファレンスについては IntRef. 5, IntRef. 6 のそれぞれの二本鎖融解開始点を対角にとり、一方シングルレファレンスについては IntRef. 1 の二本鎖融解開始点と完全二本鎖解離終了点を対角にとった(Fig. 3-2)。

TGGE はダブルレファレンスおよびシングルレファレンスについてそれぞれ 4 回ずつ ( $W_1 \sim W_4, S_1 \sim S_4$ ) 行った。ダブル内の 6 通りの組み合わせ ( $W_1/W_2, W_1/W_3, W_1/W_4, W_2/W_3, W_2/W_4, W_3/W_4$ ) とシングル内の 6 通り組み合わせ ( $S_1/S_2, S_1/S_3, S_1/S_4, S_2/S_3, S_2/S_4, S_3/S_4$ ) の *PaSS* 計算を行い、*PaSS* 計算結果からダブルとシングルの安定度の比較を *PaSS* の平均値と標準偏差をとって行った。つまり同一サンプルを用いて同一操作で繰り返しているので、理想的には *PaSS* 平均値は 1 に近づき、標準偏差は 0 に近づくことが望ましい。

各 6 通りの組み合わせの *PaSS* 値はダブルレファレンスでは、 $W_1/W_2 = 0.994$ ,  $W_1/W_3 = 0.985$ ,  $W_1/W_4 = 0.988$ ,  $W_2/W_3 = 0.987$ ,  $W_2/W_4 = 0.992$ ,  $W_3/W_4 = 0.993$  となり、平均値は 0.990 であった。一方、シングルレファレンスでは、 $S_1/S_2 = 0.965$ ,  $S_1/S_3 = 0.978$ ,  $S_1/S_4 = 0.985$ ,  $S_2/S_3 = 0.971$ ,  $S_2/S_4 = 0.979$ ,  $S_3/S_4 = 0.983$  となり、平均値は 0.977 であった。平均値において有意な差が表れていた。また標準偏差については、ダブルレファレンス標準偏差は 0.003438, シングルレファレンス標準偏差は 0.007727 となり、ダブルレファレンス使用時のデータ

処理におけるバラツキが減少して、データの精度が上昇したことが明らかになった。

今回、2 章の応用研究で使用した内部標準試料は、すべてダブルレファレンスを使用した。したがって精度の向上したデータを扱っていることになり、信頼できるものとなっている。

## 2章 カルチャーコレクションに向けての GP 法の応用

### 1節 緒言

一般に細菌や真菌の同定には、理想的条件から外れた条件下で培養することが多く成長速度が遅くなることもあって、専門家でも結果を得るまでに数週間から数カ月かかることが少なくない。とりわけ真菌の同定には時間がかかり煩雑な作業を必要とする。1つの理由としては真菌には形態学的に酵母型と菌糸型があり、*Candida albicans* などは外界では菌糸型を、また体内では酵母型をとり二形成真菌と呼ばれるタイプであり、また、有性世代期と無性世代期でその表現形を変えるものもあり、同定・分類は難しい。

*Trichosporon* 属は、環境中とくに土壌や腐朽樹木などに広く生息している担子菌系不完全酵母の分類群に属する。形態学的には分節型分生子、仮性菌糸及び真性菌糸を産生するのを特徴とする。一方で、この菌はヒトの咽頭、消化管、皮膚などに一過性に（場合によっては恒久的に）定着することがあり、宿主の免疫低下(好中球減少)の条件が重なると侵襲型の全身感染である深在性トリコスポロン症を引き起こす(いわゆる日和見感染症)。深在性トリコスポロン症は、深在性カンジダ症と良く似た病態を呈する。すなわち真菌血症を介して全身臓器に播種し、肺炎、多発性の真菌性肝膿瘍、眼内炎、あるいは皮膚に転移性の真菌膿瘍などを形成する。深在性トリコスポロン症の死亡率は 60-80% と高く、同じ酵母様真菌感染症であるカンジダ症の死亡率が 40%前後とされているのに比べても、予後が不良である<sup>(39,40)</sup>。

*Trichosporon* 属の分類については、その長い研究の歴史のなかでさまざまな菌種名が提案され、時代とともに著しく変遷してきた。特に大きく変わった

のは積極的に分子生物学的手法が用いられるようになった 1990 年前後からである。それ以前は形態学的あるいは生化学・生理学的性状のいわゆる表現型の同定法によるしかなかった。形態学的同定法は臨床材料から菌株を分離培養するのに数日間を要し、その後の同定検査にさらに数日間が必要であった。また、顕微鏡操作にも精通した高度な技術が必要であった。一方、生化学・生理学的性状による同定方法は市販の検査キット<sup>(41)</sup>を利用することができたが、一部の形質が変容した菌株では誤った結果を招く可能性もあった。表現型による方法ではとりわけ *Trichosporon* 属の菌種の検出方法は煩雑で、判定までに長時間(平均 7~14 日間)を要し、さらには多量の試料を必要とする等の問題点が多かった。

表現型による同定法しかなかった時代には *Trichosporon* 属は *Trichosporon beigelii* または *Trichosporon cutaneum* の 2 菌種名のみで扱われていた。しかし、1990 年以降、Gueho らおよび杉田らは、DNA の G-C 含量、DNA 再会合率、リボゾーム RNA 遺伝子特定領域の塩基配列などに基づいて、従来の *T. beigelii* をいずれもヒト病原性をもつ 6 種、すなわち① *T. asahii*, ② *T. asteroides*, ③ *T. cutaneum*, ④ *T. inkin*, ⑤ *T. mucoides*, ⑥ *T. ovoides* に細分類した<sup>(41~44)</sup>。さらに両研究グループはこれら 6 種に環境由来の腐生性菌種を加えた計 17 の菌種と、5 つの変種を含む *Trichosporon* 属の新しい分類を提案した。さらに杉田らは、2002 年に 2 種の潜在的病原菌種 (*T. domesticum* と *T. montevidense*) を加えた 25 菌種<sup>(45)</sup>を、次いで 2004 年には Middelhoven ら<sup>(46)</sup>が提唱した 5 種の新種 (*T. vadens*, *T. simithiae*, *T. dehoogii*, *T. scarabaeorum*, *T. gamsii*) を含む 36 菌種<sup>(47)</sup>を、各々提示した。その後、昆虫

から分離された新菌種 2 種が追加され、現在 *Trichosporon* 属には 40 菌種が登録され、この流動的状況は今も続いている。そのような経緯から *Trichosporon* 属は分類学的に興味深い対象であると考えられ、今回 GP 法の菌株同定の有効性検証に用いた。

真菌症の代表格として *Candida* 症がある。*Candida* 症の多くの原因菌は *Candida albicans* であるが、病原性がある *Candida* には *C. kefyr*, *C. krusei*, *C. guilliermondii*, *C. parapsilosis*, *C. glabrata*, *C. tropicalis* などが挙げられ、現在病原性が認められていない菌種も含めると、*Candida* は 200 以上の菌種が存在する<sup>(48)</sup>。*Candida* の同定方法には、表現型として形成されたコロニーの色が白色～クリーム色で光沢があり、仮性菌糸の形成の有無などを観察する培養法をはじめ、市販の検定キット、例えば API ID 32C (BioMerieux) (炭水化物 22 種類、有機酸 5 種類、アミノ酸 2 種類の資化性とその他の 31 種類の試験により、*Candida* 37 菌種を含む病原性酵母 69 菌種を同定するシステム) などがある。しかし、臨床から分離される *Candida* は多様化しており、キットでは同定が困難なケースが往々にして発生する。そのようなケースの場合、rRNA 遺伝子領域の 26S rDNA D1/D2 や ITS (Internal transcribed spacer) の塩基配列解析が行われる。今回、先に挙げた病原性 *Candida* 7 菌種に絞り、株の多様性も観る目的で *C. albicans* 12 株と *C. tropicalis* 9 株について GP 法の検証を行った。

サワードウ (sourdough) は、小麦やライ麦の粉と水を混ぜてつくる生地に、乳酸菌と酵母を主体に複数の微生物を共培養させた伝統的なパン種である。そこに生育している主な乳酸菌は *Lactobacillus sanfranciscensis* であり、酵母

は *Saccharomyces cerevisiae* である。酵母についてはその他、*Candida milleri* や *C. krusei* などの *Candida* 属も含まれている。北原らは 5 つのパン種から分離した *L. sanfranciscensis* の 21 株についてその多様性をリボタイピング(リボソーム RNA 遺伝子の周辺領域の多型性を利用した遺伝子フィンガープリント解析)により解析を行った<sup>(49)</sup>。今回、北原らの解析に使用した乳酸菌 (*L. sanfranciscensis* 22 株, *L. plantarum*, *Pediococcus pentosaceus*) 計 24 株についても、GP 法による多様性解析を行い、リボタイピング法との比較を行なった。

現在、全生物種に対して統一的な同定法は GP 法以外には存在しない。rRNA 遺伝子にしても、eukaryote と prokaryote とではその塩基配列構造は 16S rRNA と 18S rRNA もしくは、23S rRNA と 26S または 28S rRNA において大小に異なるために、同一のユニバーサルプライマーによるターゲット DNA 断片を取得することは不可能である。また同一種に対して、変異などの影響により同一プライマーでも断片が増幅されないケースもあり、プライマーを再設計する煩わしさが伴っている。一方、操作的には簡便な RAPD や AFLP などは、近縁な種内変異を観る場合には適しているが、属レベル以上の生物に対しては十分な同定精度が得られず、広範な利用には限界がある。今回、先の項目で使用する *Trichosporon* 属 38 株、*Candida* 属 26 株、*L. sanfranciscensis* を含む他 24 株と新たに *Saccharomyces cerevisiae* 9 株を加えた全 97 株について、GP 法による広域分類を試みた。

## 2 節 実験材料および実験方法

### 2-3 使用菌株

用いた菌株は次のとおりである。千葉大学真菌医学研究センターより真菌の *Trichosporon* 属 16 種 38 株(Table 1)、理研バイオリソースセンター微生物菌株保存施設(JCM)より真菌の *Candida* 属 7 種 26 株(Table 2)および細菌の *L. sanfranciscensis* 22 株、*P. pentosaceus* 1 株と *L. plantarum* 1 株(Table 3)、さらに酒類総合研究所より真菌(酵母) *S. cerevisiae* 9 株(Table 4)の提供を頂いた。

### 2-2 DNA 抽出

ゲノム DNA の抽出は、病原性菌株の *Trichosporon* 属と *Candada* 属については上記専門機関において市販 DNA 抽出キット Gen とるくん™(酵母用)(タカラバイオ)を使用し、そのプロトコールに従った。すなわち、菌体沈殿に GenTLE™ 酵母溶液 A を 500 µl 加えて懸濁した後、37°C で 1 時間、時々混合しながらインキュベートした。次に GenTLE™ 酵母溶液 B 100 µl を加えて穏やかに混合した後、70°C で 10 分間加熱した。その後、GenTLE™ 溶液 C 200 µl を加えて穏やかに混合した後、氷中で 5 分間冷やした。次に 4°C で 12,000 rpm、5 分間遠心した後、沈殿を吸い上げないように注意深く上清を分取し、新しいマイクロチューブに移した。回収した液量の 1/2 量(約 400 µl)のイソプロパノールを加え、転倒混和により十分に混合し、DNA の凝集反応を促した。4°C で 12,000 rpm、5 分間遠心し DNA を沈殿させ上清を捨てた後、500 µl の 70% 冷エタノールの添加により DNA の沈殿を懸濁洗浄

し、再度 4°Cで 12,000 rpm、5 分間遠心した。マイクロピペットで上清を十分取り除き、DNA の沈殿を風乾し、適量の TE バッファーに溶解した。

乳酸菌 *L. Sanfranciscensis* と酵母 *S. cerevisiae* ゲノム DNA 抽出には、市販の ISOPLANT™ (ニッポンジーン) を使用し、そのプロトコールに従った。すなわち、菌体沈殿に溶液 I を 500 µl 加えて懸濁し、続けて溶液 II を 150µl 加えて懸濁後、50°Cで 10 分間加熱した。次に溶液 III-A を 100µl と溶液 III-B を 120µl 加えて懸濁後、氷中で 10 分間冷やした。次に 4°Cで 14,000×g で 10 分間遠心した後、水層を採取し、2 倍量エタノールを加えて混合した。室温にて 6,000×g で 5 分間遠心した後、上清を捨て、沈殿に 70%エタノールを 1ml 加えて混合し、室温にて 6,000×g で 1 分間遠心した。マイクロピペットで上清を十分取り除き、DNA の沈殿を風乾し、適量の TE バッファーで溶解し、PCR 用テンプレート溶液とした。



## 2-3 ゲノムプロファイリング法 (Genome profiling (GP))

GP法の操作手順は次のステップから構成されている。 i) ランダム PCR、  
ii) マイクロ温度勾配ゲル電気泳動法： $\mu$ TGGE、iii) コンピュータデータ処理  
である (Fig. 1 & 2)。

### i) ランダム PCR

ランダム PCR は任意配列のプライマーを単一で用いて比較的低温のアニー  
リング温度条件で行う一般化 PCR である<sup>(31)</sup>。そのために使用するプライマー  
配列に依存してゲノム中の方々からランダムに DNA 断片が増幅・回収される  
(33)。ランダム PCR は RAPD<sup>(16)</sup>や ap-PCR<sup>(17)</sup>と同義であり、これらは同時期か  
つ独自に考案された<sup>(21)</sup>。

*Trichosporon* 属の GP 法においては、プライマーは 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) と、  
B89(5'-Cy3-dACTAACCTGGAC)を個々に用いて行った。

*Candida* 属および乳酸菌 *L. Sanfranciscensis* の GP においては pfm12,  
pfm19, Hunt, B81, TeeGee (5'-Cy3-dTTTGGGTTTGGG) を個々に用いた。

PCR の反応溶液組成は最終濃度として 200  $\mu$ M dNTP (N=G, A, T and C),  
0.5  $\mu$ M プライマー, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1  
ユニットの rTaq DNA ポリメラーゼ (TaKaRa, Japan) と、1  $\mu$ l のテンプレ  
ート DNA とし、全量 30  $\mu$ l とした。サーマルサイクラーは Mastercycler  
gradient (Eppendorf, Germany) を使用して、95°C, 2 分の変性の後、変性

94°C, 15 秒、アニーリング 40°C, 30 秒、鎖伸長 50°C, 30 秒を 30 サイクル行い、最後に 72°C, 30 秒を行った。

## ii) マイクロ温度勾配ゲル電気泳動法 ( $\mu$ TGGE)

温度勾配ゲル電気泳動法 : Temperature gradient gel electrophoresis (TGGE)<sup>(50)</sup>により、その DNA がもつ塩基配列固有の融解温度がわかる<sup>(34)</sup>。しかも現在は TGGE に用いるゲルは 1 インチと小型化されたうえ、その精度は維持したままである<sup>(26)</sup>。本研究では  $\mu$ TGGE 装置として、 $\mu$ -TG (タイテック株) を使用した。

ゲル組成は最終濃度として、6% (w/v) アクリルアミド (19 : 1), 6.5 M 尿素, 1×TBE buffer (0.1 M Tris, 0.09 M ホウ酸, 0.001 M EDTA [pH 8.4]) となるように調製し、 $\mu$ -TG 専用ゲルカセット(タイテック株) に充填し重合させた。

ランダム PCR 産物 6.5 $\mu$ l を新しいマイクロチューブに取り、それに別個に作製した内部標準試料 DNA (IntRefs. 5 and 6 DNA) 1  $\mu$ l と泳動用色素 (6×Loading Buffer Triple Dye ; ニッポンジーン) 1.5  $\mu$ l を加えて混合したものを  $\mu$ TGGE 泳動サンプルとした。次に重合したゲルを含むゲルカセットを専用のゲルホルダーに載せ、正負極それぞれにバッファウィックを入れ、各バッファウィックに 5×TBE 800 $\mu$ l を注いだ。泳動サンプルをウェルにチャージし、低温側 15 °C、高温側 55 °C に設定した  $\mu$ -TG にゲルホルダーセットを載せて、電圧 100V で 9 分間の  $\mu$ TGGE 泳動を行った。

$\mu$ TGGE 泳動後、バンドの検出には蛍光イメージャー Typhoon 9400 (GEヘルスケアサイエンス, USA)を使用した。

内部参照試料 DNA (IntRefs. 5, 6) はプラスミド pBR322 DNA をテンプレ

ートとして、プライマーセット Ref5F(5'-Cy3-dAGTGGTCCTGCAACTTT ATC)と Ref5R(dAACATGGGGGATCATGTAAC)、および Ref6F(5'-Cy3-d GCCGGCATCACCGGCGCCACAGGTGCGGTTG)と Ref6R (dTAGCGAG GTGCCGCGGCTTCCATTCAGGTC) を使用し、次の一般的な PCR 条件により作製した。95°C, 2 分の後、変性 94°C, 15 秒、アニーリング 55°C, 30 秒、鎖伸長 72°C, 30 秒を 25 サイクルの後、72°C, 30 秒とした。この PCR により IntRef. 5: 200bp と IntRef. 6: 900bp の dsDNA 断片が作製された。これらの DNA 断片はそれぞれプラスミド pBR322 DNA 中の 3501-3700 と 401-1300 の部位に由来しており、泳動条件下 6%アクリルアミドゲル(Bis 5%T, 6.5M Urea, 1×TBE), 5×TBE バッファー溶液, 15~55°C 温度勾配, 定電圧 100V, 泳動時間 8 分間で、各々 52.2°C, 61.4°C で融解開始する。

### iii) コンピュータデータ処理

データに再現性と普遍性をもたせるために、内部標準試料 DNA を使用した上で、コンピューター処理で標準規格化することが重要である。蛍光イメージャーで取り込んだ  $\mu$ TGGE の画像ファイルをコンピューター画面上で開き、二本鎖 DNA 断片の第一融解開始点を特徴点として打点すると特徴点群ができる。この特徴点群は、あらかじめ  $\mu$ TGGE で共泳動した 2 種類の内部標準試料 DNA (IntRef. 5, 6) のそれぞれの特徴点を利用することで泳動毎のゆらぎ (装置の影響やゲル毎の不均一性、環境温度など) が排除され、標準規格化された座標上の点群となり、これを *spiddos* (species identification dots) と称している<sup>(25)</sup> (Fig. 2)。そのため実験毎のゆらぎを排除でき、各 *spiddos* の点群は厳密

に比較可能となっている。2 生物種のゲノムの類似度は *spiddos* のパターン類似度 *PaSS* (pattern similarity score) として、以下の式で計算される。

$$PaSS = 1 - \frac{1}{n} \sum_{i=1}^n \frac{|\bar{P}_i^{(1)} - \bar{P}_i^{(2)}|}{|\bar{P}_i^{(1)}| + |\bar{P}_i^{(2)}|} \quad 0 \leq PaSS \leq 1 \quad (5)$$

ここで  $\bar{p}$  は個々の *spiddos* (1 to n) のベクトルであり、温度 T と移動度 m の関数 ( $|\bar{p}| = P(T, m)$ ) である。式(5) 中の肩書き 1 と 2 は、ゲノム 1 とゲノム 2 を表している。両者の *spiddos* が完全一致すると *PaSS* は 1 になる。それぞれの DNA はその塩基配列に依存して移動度と温度の 2 つの変数からなる固有の座標を占める *spiddos* となる。その転移点の座標 (移動度と温度) はそのときに使用したプライマーとテンプレートの配列の組み合わせに依存しており、プライマーは既知であるから、得られる *spiddos* はゲノムテンプレートの塩基配列を反映したものとなっている。

### Phylogenetic analysis

系統樹作成には、UPGMA 法を使用したプログラム Web サイト; DendroUPGMA (<http://genomes.urv.cat/UPGMA/>)<sup>(51)</sup> を利用した。すなわち、*PaSS* 値の Matrix テキストファイルを投じることで、自動で距離  $d_g = 1 - PaSS$  に置換され、距離を基に無荷重平均距離法: UPGMA (Unweighted Pair Group Method using arithmetic) により系統樹テキストファイル形式が得られる。このテキストファイルを系統樹描写ソフトウェア: TreeView<sup>(52)</sup> で系統樹に変換した。

### 3節 結果

#### 3-1 *Trichosporon* 属

*Trichosporon* 属 16 種 38 株の相互関係を表現するために、7 種類のプローブを単独使用して GP を行い、得られた *PaSS* 値からゲノム距離  $d_g (=1-PaSS)$  を求めて系統樹を作成した(Fig. 4; その元データは Fig. 6)。作成した系統樹で、ラベル種名が同一でありながらペアを組まない株、同一グループに属さない株があることが確認された。そのような理由からラベル種名が疑わしいと思われる株、すなわち上から T-20, 12, 14, 13, 15, 16, 17, 35, 19, 34, 11, 36, 33, 18, 21 を色付けした。T-20 は *T. cutaneum* とラベルされている。一方、T-19, T-18, T-21 も *T. cutaneum* とラベルされているが、T-18 と T-21 はペアを組んだが、T-19 は離れて位置した。同様に、共に *T. montevidense* とラベルされている T-35 と T-36 や、*T. moniliiforme* とラベルされている T-34 と T-33 も系統樹の中では離れて位置した。これらのラベル種名が疑わしい株が形成するクラスターの中に *T. coremiiforme* とラベルされる T-17 と T-16 が位置した。この時点ではどれが誤っているか（もしくはいずれも誤っているか）は、判断できなかった。しかし、GP 法は系統樹の作成に有用な *spiddos* 情報だけではなく、泳動パターン全体から多くの情報を得ることができる。つまり温度もしくは変性剤によって引き起こされる二本鎖 DNA の融解過程を表したバンドパターン<sup>(53, 54)</sup>そのものである。

Fig. 7 にプロファイルの一部を示した。*T. cutaneum* とラベルされている T-20 は T-18 (*T. cutaneum*) よりも T-6 (*T. asahii*) に明らかに類似しており、特に図中に示したバンド a, b, c は類似している (Fig. 7-A)。同様なことが

T-35 (Fig. 7-B), T-19 (Fig. 2-5-C), と T-33 (Fig. 7-D) でも認められた。たとえば T-35 と T-19 のバンド d, e, f はともに T-17 および T-16 の *T. coremiiforme* に極めて類似しているし、T-33 のバンド g, h, i は T-23 *T. faecale* のそれと類似していた。これらの異なるラベル種の間で観られる類似バンドは著者や Naimuddin M. らの先行研究で *ccgf* (commonly conserved genetic fragments: 共通保存遺伝子断片) として明らかにされており<sup>(24, 55)</sup>、*ccgf* は定性的かつ主観的ではあるが、これらの疑わしい種同定(とりわけ近縁種間において)の手助けとなっている。T-11 *T. asahii* var. *faecale* はその表現型から *T. faecale* の特徴をもつ *T. asahii* の変種としてラベルされているが、Fig. 7-E で示すように T-11 は *T. asahii* および *T. faecale* のいずれのバンドパターンとも類似していなかった。いずれも *T. asteroides* のラベル名である T-12 ~ T-15 は、プローブ B89 を使用したときのプロファイルが Fig. 7-F に示したようになり、T-13 と T-15 に共通して観られるバンド j は T-12 および T-14 には観られなかった。

今回の系統樹作成では、*Trichosporon* 属のアウトグループとして *Candida* 属の 2 種を加えて作成した。その結果 *Candida* の 2 種はペアを組んで *Trichosporon* 属とは離れて矛盾しない結果となり、Fig. 4 および Fig. 8 の系統樹はその意味で信頼できるものである。その上に対応する *ccgf* の比較を加味し、系統樹で推量された種名を、より支持する結果となっている。

系統樹および *ccgf* 解析からそのラベル種名が疑わしく、かつ真のラベル種名ではないかと推定される種名を Table 5 に示した。これらの推量をさらに確認するため、全 38 株について 26S rDNA の D1/D2 領域の塩基配列解析を行

い、その塩基配列を DDBJ データベースに対して BLAST サーチを行った。その結果を Table 5 中およびシーケンスの一部を Fig. 9 に示した。この結果によれば、GP 法による解析で、そのラベル種名の信憑性が疑われた 6 株(T-11, -19, -20, -23, -33, -35)と別のグループ 4 株(T-12, -13, -14, -15)のグループはすべてミスアサインであることが判明した (T-12 と T-14 はそのまま *T. asteroides* であった)。残りすべての 28 株については GP 法による解析と 26S rDNA シーケンス解析でその信憑性が再確認され、GP 法正当性が示された。このことは、GP 法による解析は菌株データベースでの種同定における迅速スクリーニングに優れていることを初めて実証したことになる。

さらに興味深いことに、Fig. 6 で示すようなゲノム距離の三角マップは、これまでのデータの扱いとは異なる表現であり、同定を確認するための付加情報を与えるものである。たとえば、ゲノム距離としては極めて近いが、系統樹では離れて位置している組み合わせ(白抜きの T-20/T-6, T-20/T-8, T-33/T-23, T-35/T-17, および T-35/T-19 の 5 組)が対角線から離れて位置している。このことは T-6, T-8, T-20 が最終的に *T. asahii* とアサインされて同一種であったこと、同様に T-23 と T-33 が *T. faecale* であったこと、T-17, T-19, T-35 が *T. coremiiforme* であったという Table 5 の推量と完全に一致し、これまでの推量をさらに支持するものとなっている。新たに浮び上がったこととして、図において青線でマーキングした T-4, T-13, T-18, T-29 の 4 株は、ゲノム距離  $d_g$  の値が大きい傾向にあり、異質な振る舞いをとっていることである。これはこれらの株は頻繁な変異を経て、他の *Trichosporon* 属の株から遠ざかったことが示唆される。これは *spiddos* のもつ情報のベクトル性を利用した解析といえ、今

後さらにこの解析法を発展させることが期待される。

同様に T-35 (最終的に *T. coremiiforme* とアサインされた) 株は、他の株とのゲノム距離  $d_g$  が全体的に小さくことから、系統樹の中央に位置しているように観られる。三角距離図から得られるこのような解釈は、他の解析法では困難であり、カルチャーの管理・解析において役立つものと思われる。したがって三角距離図から全生物種のペアの相互の近さについての深い情報を得ることができる。以上、ここで述べた解釈の有用性についてさらなる研究が必要ではあるが、生物種同定において、GP 法によってより豊富な情報が得られることが明らかになった。

### 3-2 *Candida* 属

*Candida* 7種 26株の相互関係を表現するために、5種類のプローブを単独に使用して行った GP から得られたゲノム距離  $d_g (=1-PaSS)$  の平均値を使って系統樹を作成した (Fig. 10; 生データ Fig. 11)。まず、この系統樹を観て気になるのは、C-21 *C. tropicalis* JCM 5941 のみが *tropicalis* のクラスターb から外れて、異種名の C-22 *C. parapsilosis* JCM 1785 とペアを組んだことである。C-21についてそのプロファイルを観ると、プローブ Hunt を使用した場合は、C-21 は他の *C. tropicalis* のパターンとは明らかに異なり、C-22 のパターンと極めて類似した (Fig. 12 および Fig. 13)。両者のゲノム距離 :  $d_g (=1-PaSS)$  は 0.015 と小さく、極めて近いことが伺えた。

この時点で、C-21 のラベル種名 *C. tropicalis* は疑わしく、真の種名は *C. parapsilosis* と推測された。この追跡調査として、C-21 の 26S rDNA D1/D2



ドメインについて塩基配列解析を行い、DDBJ データベースに対して BLAST サーチを行った。その結果、C-21 は *C. parapsilosis* とアサインされ、これまでのラベル種名 *C. tropicalis* はミスラベルであったことが判明した。

次にプローブ pfM12 を使用した場合のプロファイルにおける *C. albicans* と *C. tropicalis* の間で特徴的な *ccgf* と見られるバンド k に注目し (Fig. 14)、塩基配列解析を行った。μTGGE 後、C-1 *C. albicans* および C-13 *C. tropicalis* のゲルからそれぞれバンド k を含むゲル断片を切り出し、その断片からバンド k の DNA を抽出した。バンド k の DNA の両端は pfM12 の配列になっているため、バンド k の DNA をテンプレートとして PCR を行う場合、pfM12 は完全相補のプライマーとなっている。ランダム PCR ではアニーリング温度は低めの 40°C で行っていたが、ここにおける re-PCR ではアニーリングは 60°C に設定し、サイクル数を 10 サイクルで行った。バンド k の DNA 単離の確認を μTGGE で行った (Fig. 15)。バンド k の DNA の単離が確認されたことから、次のステップとしてバンド k の DNA を TA クローニングし、その後塩基配列決定をした。Fig. 16 は C-1 および C-13 のバンド k の塩基配列を Clustal W で解析したものである。バンド k はそのサイズが 571bp で、C-1 と C-13 との間では 10 塩基の違いがあり、そのホモロジーは 98.2% であった。このことからバンド k が *ccgf* であるとした推定は正しいことが示された。またこの配列を DDBJ データベースに対してクエリーしたところ、25S rRNA 遺伝子の配列であることがわかった。

プローブ毎の *spiddos* と系統樹を Fig. 17~21 および Fig. 22~26 に示した。まず、使用したプローブの種類に応じて同一ゲノムに対して *spiddos* の変化が

見られた。複数の株がある *C. albicans* と *C. tropicalis* については、プローブ pfM19 を使用したときは *C. albicans* は各株のパターンの違いが顕著に表れている一方で、*C. tropicalis* の各株のパターンの違いはさほど大きくなく、いずれも類似していた(Fig. 18)。このときの系統樹は、*C. albicans* の 12 株は 1 つのクラスターにまとまらずに 2 つのクラスターに分かれた(Fig. 23)。これは *C. albicans* のゲノムの中には、株の多様性を特徴づけるシーケンス領域があつて、その領域をプローブ pfM19 はサンプリングしてきた可能性が考えられる。このように GP 法ではプローブを換えることで簡便に、同一ゲノム (個) に対して多面的なプロフィールの観測が可能である (このことがゲノムプロファイリングという名称の由来) し、集団に対しても多面的な観測が可能であることが確認された。

### 3-3 *L. sanfransicensis*

北原らのリボタイピングによる先行研究<sup>(49)</sup>で用いられた *L. sanfransicensis* 22 株、およびアウトグループとして用いられた *L. plantarum* C2-32 と *P. pentosaceus* F3-24 の系統関係について、5 種類のプローブ (pfM12, pfM19, Hunt, TeeGee, B81) を単独に使用して行った GP 法から得られた *PaSS* 値からゲノム距離  $d_g (=1-PaSS)$  の平均値を使って UPGMA 法で作成した系統樹を Fig. 27 に、北原らがリボタイピングから作成した系統樹を Fig. 28 に示した。GP 法では *L. sanfransicensis* 22 株は 3 つの Cluster A~C に分かれた。一方リボタイピングでは 4 つの Cluster I~IV に分かれたとしている。Cluster A は Cluster I に、Cluster B は Cluster II、Cluster C は Cluster III + IV に対応していた。リボタイピングでは *L. sanfransicensis* JCM 5668<sup>T</sup> は、他の株とはペアを組まず、他とは異質のように思われるが、GP では *L. sanfransicensis* JCM 12444 とペアを組むなどの違いはあるが、大方類似した系統関係を示していた。解析法の違いにより、系統樹のトポロジーが多少変化することは一般的と考えられている<sup>(11)</sup>。

### 3-4 GP 法による大局的系統樹作成

現在、その rRNA 遺伝子の配列構造の違いから、真核細胞生物(eukaryote)と原核細胞生物(prokaryote)を同時に系統解析にかけられる手法は存在しない。今回、GP 法の本質的ロバスト性から大域的分類同定が可能であることの一証左として、これまでに用いた *Trichosporon* 属 38 株、*Candida* 属 26 株、*L. sanfransicensis* を含む他 24 株と新たに *Saccharomyces cerevisiae* 9 株

(Table 4)を加えた全 97 株について、4 種類のプローブ (pfM12, pfM19, Hunt, B81) を個々に用いた GP 法から得られたゲノム距離  $d_g$  (= 1-*PaSS*)の平均値を基に系統樹作成を試みた。その結果を Fig. 29 に示す。併せて 4 種の各プローブ単独 GP で得られた  $d_g$  値で作成した系統樹も Fig. 30-1~30-4 に示す。

まず、第一に細菌界と真菌界で大きく二分する系統樹になると当初予想していたが、その予想に反して *L. sanfranciscensis* の Cluster が *S. cerevisiae* の Cluster とペアを組み、そこに *Candida* 属の種毎の Cluster が覆いかぶさって異なる界が混合した大きな Cluster を形成し、それと *Trichosporon* 属の Cluster が二分する系統樹の結果となった。

属あるいは種毎にみると、まず、*Trichosporon* 属は同一種名の株はペアあるいは Cluster を形成し、3-1 *Trichosporon* 属のところで前述した種名ラベルの誤りを指摘したときと同様なペアリングをした (T-11, T-19, T-20, T-33, T-35)。*T. asteroides* の T-12 と T-14 がペアを組まなかった。これは 3-1 *Trichosporon* 属のときは系統樹作成に使用したプローブ数が 7 種類であったのに対し、今回は 4 種類と少なかったため精度が不十分であったと思われる。

*Candida* 属については、前述と同様に C-21 *C. tropicalis* JCM 5941 が C-22 *C. parapsilosis* JCM 1785 とペアを組み、ラベル種名の誤りを指摘できた。ところが、乳酸菌 L-23 *L. plantarum* C2-32 が *Candida* 属の Cluster に入るといふ不自然な動きも観られた。L-23 はその属名からして *L. sanfranciscensis* の Cluster の近くに位置するのが自然と思われるが、距離をとったところに位置した。各プローブで作成した系統樹 Fig.30-1~30-4 においても L-23 *L. plantarum* C2-32 は *L. sanfranciscensis* とは離れて位置している。このこと

からすると、単にエラーによるものではないと思われるが、その原因追跡は今後の課題である。

系統樹の位相でいくつかの矛盾はあるにせよ、ここで重要なのは、異なる属以上（今回のように異なるドメイン・界）の生物に対して、機械的に同一の操作で簡便に生物種（株）の同定・分類が、十分な情報量と一定の信頼性をもって可能であることを示したことである。すなわち、多種多様な菌株を管理する CC にとって、菌株管理ツールとしての GP 法の有効性を明らかにしたといえる。

#### 4 節 考察

今回、対象とした真菌、すなわち *Trichosporon* および、*Candida* の GP 法によるクラスタリングで、実際のカルチャーコレクション(CC)のラベル種名の誤記を偽陰性も偽陽性もなく**完璧な精度**で明らかにした。更に 26S rDNA の塩基配列決定をして、その結果を確証した。GP 法は *spiddos* だけではなく、*ccgf* の情報<sup>(24, 55)</sup>をも得られ、近縁種間の比較に有効であることも確認できた。便利なことに *ccgf* の DNA 断片は、塩基配列の比較において非常に便利なサンプルである。つまり、シーケンシング用に調製する DNA は、ランダム PCR で使用されたプライマーと  $\mu$ TGGE ゲルから切り出した *ccgf* バンドを鋳型にそのまま PCR を行なって調製できるからで、既に用意できているようなものである。通常のシーケンシングのような PCR プライマーを設計する手間が省略出来ている<sup>(55)</sup>。GP 法の推量の正否を確認するために、慣例的な方法の D1/D2 26S rDNA のシーケンシングを今回実施した。CC において注目すべき生物が出てきた場合に詳細な解析、すなわちシーケンシングが必要となるが、*ccgf* を使ったシーケンシングのほうが日常業務においてずっと扱いやすいアプローチになると考えられる。

今回の実験では経験的に使われてきた 7 種類の異なるプローブを用いてゲノム距離を得た。結果としてそれらのプローブはこの *Trichosporon* 属には有効であっただけでなく、必要であった。今回の実験で使われたプローブプライマー数は、任意であった。既に理論的にも経験的にも示されているが、プローブ数が多ければ多いほど GP 法ではゲノムの全体像を理解することができる。したがって、できるだけ多くのプローブを用いるようにした。

しかしながら、1種類のプローブの実験（いつもは pfM12 を使用するが、今回対象の真菌、すなわち *Trichosporon*、*Candida* 属についてはプライマー Hunt がベストであった）でも種を分類するのに十分であった<sup>(27, 56)</sup>。つまり、これは1回の GP 法の操作で得られる情報量が既に十分に多いことを意味している<sup>(31)</sup>。このことは統計学的に道理にかなっている。つまり、情報量は実験操作の数にほぼ比例して増す。実験操作に要する労力（今回はサンプル毎に7種類のプローブを使った操作）は、シーケンシングに要する労力と比しても、ずっと少ない(Table 6, 7)。最適なプローブのタイプと数は、対象とする生物種に応じて決まってくると思われる。経験的に菌類を含め大抵のケースでは、正確な種同定には1種類もしくは2種類のプローブを使用した実験操作で事足り、大幅に労力を抑制できる。

大規模な CC や従来の微生物学実験教室のような小規模の CC でも抱えている課題は、低コストと低人件費で十分な信頼性を確保しつつ、カルチャーの品質チェックを日常業務化することである。ある意味では、これらの施設ではコスト効果が最優先される。Table 6, 7 で GP 法とシーケンシングの労力とコストについての試算を行なった。これによれば、実験環境に依存するところもあるが、シーケンシングに較べて GP 法のコスト効果は5倍以上、時間は1.7倍抑制できるという結果であった。GP 法のコスト効果と迅速化の原因は、i) シーケンシングでは十分なレベルまでに PCR 産物の精製が必要となってくるが、GP 法ではそれが必要なく、その分手間が省け、細胞からの簡易 DNA 抽出だけでよく、その後のランダム PCR を行い、次にランダム PCR 産物を精製せずに TGGE に処するだけでよい。ii) TGGE 解析は、

従来のゲル電気泳動法よりも1回の操作で得られる情報量が格段に多いためである<sup>(31)</sup>。またゲノム情報生成技術として一定レベルを維持するために、TGGEはその属性から実験的ゆらぎの問題があったが、その実験的ゆらぎを排除する内部参照試料の導入<sup>(25, 38)</sup>により、高い信頼性と再現性のデータ生成を可能にしたことも併せて重要な要素である。

さらにGP法の別のメリットとして、ランダムPCRに用いるプローブプライマーを換えることで、効率的に無制限の情報量が得られることも重要な点である。例えば、今回の研究で使用した塩基長12merのプライマーの場合では $4 \times 10^{12}$ つまり $1.6 \times 10^7$ を超える種類のプローブがあるということになる。今回の研究で使用したプローブ数は、その膨大な種類の中から、たった7種類あるいは5種類である。もし必要ならば、プローブの種類を換えることで、同一の操作により、十分なレベルのまったく別の情報を得ることが可能である。したがって、rDNA領域で分類・同定が出来てない場合にシーケンシングアプローチで慣例的に行なわれているcytochromeやoxidase、gyrase等のような遺伝子をターゲットに新たにプライマーを設計・合成する必要はないのである。その上、GP法で得られたDNA断片は、ランダムPCR時に使用したプライマーで容易にシーケンシングが可能である。つまりre-PCRを行うことであるから確実にシーケンシングができるという具合である。これらのGP法の特徴は、頻繁かつ大規模な解析が必要となる大規模CCにおいては、特に便利であるといえよう。

いずれの方法論でもそうであるが、GP法にも技術的な限界はある。すなわち、*PaSS*値(あるいはゲノム距離)計算上の確率論的特性を有しているこ



とである。2 生物種が属内のような近縁種であれば確率論的よりはむしろ決定論的になるが (*spiddos* の対応関係がユニークに定まるため)、遠縁種の場合はその効果は顕著となる<sup>(56)</sup>。しかしながら、Ahmed らの研究(この時、昆虫(トンボ、バッタ、セミ、蝶、カブトムや他の分類群)を用いた例では、古典的な表現型に基づいた方法で得られた分類結果と GP 法による結果が一致した)により、GP 法の効果の応用が効く範囲は、意外にも広範囲であることが経験的にわかっている<sup>(57)</sup>。今回の研究の成功事例をとおして、GP 法の適用範囲の限界よりも *Trichosporon* 属の種や *Candida* 属の種は遺伝的に近いことが示された。そのような限界の理論的扱いができるようにするためには、この便利な GP 法を遠縁関係にある種のクラスタリングに適用するには特に注意を払わなければならない。この適用範囲が意外にも広い背景には生物の進化系統の構造問題があり、興味深い。幸運にもこれまでに蓄積された結果には、GP 法が真菌や細菌を含むより広範囲に適用できるという楽観的な期待が支持されてきた。当然ながら全プロセスが本質的に確率論的ではなく決定論的なプロセスで支配されている近縁種に対して GP 法を適用することは問題ないことは、明白である<sup>(28)</sup>。

Fig. 4 と Fig. 5 で、GP 法と D1/D2 26SrDNA シーケンシングによる *Trichosporon* 属の系統樹は似たようになったが、同一ではなかった(D1/D2 26SrDNA シーケンシングから作成された系統樹の Cluster I の中に *T. asahii* からなる Cluster A が一部として入っている)。これはそれぞれの方法論の違いによるものであり、系統樹作成において用いる方法によって系統樹の形状が同一にならないのはごく普通のことである<sup>(41)</sup>。注目すべきは、

D1/D2 26SrDNA シーケンシングの結果である Fig.5 中の Cluster I では、各株の差異はほとんどない。つまり点突然変異レベルであるのに対し、差異の程度がどのくらいであるかの未解決部分があるものの GP 法の解析から得られた Fig.2 の系統樹では、株の差異が明確に表れている。これまでに *Trichosporon* 属の種間の違いは ITS<sup>(41)</sup>や *Cyt b*<sup>(58)</sup>, 18S rDNA<sup>(59)</sup>, D1/D2 26S rDNA<sup>(60)</sup>, IGS<sup>(61~63)</sup>のシーケンス解析で行われてきた。一方、種内変異の解析には RAPD<sup>(64)</sup>, AFLP<sup>(65)</sup>, MLST<sup>(66)</sup>の解析法が利用されてきた。これらの解析法は種間の差異を観るのか、それとも種内(株)の差異を観るのか、その目的に応じて使い分けられ、統一的な標準方法とはなっていない。今回、GP 法単独で種間も種内も同時に分類できた結果を示すことができたのは非常に意義が大きい。

カルチャーコレクションにおいて GP 法の利点を活かす最適な使用法は、次の 2 点が考えられる。すなわち、1) 日常業務における全生物を対象とした予備同定 と、2) Fig. 6 の三角距離図で示したように、また *ccgf*のシーケンシングを含め、複数のプローブを使用して解析の深化を行うこと である。今回の研究において、GP 法の解析から得られたデータは他の手法から得られたデータと一致し、補完するものとなっていることから明らかなように、GP 法の使いやすさと応用性を考慮すると、ほとんどの CC にとって標準アプローチとして最適であると考えられる。

結論として、GP 法によりカルチャーコレクション(CC)における真菌では *Trichosporon* と *Candida* を、細菌では *Lactobacillus* を種および株のすべてにわたって正確に同定・分類ができたことを確認した。そのなかで、

*Trichosporon* と *Candida* の株の中にそのラベル種名の誤りがあるものを指摘し、その指摘が正しかったことも 26S rDNA シーケンス解析で確認することができ、GP 法の解析能力の高さを示すことができた。さらに GP 法の解析用途として、三角距離図は見た目は距離のあると思われた種が意外にも近いことを明らかにすることができ (T-35 *T. montevidense* → *T. coremiiforme*)、*ccgf* はシーケンスレベルの容易な比較を可能とすることを示すことができた。そして、真菌および細菌を含め、GP 法は株レベルからドメイン (界) レベルを超える広域の生物を対象に比較することが可能であることを示すことができた。

### 3章 総括

カルチャーコレクション(CC)において菌株の品質維持・管理は極めて重要であり、生命線とも言える。品質維持・管理のためのチェック方法、つまり同定法には、形態観察をはじめ、生理・生化学的手法等の古典的な表現型によるものと、PFGE の他、そのベースに PCR 技術を利用した遺伝子型によるものがある。CC の日常業務において表現型の同定法を用いるのは大変な労力と時間がかかり、PCR 技術が浸透した現在、品質チェック法としては相応しいとは言い難い。PCR 技術を利用した同定法には 16S rDNA<sup>(6)</sup>, 18S rDNA<sup>(7)</sup>, 26S/28S rDNA<sup>(8)</sup>, ITS<sup>(11)</sup>, IGS<sup>(12)</sup>の rDNA 領域や、特定遺伝子 *gyr B*<sup>(13, 14)</sup>、*rpoD*<sup>(14)</sup>、*Cyt b*<sup>(15)</sup> 等の House keeping 遺伝子のシーケンス解析による同定法があるが、これらによる同定は種レベルである。一方、RAPD<sup>(16, 17)</sup>、AFLP<sup>(18)</sup>、MLST<sup>(20)</sup>は株レベルが対象である。実際の同定は求めるレベル(属, 種, 株)に応じて方法を選択しており、すべての菌株に共通した同定法はこれまでなかった。

今回、*Trichosporon* および *Candida* 株を使った研究では、GP 法から得られたゲノム距離を使って、種レベルと株レベルを一括してクラスタリングした系統樹を作成することができ(Fig. 4, Fig. 10)、その系統樹からラベル菌種名の誤りを指摘することに成功した。併せてゲノムプロファイル(GP)に検出された共通保存遺伝子断片(*ccgf*)からもラベル菌種名の誤りを確認した。追跡調査として D1/D2 26S rDNA 解析を行なった結果、GP 法による指摘が正しいことを追認した(Table 5)。その後、菌株を保有するそれぞれの CC において正式にラベル菌種名の変更が実行された(当該センターの Web に公

示)。

また、今回ゲノム距離マトリックスを一定の距離範囲ごとに色分けを施した三角距離図(Fig. 6)を作出することで、一見距離のあると思われた種が以外にも近い関係にあることを示す新たな事実を浮上させ、GP法の解析方法としての用途幅の広さを示すことができた。

すべての菌株に共通した同定・分類法がない現在、真菌および細菌を含めた株レベルからドメイン(界)レベルの広域生物を対象にGP法による一括分類が可能であることを示すことができた(Fig. 29, Fig. 30-1~30-4)。

同定に要する1サンプル当たりのランニングコストは、GP法は約230円でシーケンシングの約1/5にとどまり、検出までの時間は約4時間でシーケンシングの約1/2である試算となり(Table 6, Table 7)、同定精度を犠牲にすることなく、コストパフォーマンスを上げていることが示された。

したがって今回の結果は、CCにおける菌株管理ツールとしてのGP法の有効性を実証した結果となった。*Trichosporon* 38株中7株および*Candida* 26株中1株のミスラベルが今回判明した事実からすると、少ない人員で多数の菌株を従来法で管理しているために<sup>(67)</sup>、ミスラベルを見落とししたり、あるいは手付かずの状態にある菌株が存在する事情は容易に想像がつく。したがって、GP法を従来の同定・分類法と併せてCCで用いるならば、正確かつ統一された菌株の管理が可能となり、菌株のクォリティーの向上が見込まれる。

近年、次世代シーケンサー(NGS)の登場によりシーケンシングスピードは驚異的に向上している。NGSは単に塩基配列解読にとどまらず、a) 転写物

(トランスクリプト)を明らかにするトランスクリプトーム解析、b) 転写開始点の特定やゲノムワイドな遺伝子発現パターンの解析、c) DNA-タンパク質相互作用の解明、d) メチル化, アセチル化等のゲノム DNA の修飾といったエピゲノミクス解析、e) タンパク質間相互作用の解明、f) ゲノム DNA の 3 次元空間における配置の解明、g) ヒトの腸内細菌の動態や難培養性を示す菌などのメタゲノミクス解析 といった分子情報の取得が可能であり、その用途は多様かつ高次元である<sup>(68)</sup>。しかし、一方、微生物種同定に必要な情報量は、1 章 2-2-3 で議論したようにわずか 200 文字(200 ビット)の配列であり、全ゲノムシーケンスで種同定を行うことは情報過多である。技術コスト論からして、現状において NGS を菌株管理に用いることは不経済であり、現実的でない。浦和から東京に行くのに超音速ジェット機を用いる必要はなく、JR で十分であるのに類比される。

コスト問題だけでなく、GP 法では普遍的でコンパクトな *spiddos* というゲノム情報を与え、これを検索指標に用いることで生物界を横断するデータベース構築が可能であり<sup>(56)</sup>、まさに CC にとって基本的重要事項(試料の一元的管理)を提供していることになる。

以上、本論文において、CC において GP 法を用いる事の合理性、経済性、必然性を議論した。

## 謝辞

はじめに、今回の研究論文の元となる「埼玉バイオプロジェクト」(2003年4月～2006年3月)に参画の機会を与えていただいたタイテック株式会社桂澤隆会長に深く感謝申し上げます。また、研究材料のご提供とご助言をいただいた千葉大学真菌医学研究センターの西村和子名誉教授、田中玲子助教と理化学研究所バイオリソースセンター元専任研究員 鈴木基文先生に深く感謝申し上げます。

学位申請に当たり長期間にわたりご指導・ご教授下さいました埼玉大学大学院理工学研究科 西垣功一教授に心より感謝・御礼申し上げます。

そして副査を引き受けて下さいました埼玉大学大学院理工学研究科 松岡浩司教授、根本直人准教授、藤野毅准教授、齋藤伸吾准教授に御礼申し上げます。

最後に家族の協力を深く感謝申し上げます。

## 引用文献

- (1) 平成 25 年版 環境・循環型社会・生物多様性白書 環境省 139 頁
- (2) 飯島貞二. カルチャーコレクションのソフトとハード *Microbiol. Cult. Coll.* 2001; 17: 63-65.
- (3) 鈴木健一郎. 細菌分類学とカルチャーコレクションの役割 *Microbiol. Cult. Coll.* 2009; 25(2): 79-88.
- (4) 中瀬 崇. カルチャーコレクションの運営に思うこと *Microbiol. Cult. Coll.* 2001; 17: 69-72.
- (5) 渡邊 信. シアノバクテリアとカルチャーコレクション *Microbiol. Cult. Coll.* 2007; 23: 83-88.
- (6) Carl R. Woese and George E. Fox Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. USA* 1977; 74(11): 5088-5090.
- (7) Suh S.-O. and Nakase T., Phylogenetic analysis of the ballistosporous anamorphic genera *Udeniomyces* and *Bullera* and related basidiomycetous yeasts, based on 18S rDNA sequences. *Microbiology.* 1995; 141: 901-906.
- (8) Sugita T., Takashima M., Ikeda R., Nakase T. and Shinoda T., Intraspecies diversity of *Cryptococcus laurentii* as revealed by sequences of internal transcribed spacer regions and 28S rRNA gene and taxonomic position of *C. laurentii* clinical isolates. *J. Clin. Microbiol.* 2000; 38: 1468-1471.



- (9) Collins, MD, Williams, AM, and Wallbanks, S. The phylogeny of *Aerococcus* and *Pediococcus* as determined by 16S rRNA sequence analysis: description of *Tetragenococcus* gen. Nov. *FEMS Microbiol. Lett.* 1990; 70: 255-262.
- (10) Devriese, LA, Pot, B, and Collins, MD. Phenotypic identification of the genus *Enterococcus* and differentiation of phylogenetically distinct enterococcal species and species groups. *J. Appl. Microbiol.* 1993; 75: 399-408.
- (11) Sugita T., Nishikawa A., Ikeda R., Shinoda T. Identification of medically relevant *Trichosporon* species based on sequences of internal transcribed spacer regions and construction of a database for *Trichosporon* identification. *J. Clin. Microbiol.* 1999; 37: 1985-1993.
- (12) Sugita T., Nakajima M., Ikeda R., Matsushima T., Shinoda T. Sequence analysis of the ribosomal DNA intergenic spacer 1 regions of *Trichosporon* species. 2002; 40(5): 1826-1830.
- (13) Huang WM. Bacterial diversity based on type II DNA topoisomerase genes. *Annu Rev Genet.* 1996; 30: 79-107.
- (14) Yamamoto S, Harayama S. Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol.* 1998; 48(3): 813-819.

- (15) Biswas SK, Wang L, Yokoyama K, Nishimura K. Molecular analysis of *Cryptococcus neoformans* mitochondrial cytochrome b gene sequences. *J Clin Microbiol.* 2003; 41(12): 5572-5576.
- (16) Williams J. G. K, Kubelik A. R., Livak K. J., Rafalski J. A., and Tingey S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids. Res.* 1990; 18: 6531-6535.
- (17) Welsh J. and McClelland M. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids. Res.* 1990; 18: 7213-7218.
- (18) Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., et al. AFLP: a new technique for DNA fingerprinting". *Nucleic Acids Res.* 1995; 23 (21): 4407-4414.
- (19) David C., Schwartz and Cantor C.R., Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell.* 1984; 37 (1): 67-75.
- (20) Maiden MC., Bygraves JA., Feil E., Morelli G., Russell J.E., Urwin R., Zhang Q., Zhou J., Zurth K., Caugant DA., Feavers IM., Achtman M., Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA.* 1998; 95(6): 3140-3145.
- (21) Nishigaki K., Amano N., Takasawa T., Kinoshita Y., Husimi Y. DNA profiling: method and principle. *Seibutu-buturi* 1990; 30: S230.

- (22) Nishigaki K., Amano N. and Takasawa T. DNA profiling: An approach of systematic characterization, classification and comparison of genomic DNAs. *Chem. Lett.* 1991; 1991: 1097-1100.
- (23) Nishigaki K., Miura T., Tsubota M., Sutoh A., Amano N. and Husimi Y. Structural analysis of nucleic acids by precise Denaturing Gradient Gel Electrophoresis: II. Applications to the analysis of subtle and drastic mobility change of oligo- and polynucleotides. *J. Biochem.* 1992; 111: 151-156.
- (24) 浜野圭一 DNA プロファイリングにおける共通保存遺伝子断片(CCGF)の実証とそのゲノム分類学への応用. 埼玉大学修士論文 1995
- (25) Naimuddin M., Kurazono T., Zhang Y., Watanabe T., Yamaguchi M., Nishigaki K. Species-identification dots: a potent tool for developing genome microbiology. *Gene.* 2000; 261: 243-250.
- (26) Biyani M., Nishigaki K. Hundred fold productivity of genome analysis by introduction of micro-temperature-gradient gel electrophoresis. *Electrophoresis.* 2001; 22(1): 23-28.
- (27) Kouduka M., Sato D., Komori M., Kikuchi M., Miyamoto K., Kosaku A., Naimuddin M., Matsuoka A, Nishigaki K. A solution for universal classification of species based on genomic DNA. *Int J Plant Genomics.* 2007; Article ID 27894: 8 pages.
- (28) Futakami M., Salimullah M., Miura T., Tokita S., Nishigaki K. Novel mutation assay with high sensitivity based on direct measurement of

- genomic DNA alterations: comparable results to the Ames test. *J. Biochem.* 2007; 141(5): 675-686.
- (29) 西垣功一, 生物物理, 1986; 25: 39.
- (30) 浜野圭一, 高沢 努, 倉園貴至, 奥山雄介, 西垣功一 ゲノムプロファイリング-その方法論の確立と実践的評価, 日化誌, 1996; 54-61.
- (31) Nishigaki K., Mohammed N. and Hamano K. Genome Profiling: A Realistic Solution for Genotype-Based Identification of Species. *J. Biochem.* 2000; 128: 107-112.
- (32) Sakuma Y., Nishigaki K. Computer Prediction of General PCR Products Based on Dynamical Solution Structures of DNA. *J. Biochem.* 1994; 116 (4): 736-741.
- (33) Nishigaki K., Saito. A., Hasegawa T. and Naimuddin M. Whole genome sequence-enabled prediction of sequences performed for random PCR products of *Escherichia coli*. *Nucleic Acids Res.* 2000; 28: 1879-1884.
- (34) Wada A., Yabuki S., Husimi Y. Fine structure in the thermal denaturation of DNA: High temperature-resolution spectrophotometric studies. *CRC Crit. Rev. Biochem.* 1980; 9(2): 87-144.
- (35) Abrams ES, Stanton VP., Jr Use of denaturing gradient gel electrophoresis to study conformational transitions in nucleic acids. *Methods Enzymol.* 1992; 212: 71-104.
- (36) Nishigaki K., Husimi Y., Masuda M., Kaneko K., and Tanaka T.

- Strand dissociation and cooperative melting of double-strand DNAs detected by denaturant gradient gel electrophoresis. *J. Biochem.* 1984; 65: 627-635.
- (37) 江崎孝行 病原細菌の系統保存活動から見えてきた菌種の再定義への課題. *Microbol. Cult. Coll.* 2008; 24(2): 81-86.
- (38) 特許 5024784 被検体生物の同定方法、この方法に使用する内部標準用 DNA 組成物及びその製造方法. 2012.
- (39) 時松一成, 門田淳一. 深在性トリコスポロン症 トリコスポロン症の臨床, *感染症学雑誌* 2006; 80: 196-202.
- (40) 時松一成 血液疾患の真菌症感染～最近の傾向と対策～血液フロンティア 2007; 17: 1357-1364.
- (41) Guého E., Improvisi L., de Hoog G.S., et al. *Trichosporon* on humans: a practical account. *Mycoses*, 1994; 37: 3-10.
- (42) Guého E., Smith M. T., de Hoog G. S., et al. contributions to a revision of the genus *Trichosporon*. *Antonie Van Leeuwenhoek*, 1992; 61: 289-316.
- (43) Sugita T., Nishikawa A. and Shinoda T. Reclassification of *Trichosporon cutaneum* by DNA relatedness by using on the spectrophotometric and chemiluminometric method. *J. Gen. Appl. Microbiol.* 1994; 40: 397-408.
- (44) Sugita T., Nishikawa A., Shinoda T., et al. Taxonomic position of deep-seated, mucosa-associated, and superficial isolates of

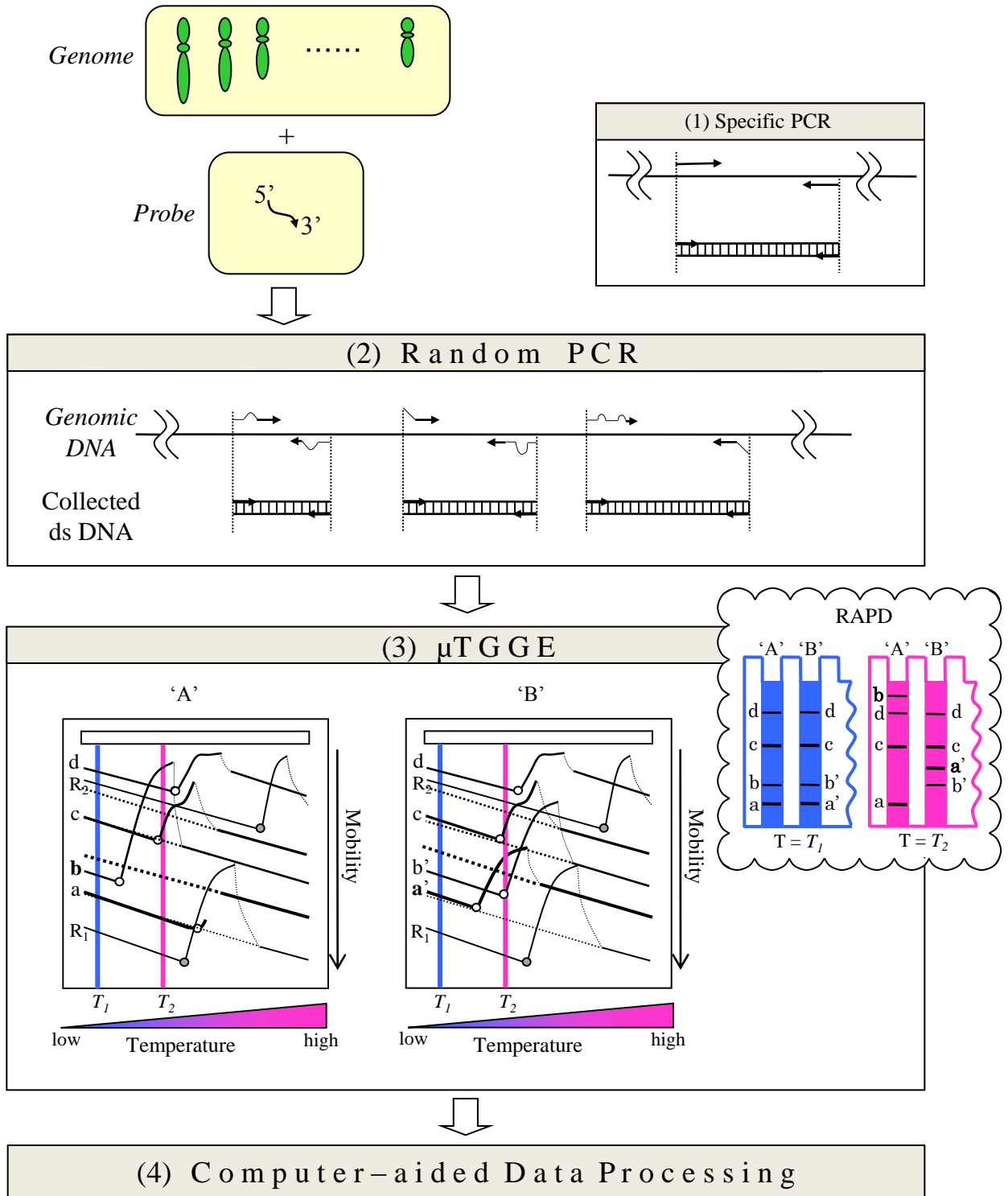
- Trichosporon cutaneum from trichosporonosis patients. *J Clin. Microbiol.* 1995; 33: 1368 -1370.
- (45) Sugita T., Nakajima M., Ikeda R., et al. Sequence analysis of the ribosomal DNA intergenic spacer I regions of Trichosporon species. *J. Clin. Microbiol.* 2002; 40: 1826-1830.
- (46) Middelhoven W. J., Scorzetti G. and Fell J.W. 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.* 2004; 54: 975-986.
- (47) Sugita T., Ikeda R. and Nishikawa A. Analysis of Trichosporon isolates obtained from the houses of patients with summer-type hypersensitivity pneumonitis. *J. Clin. Microbiol.* 2004; 42: 5467-5471.
- (48) 西川朱實 カンジダの菌学 *Jpn. J. Med. Mycol.* 2007; 48: 126-128.
- (49) Kitahara M., Sakata S. and Benno Y. Biodiversity of Lactobacillus sanfranciscensis strains isolated from five sourdoughs. *Lett. Appl. Microbiol.* 2005; 40: 353-357.
- (50) Thatcher D. and Hodson B. Denaturation of proteins and nucleic acids by thermal-gradient electrophoresis. *Biochem. J.* 1981; 197: 105-109.
- (51) Garcia-Vallve S., Palau J. and Romeu A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli*. and *Bacillus*

- subtillis. Mol. Biol. Evol.* 1999; 9: 1125-1134.
- (52) Page R. D. M. Tree View: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 1996; 12: 357-358.
- (53) Nishigaki K., Husimi Y., Masuda M., Kaneko K. and Tanaka T. Strand dissociation and cooperative melting of double-stranded DNAs detected by denaturant gradient gel electrophoresis. *J. Biochem.* 1984; 95: 627-35.
- (54) 坪田美佐, 西垣功一, 伏見讓. 変性ゲル電気泳動の精密化の試み. 生物物理. 1986; 26: 3E1145.
- (55) Naimuddin, M., T. Kurazono, and K. Nishigaki. Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution. *Nucleic Acids Res.* 2002; **30**: e42.
- (56) Watanabe, T., 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* 2002; **3**: research0010.1-0010.8.
- (57) Ahmed S., Komori M., Tsuji-Ueno S., Suzuki M., Kosaku A., Miyamoto K., Nishigaki K. Genome Profiling (GP) method based classification of insects: congruence with that of classical phenotype-based one. *PLOS one* 2011; 6, 8: e23963.
- (58) Biswas S. K., Wang L. Yokoyama K. and Nishimura K. Molecular phylogenetics of the genus *Trichosporon* inferred from mitochondrial cytochrome b gene sequences. *J. Clin. Microbiol.* 2005; 43: 5171-5178.

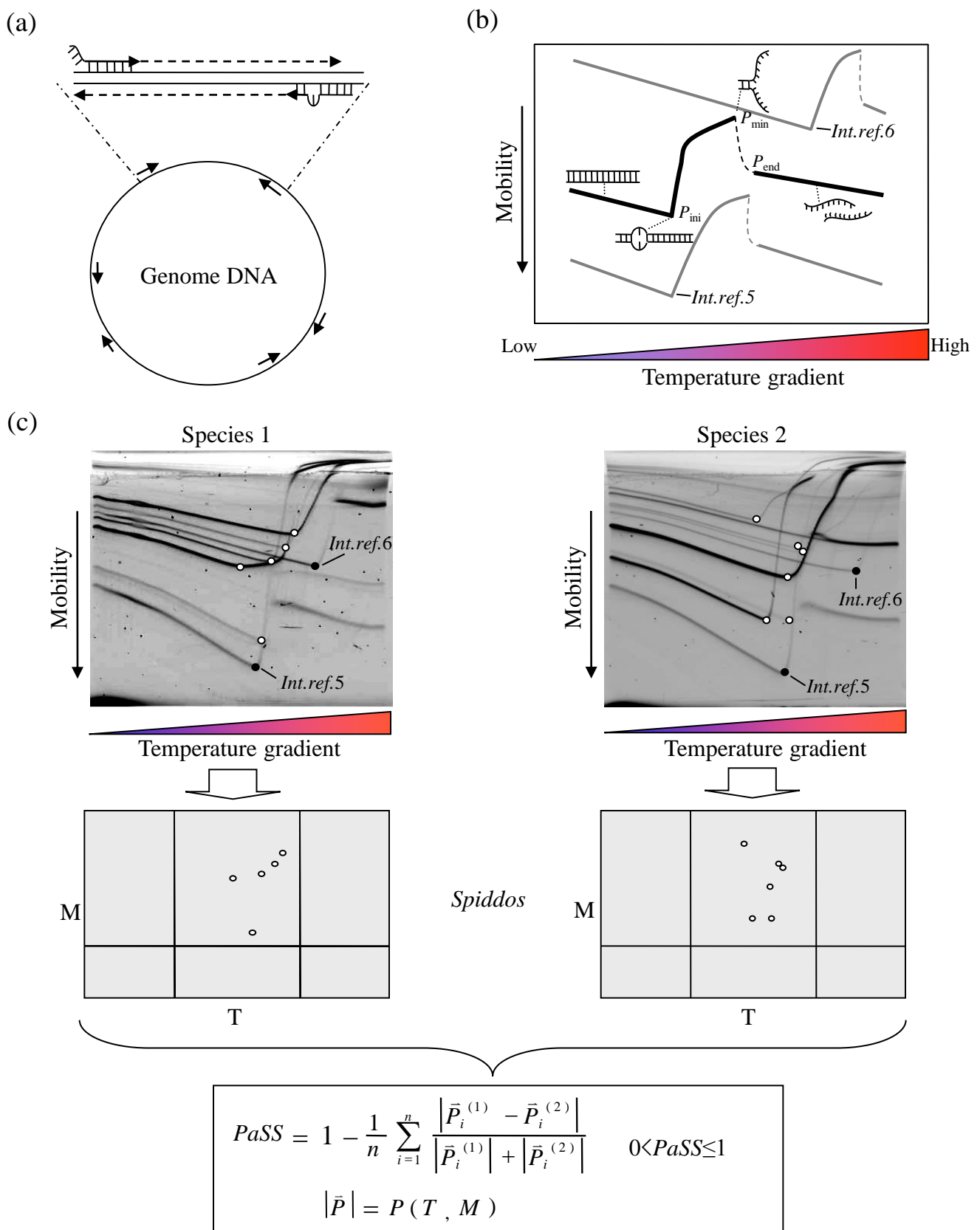
- (59) Takashima M., Nakase T., Molecular phylogeny of the genus *Cryptococcus* and related species based on the sequences of 18S rDNA and internal transcribed spacer region. *Microbiol. Cult. Coll.* 1999; 15: 35-47.
- (60) Fell J. W., T. Boekhout T., Fonseca A., Scorzetti G., Statzell-Tallman A. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *Int. J. Syst. Bacteriol.* 2000; 50: 1351-1371.
- (61) Rodriguez-Tudela, J. L., Diaz-Guerra T. M., Mellado E., Cano V., Tapia C., Parkins A., Gomez-Lopez A., Rodero L. and Cuenca-Estrella M. Susceptibility patterns and molecular identification of *Trichosporon* species. *Antimicrob. Agents Chemother.* 2005; 49: 4026-4034.
- (62) Sugita, T., R. Ikeda, and A. Nishikawa. Analysis of *Trichosporon* isolates obtained from the houses of patients with summer-type hypersensitivity pneumonitis. *J. Clin. Microbiol.* 2004; 42: 5467-5471.
- (63) Sugita, T., M. Nakajima, R. Ikeda, T. Matsushima, and T. Shinoda. Sequence analysis of the ribosomal DNA intergenic spacer 1 regions of *Trichosporon* species. *J. Clin. Microbiol.* 2002; 40: 1826-1830.
- (64) Sugita T., Ichikawa T., Matsukura M., Sueda M., Takashima M., Ikeda R., Nishikawa A. and Shinoda T. Genetic diversity and biochemical characteristics of *Trichosporon asahii* isolated from clinical specimens, houses of patients with summer-type-hypersensitivity pneumonitis, and environmental materials.



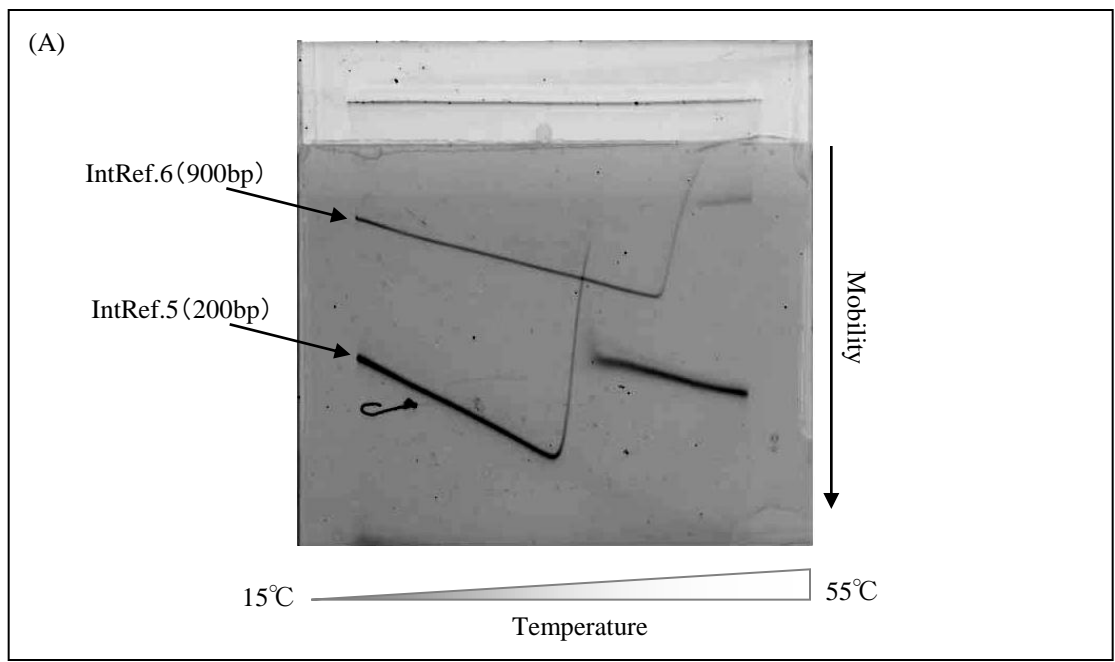
- J. Clin. Microbiol.* 2001; 39: 2405-2411.
- (65) Wolf D. G., Falk R., Hacham M., Theelen B., Boekhout T., Scorzetti G., Shapiro M., Block C., Salkin I. F. and Polacheck I. Multidrug-resistant *Trichosporon asahii* infection of nongranulocytopenic patients in three intensive care units. *J. Clin. Microbiol.* 2001; 39: 4420–4425.
- (66) Juan C. R., L. Koreen L., Park S. and Perlin D. S. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. *J. Clin. Microbiol.* 2004; 42: 2480-2488.
- (67) 波多野和徳. カルチャーコレクションー理想と現実ー *Microbiol. Cult. Coll.* 2001; 17: 97-100
- (68) 林崎良英, 八尾徹, 五條堀孝. 次世代シーケンサーは生命科学に新たな革命をもたらす. *科学.* 2009; 79: 231-244.
- (69) Hamano K., Ueno-Tsuji S., Tanaka R., Suzuki M., Nishimura K., Nishigaki K. Genome profiling (GP) as an effective tool for monitoring culture collections: a case study with *Trichosporon*. *J. Microbiol. Methods.* 2012; 89: 119-128.



**Fig. 1 Main steps of genome profiling (GP).** GP consists of DNA extraction, random PCR,  $\mu$ TGGE and computer-aided data processing (normalization). While the specific PCR where the primers hybridize to the template DNA completely complementarily and render the target region amplified, random PCR enable us to collect double stranded (ds) DNA fragments from genomic DNA by mismatch-allowing hybridization of a primer. It has a meaning of 'random sampling in Statistics' which extracts partial information from a massive one. The random PCR itself<sup>(21)</sup> is identical to the concurrently devised RAPD (random amplified polymorphic DNA)<sup>(16)</sup>, and GP has an additional process to obtain the sequence information without sequencing: TGGE which can detect the sequence-dependent double-stranded DNA melting<sup>(36)</sup>. This figure shows the weakness of RAPD which is usually operated without controlling temperature: That is, the fluctuation of temperature changes the order of DNA migration pattern as indicated in the inset. In TGGE, single stranded (ss) DNAs behave quite differently (monotonously) from the corresponding dsDNA, providing the chance to discriminate them. 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*<sup>(25)</sup> This figure was taken from Journal of Microbiological Methods<sup>(69)</sup>.



**Fig. 2 Three processes of genome profiling (GP).** Random PCR. A probe binds to various sites of genome DNA in a mismatch- and/or bulge-containing structure when the PCR is operated under lower temperature conditions, leading to the generation of a set of DNA fragments. The figure represents the whole of chromosomes conceptually but not the actual shape. (b) TGGE (temperature gradient gel electrophoresis). In TGGE, random PCR fragments are forced to migrate in a sequence-specific pattern which has featuring points (open circles) together with two internal references (shown in closed circles). (c) How to obtain the *PaSS* (pattern similarity score). The genome profiles of two species (top, left and right) are computer-processed after spotting featuring points (shown in open and closed circles). The featuring points derived from two internal references (shown in closed circles) are used to normalize the other samples' featuring points so as to obtain *spiddos* (species identification dots). The definition of the *PaSS* score is shown at the bottom (see also Ref. 25). This figure was taken from Journal of Microbiological Methods <sup>(69)</sup>.



(B)

**Intref. 5 (200bp)**

3501 AGTGGTCCTG CAACTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG  
 3551 GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG  
 3601 CCATTGCTGC AGGCATCGTG GTGTACGCT CGTCGTTTGG TATGGCTTCA  
 3651 TTCAGCTCCG GTTCCAACG ATCAAGGCCA GTTACATGAT CCCCCATGTT

**Intref. 6 (900bp)**

401 GCCGGCATCA CCGGCGCCAC AGGTGCGGTT GCTGGCGCCT ATATCGCCGA  
 451 CATCACCGAT GGGGAAGATC GGGCTCGCCA CTTCGGGCTC ATGAGCGCTT  
 501 GTTTCGGCGT GGGTATGGTG GCAGGCCCG TGGCCGGGGG ACTGTTGGGC  
 551 GCCATCTCCT TGCATGCACC ATTCCTTGCG GCGGCGGTGC TCAACGGCCT  
 601 CAACCTACTA CTGGGCTGCT TCCTAATGCA GGAGTCGCAT AAGGGAGAGC  
 651 GTCGACCGAT GCCCTTGAGA GCCTTCAACC CAGTCAGCTC CTCCGGTGG  
 701 GCGCGGGGCA TGACTATCGT CGCCGCACTT ATGACTGTCT TCTTTATCAT  
 751 GCAACTCGTA GGACAGGTGC CGGCAGCGCT CTGGGTCATT TTCGGCGAGG  
 801 ACCGCTTTCG CTGGAGCGCG ACGATGATCG GCCTGTCGCT TGCGGTATTC  
 851 GGAATCTTGC ACGCCCTCGC TCAAGCCTTC GTCACTGGTC CCGCCACCAA  
 901 ACGTTTCGGC GAGAAGCAGG CCATTATCGC CGGCATGGCG GCCGACGCGC  
 951 TGGGCTACGT CTTGCTGGCG TTCGCGACGC GAGGCTGGAT GGCCTTCCCC  
 1001 ATTATGATTC TTCTCGCTTC CGGCGGCATC GGGATGCCCG CGTTGCAGGC  
 1051 CATGCTGTCC AGGCAGGTAG ATGACGACCA TCAGGGACAG CTTCAAGGAT  
 1101 CGCTCGCGGC TCTTACCAGC CTAACCTCGA TCACTGGACC GCTGATCGTC  
 1151 ACGGCGATTT ATGCCGCCTC GGCGAGCACA TGGAACGGGT TGGCATGGAT  
 1201 TGTAGGCGCC GCCCTATACC TTGTCTGCCT CCCC CGTTG CGTCGCGGTG  
 1251 CATGGAGCCG GGCCACCTCG ACCTGAATGG AAGCCGGCGG CACCTCGCTA

**Fig. 3-1 TGGE band pattern and sequences of double internal references.**

(A)  $\mu$ TGGE band pattern of double references (IntRef. 5 and IntRef. 6). The conditions of  $\mu$ TGGE are as follows, 6% (w/v) acrylamide gel (acrylamide: BIS = 19:1, 6.5 M urea, 1  $\times$  TBE buffer),  $\mu$ 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. (B) Sequences of double references. Double references are extracted and amplified from pBR322 DNA by PCR. Left numbers are site numbers of pBR322 DNA.

<b>Double Reference (IntRef. 5 and 6)</b>		First( $W_1$ )	Second( $W_2$ )	Third( $W_3$ )	Fourth( $W_4$ )
	profile				
	Spiddos				
<b>Single Reference (IntRef. 1)</b>		First( $S_1$ )	Second( $S_2$ )	Third( $S_3$ )	Fourth( $S_4$ )
	profile				
	Spiddos				

**Fig. 3-2 Comparison of double internal references and single internal reference used for  $\mu$ TGGE.** The double references (IntRef. 5, 6) and the single reference (IntRef. 1) systems were compared employing the same random PCR products to examine the effect as the internal reference for  $\mu$ TGGE. The blank circles and the filled circles on the genome profile indicate the featuring points of random PCR products and the internal references, respectively.

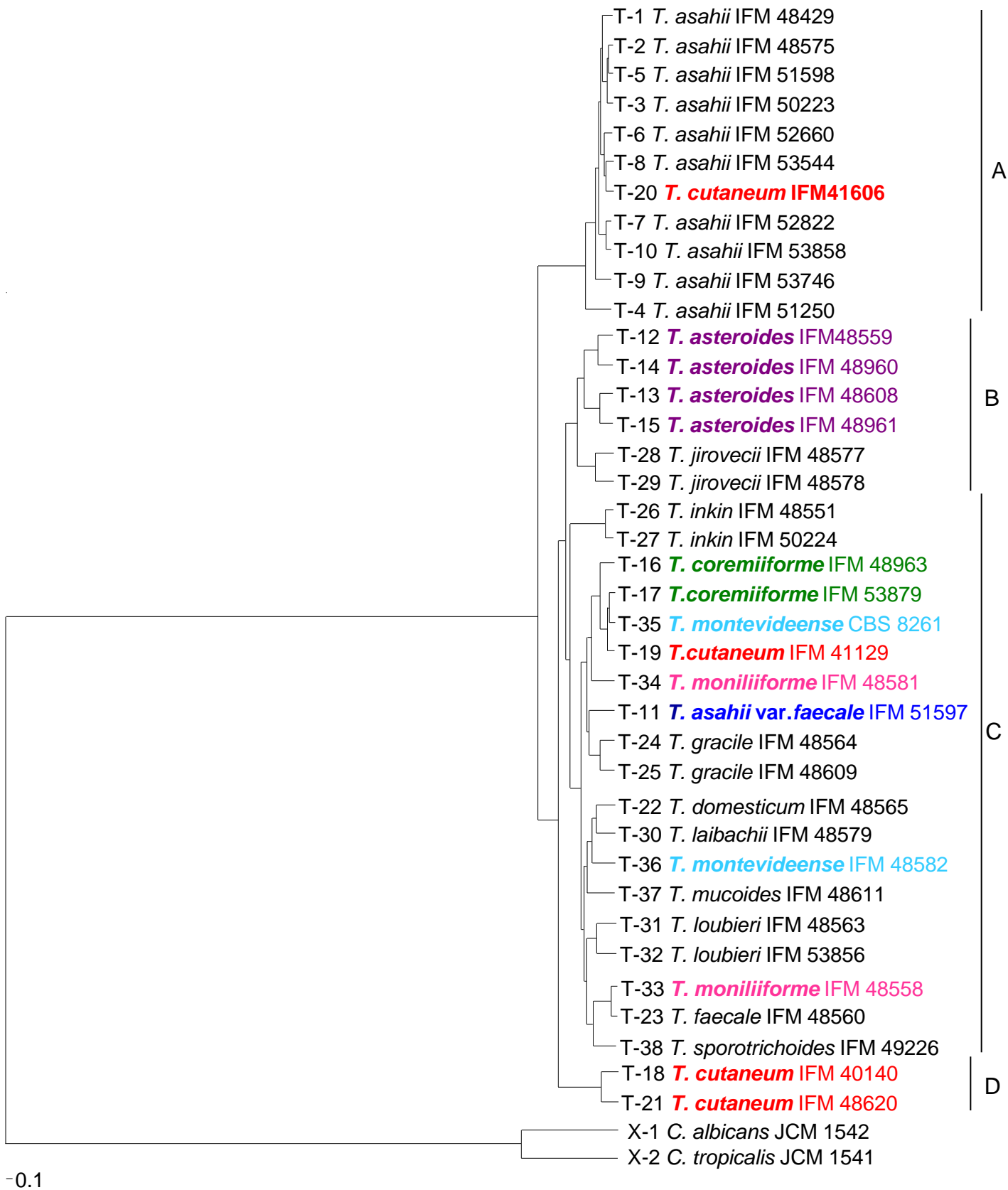


Fig. 4 Phylogenetic tree drawn based on the genome distance  $d_g (=1-PaSS)$ . The genome distance adopted here is the average value over the values obtained with the seven different probes. UPGMA algorithm was used to make this tree. Those species that are disputably assigned are shown in bold. T-serial numbers are the same used in Table 1. The distinct clusters are designated A ~ D. This figure was taken from Journal of Microbiological Methods <sup>(69)</sup>.

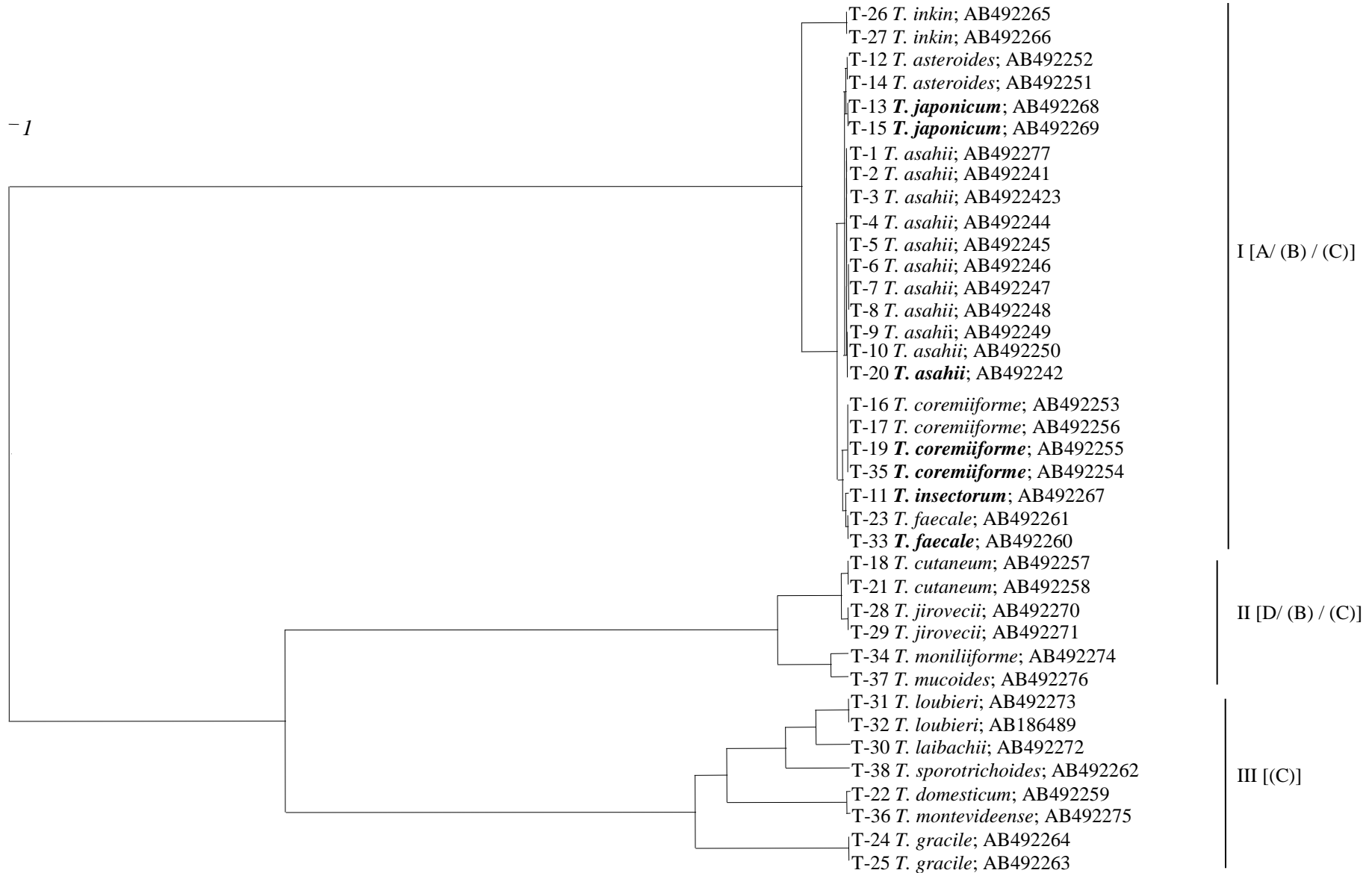


Fig. 5. Phylogenetic tree based on D1/D2 26S rDNA sequences using the UPGMA algorithm. Each sequence was assigned based on BLAST searches. Those species that led to the decision of misassignment and re-named are indicated in bold. The distinct Clusters I ~ III are indicated at the right side with Cluster names obtained from the GP-based tree (Fig. 4) where Clusters B and C are divided into subclusters (shown in parentheses). This figure was taken from Journal of Microbiological Methods <sup>(69)</sup>.





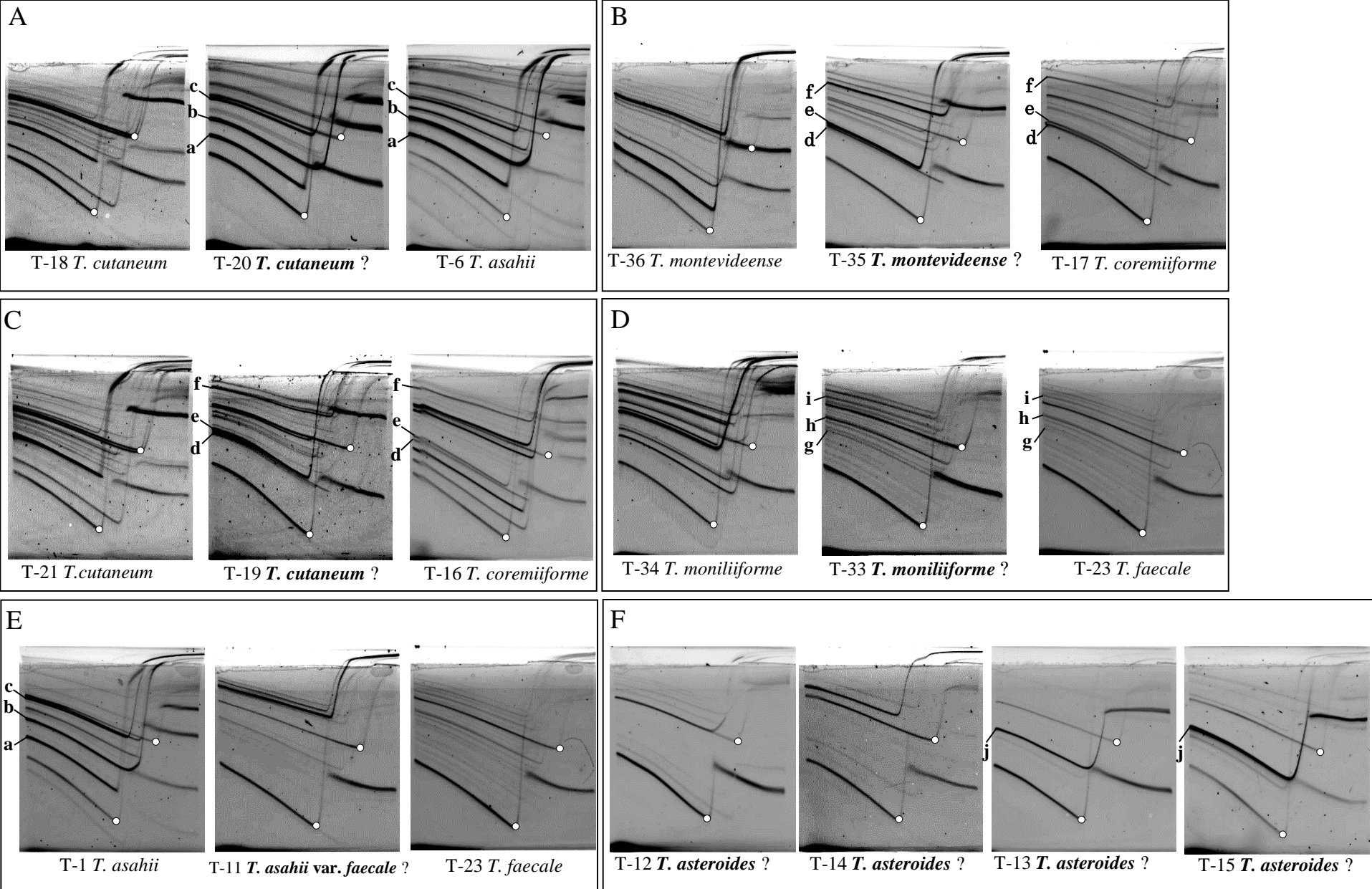


Fig. 7 Detailed genome profile analysis (*cggf* analysis)

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 *cggfs* (i.e., similar characteristic bands between two species possibly originated from a common gene or so<sup>(55)</sup>) are designated with letters a ~ j. This figure was taken from Journal of Microbiological Methods<sup>(69)</sup>.

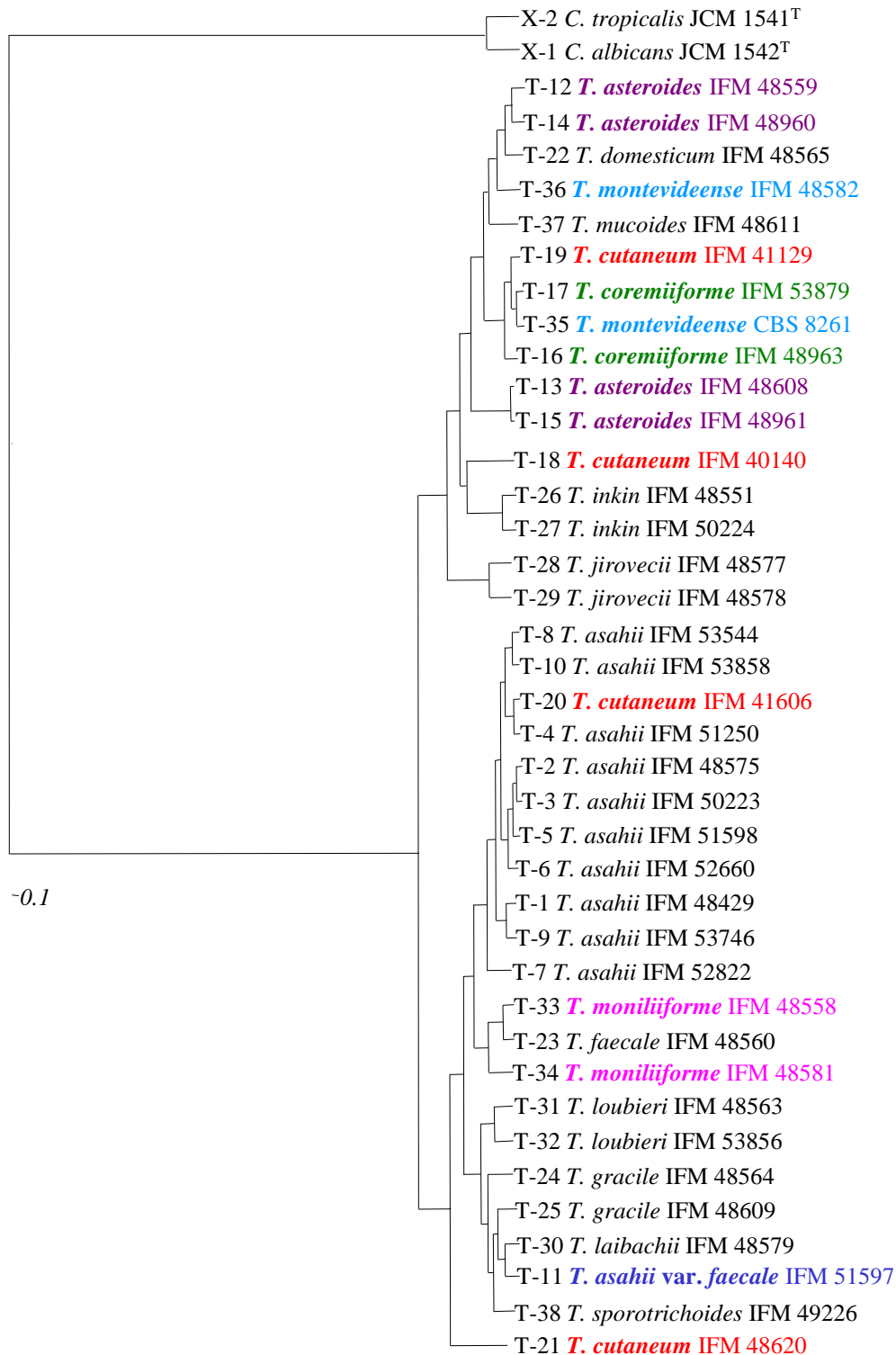


Fig. 8. Phylogenetic tree based on genome distance  $d_g$  ( $1-PaSS$ ) obtained by the single probe method (Hunt ; 5' TGCTGCTGCTGC 3') with the addition of outgroups (*Candida albicans* JCM 1542<sup>T</sup> and *Candida tropicalis* JCM 1541<sup>T</sup>) to the set of *Trichosporon* species. The UPGMA algorithm was used to make the tree. Those species in question are shown in bold (see text). This figure was taken from Journal of Microbiological Methods <sup>(69)</sup>.

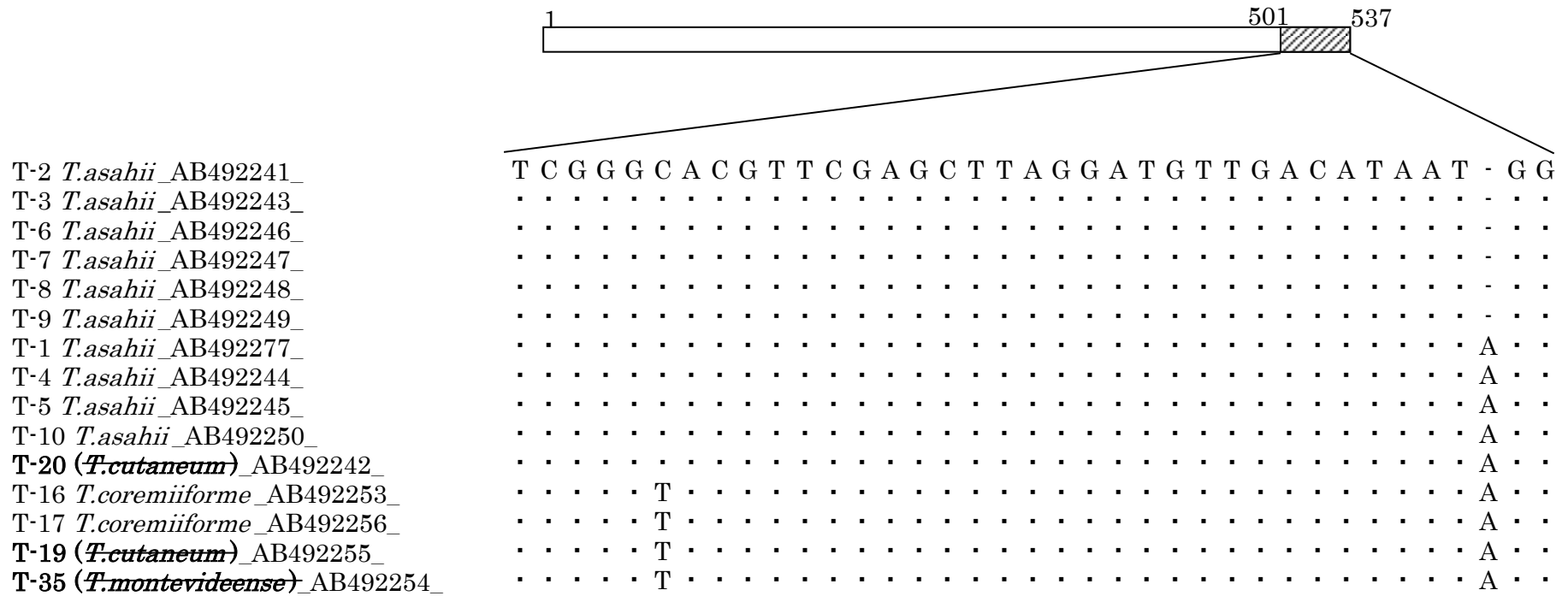


Fig. 9. Sequences used for discriminating species. *Trichosporon* D1/D2 26S rDNA sequences (~600 nts) were used for tree-making. In this figure, those partial sequences are shown, which could represent mutual differences (501-537). In this study, the strains of T-20, T-19 and T-35 were renamed to *T. asahii*, *T. coremiiforme* and *T. coremiiforme*, respectively.

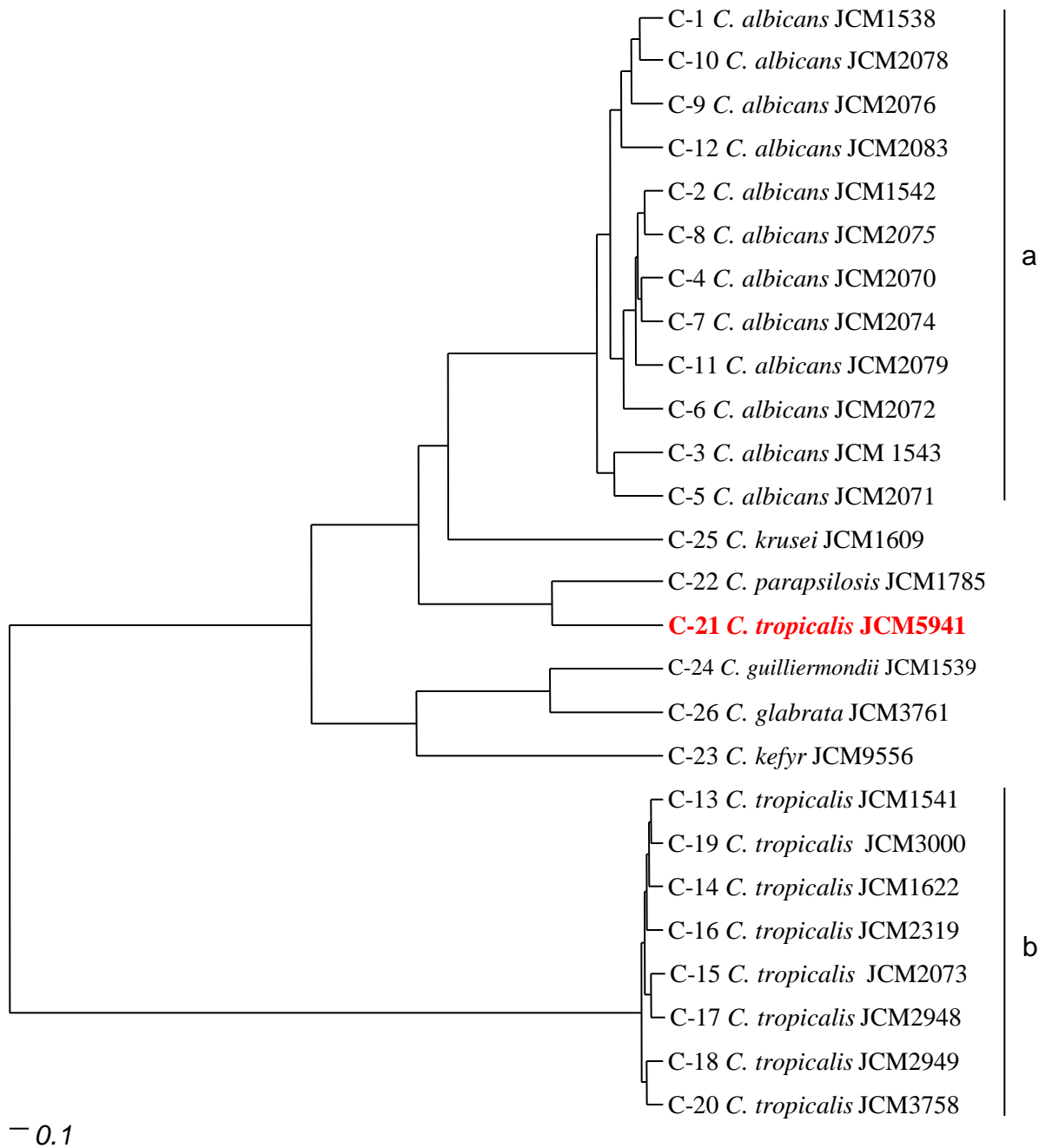


Fig. 10. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$ . The genome distance adopted here is the average value over the values obtained with the five different probes. UPGMA algorithm was used to make this tree. The strain C-21 that is disputably assigned is shown in bold. C-serial numbers are the same used in Table 2. The distinct clusters are designated a, b.

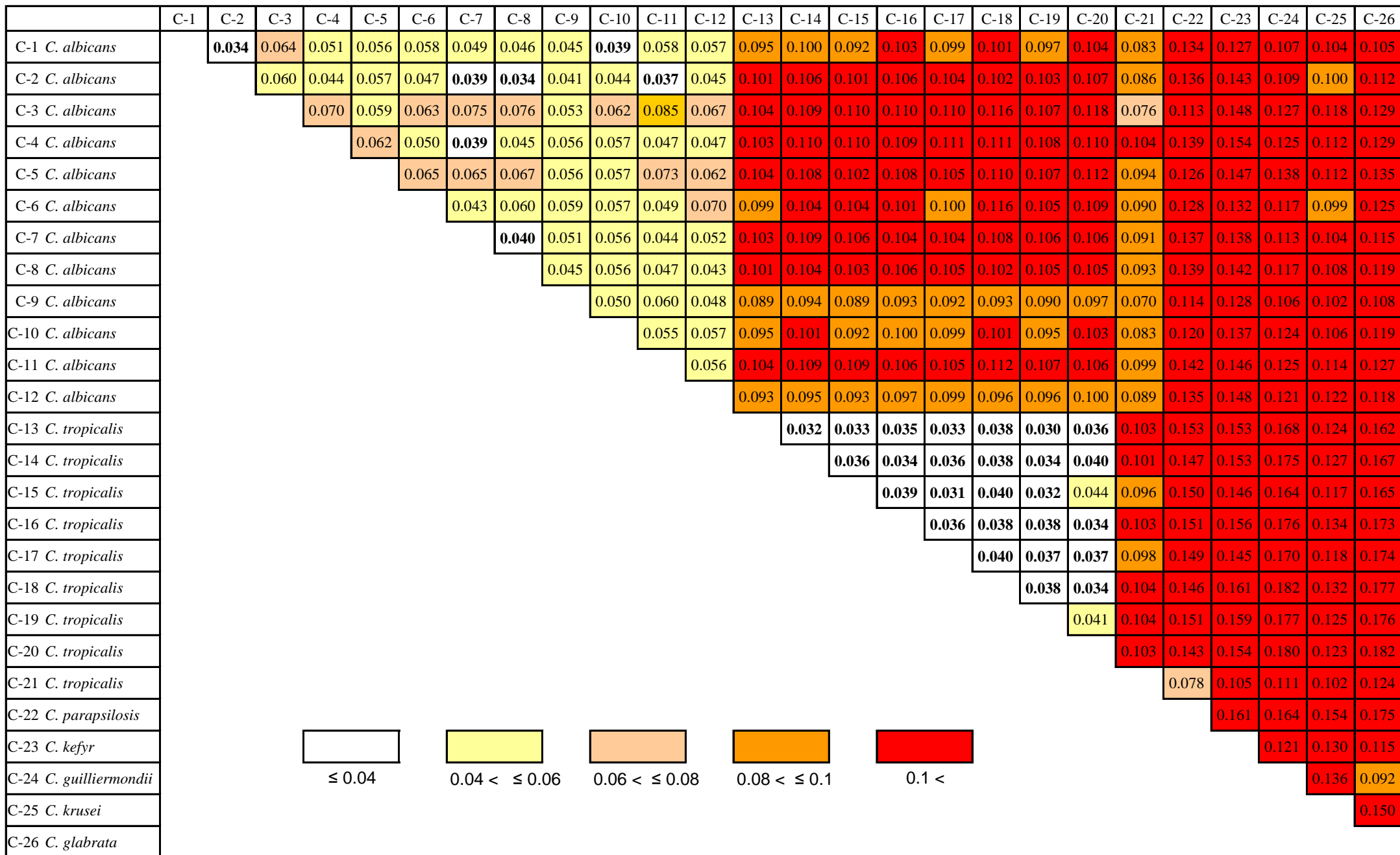


Fig. 11. **Triangle distance map** (tabular presentation of genome distance  $d_g$ ). This matrix is the average of  $d_g (=1-PaSS)$  obtained from the GP experiments using 5 different probes with respect to 26 strains of *Candida*. In this map, genome distances for closely related species (strains) appear near the diagonal in a pale colored box.

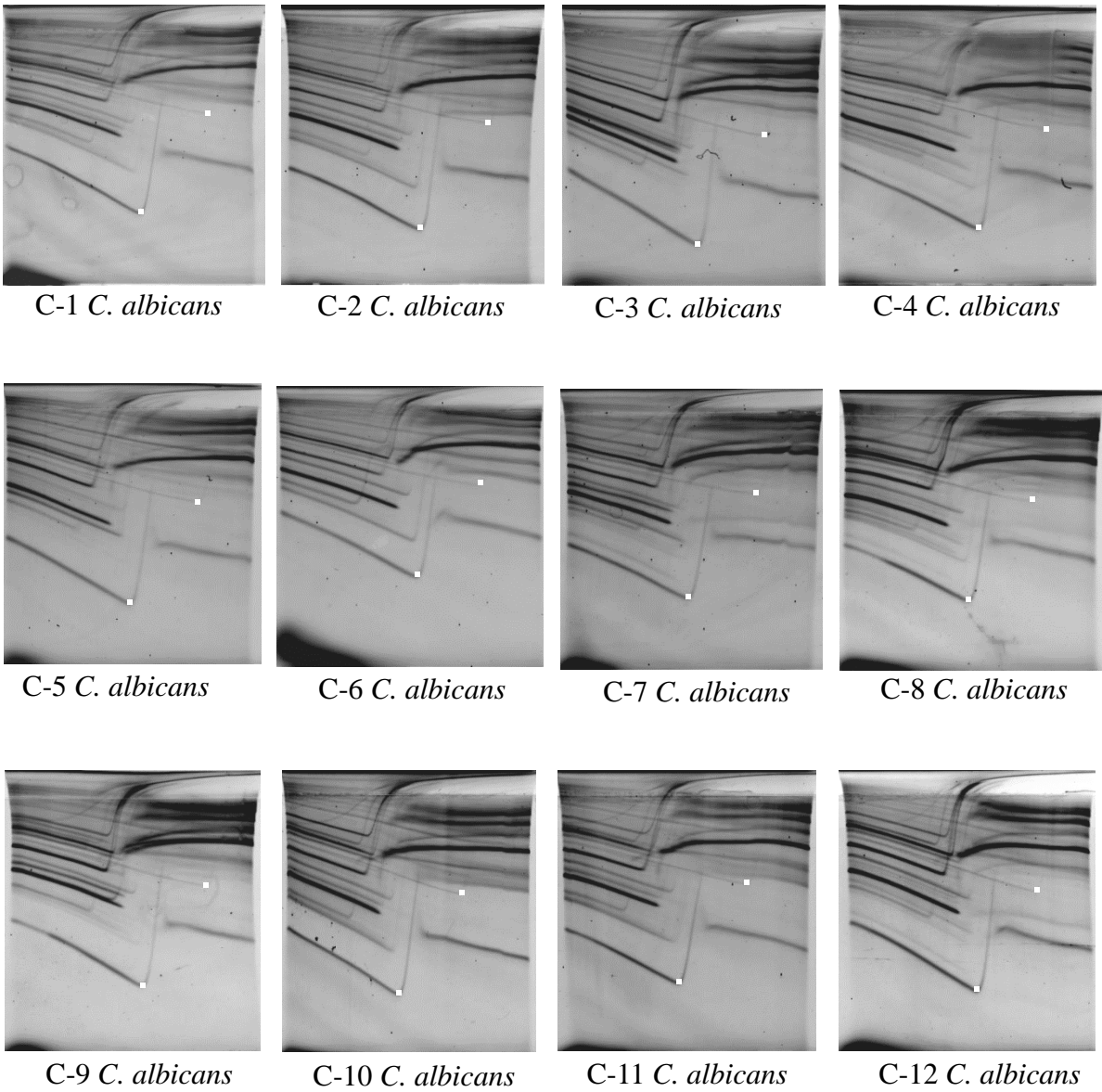
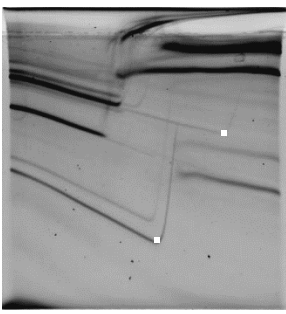
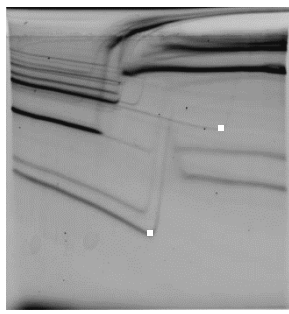


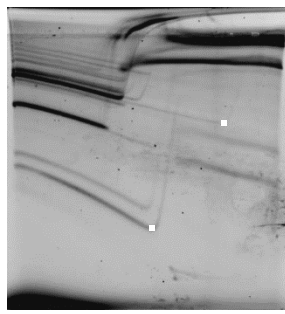
Fig. 12-1 Patterns of *Candida* genus GP using Hunt probe. Hunt probe sequence is 5'dTGCTGCTGCTGC. White dots are internal references (IntRef. 5, 6) in each profile.



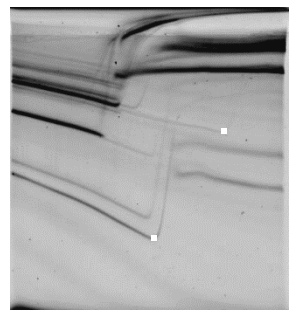
C-13 *C. tropicalis*



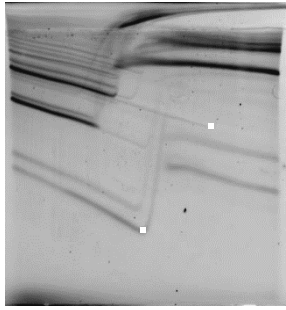
C-14 *C. tropicalis*



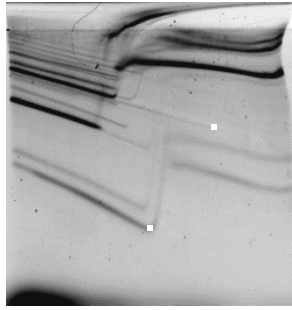
C-15 *C. tropicalis*



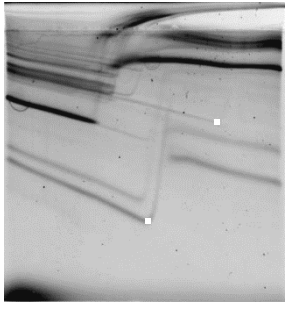
C-16 *C. tropicalis*



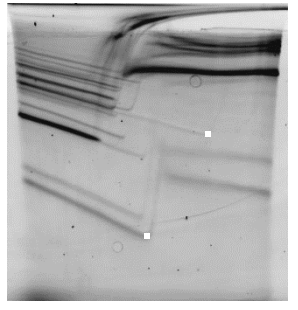
C-17 *C. tropicalis*



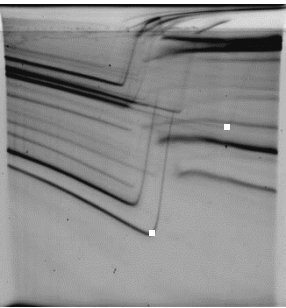
C-18 *C. tropicalis*



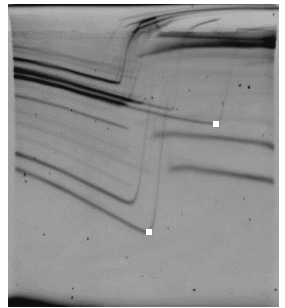
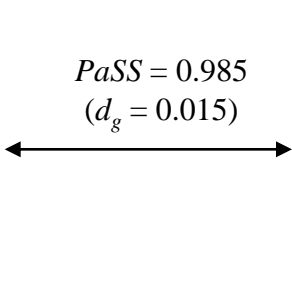
C-19 *C. tropicalis*



C-20 *C. tropicalis*

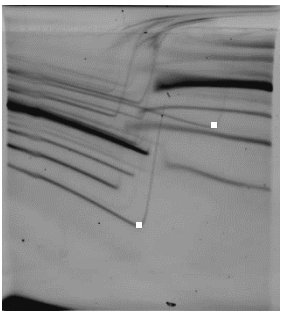
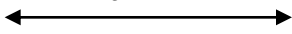


C-21 *C. tropicalis*

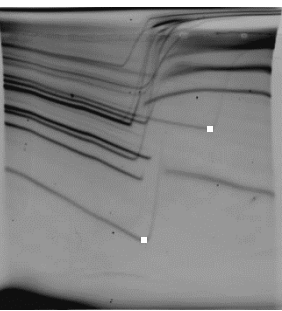


C-22 *C. parapsilosis*

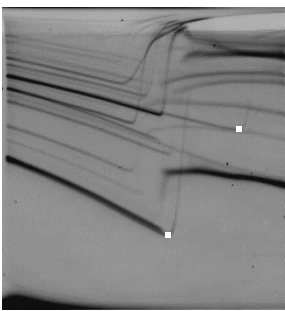
$PaSS = 0.985$   
 $(d_g = 0.015)$



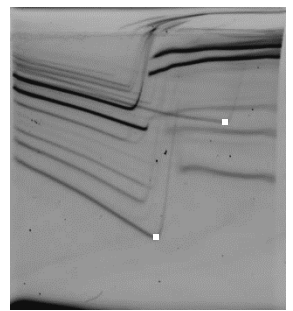
C-23 *C. kefyr*



C-24 *C. guilliermondii*



C-25 *C. krusei*



C-26 *C. glabrata*

Fig. 12-2 Continued. C-21 strain registered *C. tropicalis* is different from other *C. tropicalis*, but is similar to C-22 *C. parapsilosis*, and has high  $PaSS$  ( $=0.985$ ).

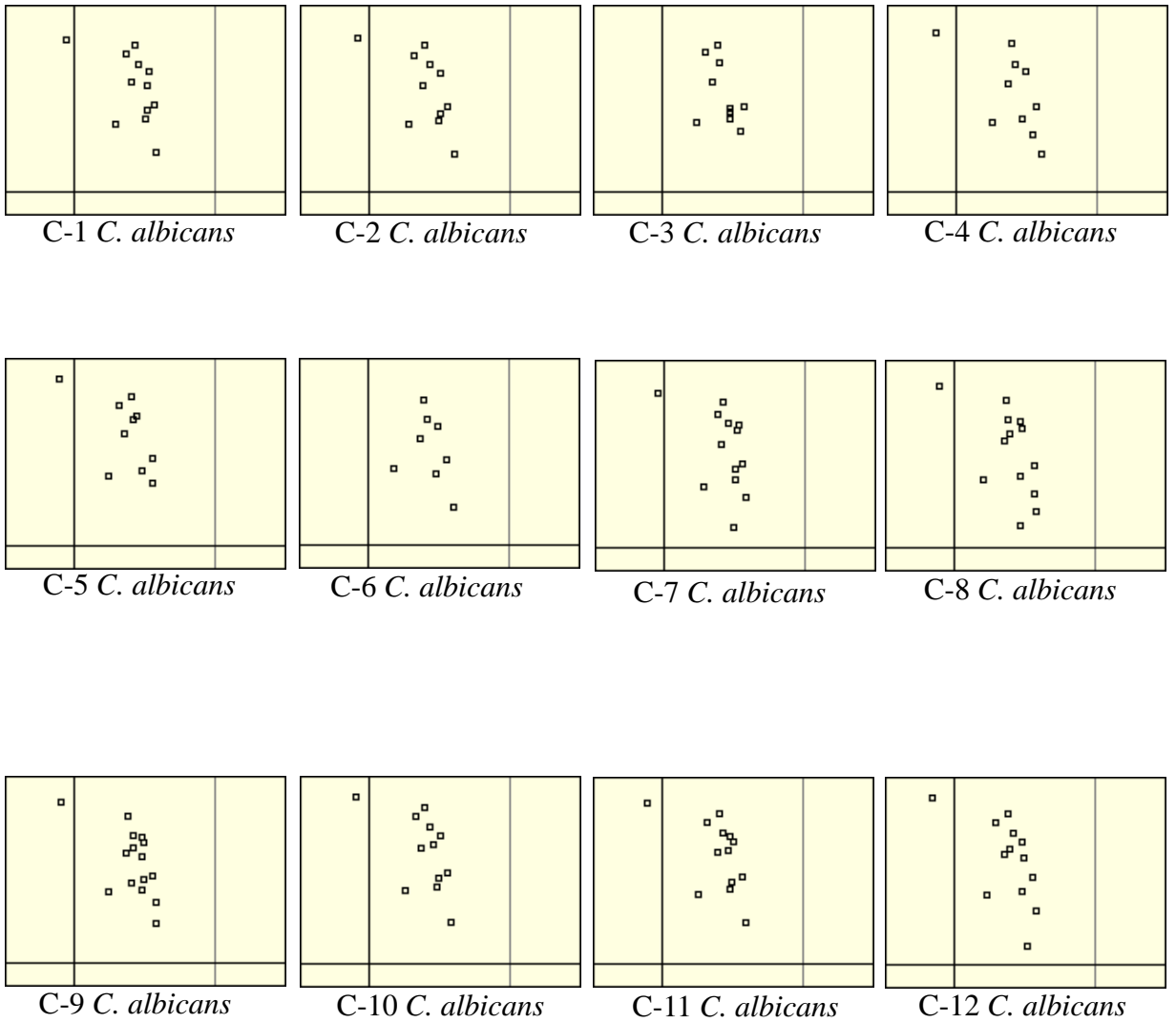
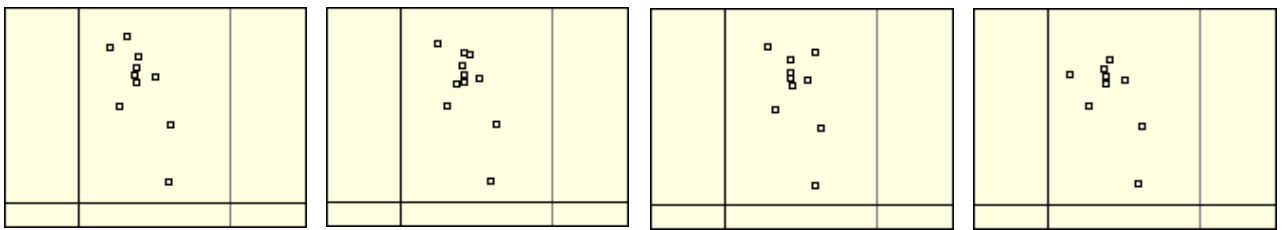


Fig. 13-1 *Spiddos* of *Candida* genus GP using Hunt probe. Hunt probe sequence is 5'dTGCTGCTGCTGC.



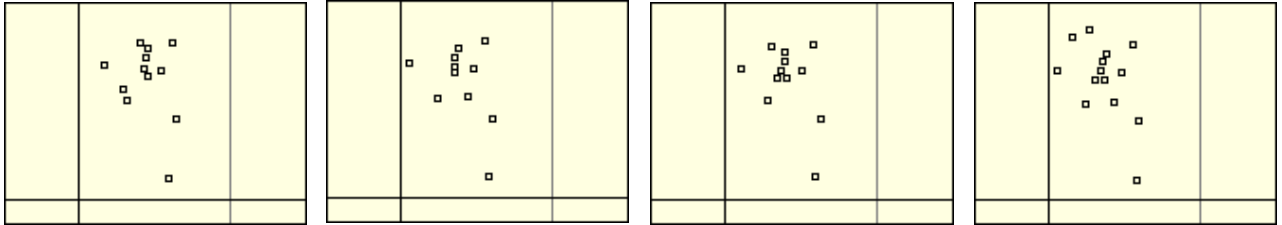


C-13 *C. tropicalis*

C-14 *C. tropicalis*

C-15 *C. tropicalis*

C-16 *C. tropicalis*

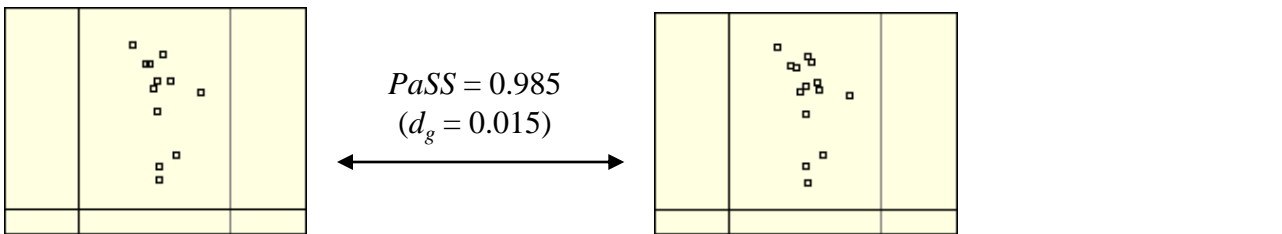


C-17 *C. tropicalis*

C-18 *C. tropicalis*

C-19 *C. tropicalis*

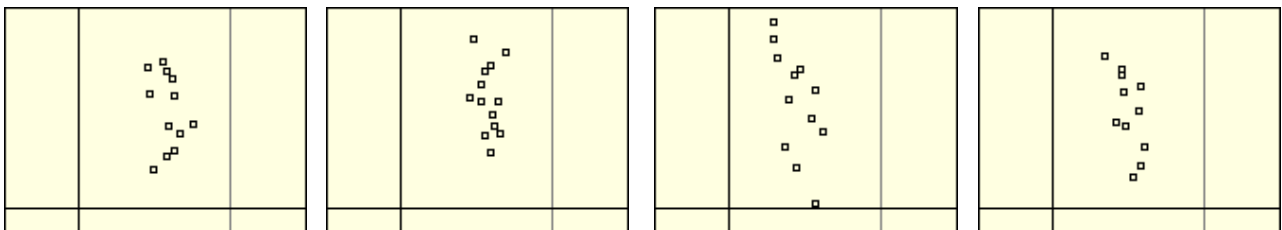
C-20 *C. tropicalis*



C-21 *C. tropicalis*

$PaSS = 0.985$   
( $d_g = 0.015$ )

C-22 *C. parapsilosis*



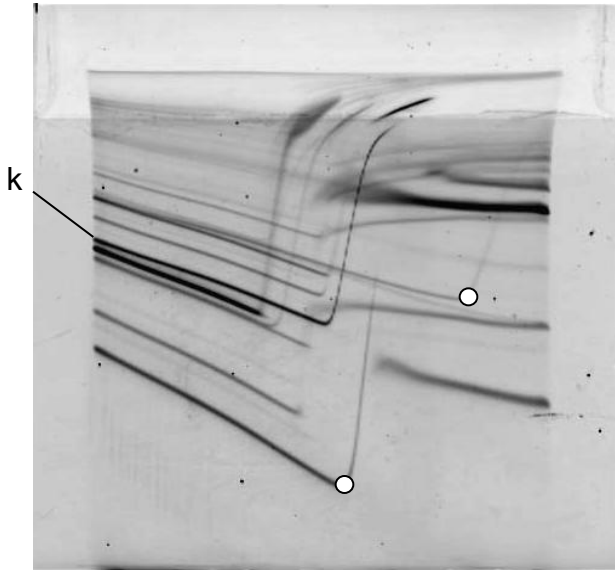
C-23 *C. kefyr*

C-24 *C. guilliermondii*

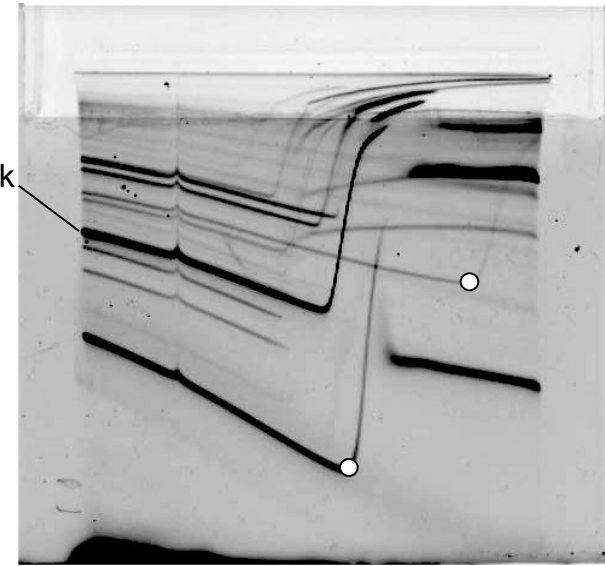
C-25 *C. krusei*

C-26 *C. glabrata*

Fig. 13-2 Continued. C-21 strain registered *C. tropicalis* is different from other *C. tropicalis*, but is similar to C-22 *C. parapsilosis*, and has high  $PaSS$  ( $= 0.985$ ).

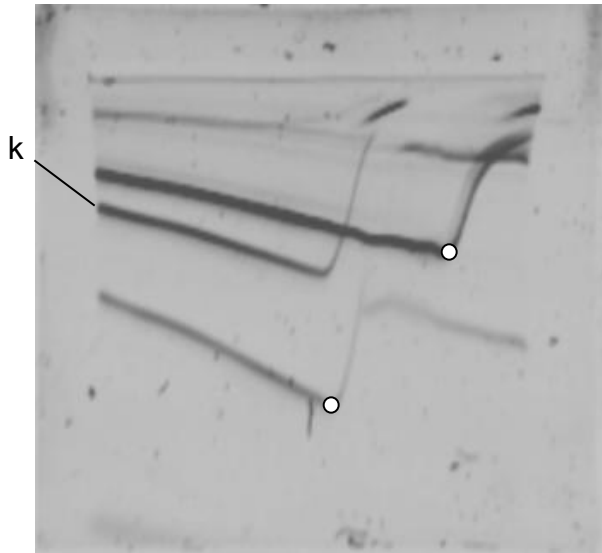


C-1 *C. albicans* JCM 1538

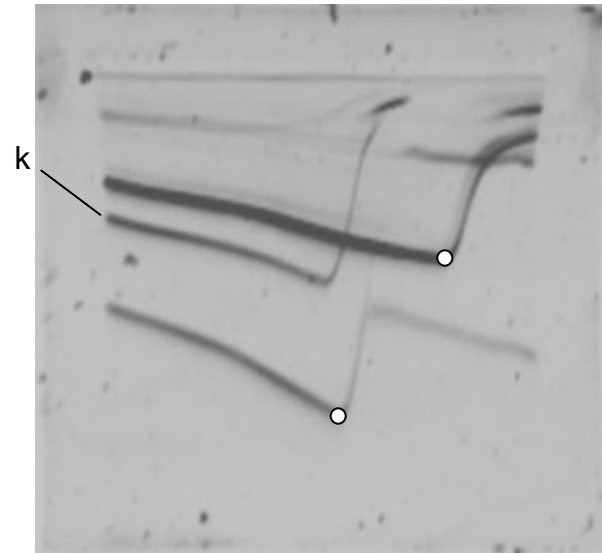


C-13 *C. tropicalis* JCM 1541

Fig. 14 The *ccgf* between *C. albicans* and *C. tropicalis*.  
The band k in this figure seems to be a *ccgf*.



C-1 *C. albicans* JCM 1538

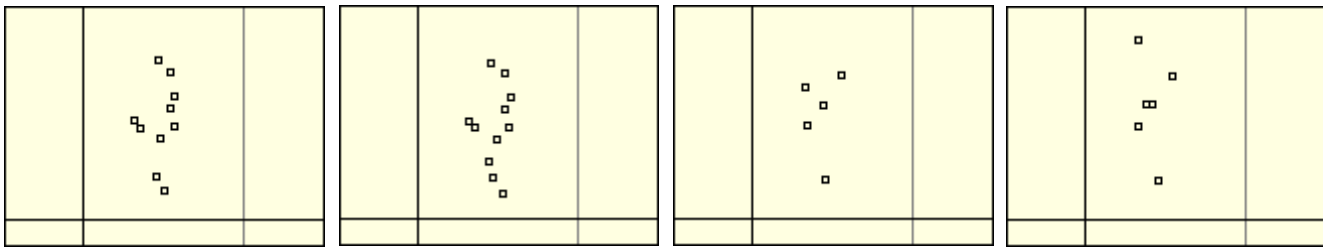


C-13 *C. tropicalis* JCM 1541

Fig. 15 The re-PCR product after extracting band k from TGGE gel was confirmed.

C-13	<i>C. tropicalis</i>	AGAACGCGCCTGAATAGGGTGAAGCCAGAGGAAACTCTGGTG <b>A</b> AGGCTCGTAGCGGTTCT	60
C-1	<i>C. albicans</i>	AGAACGCGCCTGAATAGGGTGAAGCCAGAGGAAACTCTGGTG <b>G</b> AGGCTCGTAGCGGTTCT	60
C-13	<i>C. tropicalis</i>	GACGTGCAAATCGATCGTCTGAATTTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAG	120
C-1	<i>C. albicans</i>	GACGTGCAAATCGATCGTCTGAATTTGGGTATAGGGGCGAAAGACTAATCGAACCATCTAG	120
C-13	<i>C. tropicalis</i>	TAGCTGGTTCCTG <b>T</b> CGAAGTTTCCCTCAGGATAGCAGAAGCTCGTATCAGTTTTATGAGG	180
C-1	<i>C. albicans</i>	TAGCTGGTTCCTG <b>C</b> CGAAGTTTCCCTCAGGATAGCAGAAGCTCGTATCAGTTTTATGAGG	180
C-13	<i>C. tropicalis</i>	TAAAGCGAATGATTAGAAGT <b>A</b> TTGGGGTTGAAATGACCTTAACTTATTCTCAAACCTTTAA	240
C-1	<i>C. albicans</i>	TAAAGCGAATGATTAGAAGT <b>C</b> TTGGGGTTGAAATGACCTTAACTTATTCTCAAACCTTTAA	240
C-13	<i>C. tropicalis</i>	ATATGTAAGAAGTCCTTGTTGCTTAATTGAACGTGGACAATTGAATGAAGAGCTTTTTAGT	300
C-1	<i>C. albicans</i>	ATATGTAAGAAGTCCTTGTTGCTTAATTGAACGTGGACAATTGAATGAAGAGCTTTTTAGT	300
C-13	<i>C. tropicalis</i>	GGGCCATTTTTGGTAAGCAGAACTGGCGATGCGGGATGAACCGAACGTGAAGTTAAAGTG	360
C-1	<i>C. albicans</i>	GGGCCATTTTTGGTAAGCAGAACTGGCGATGCGGGATGAACCGAACGTGAAGTTAAAGTG	360
C-13	<i>C. tropicalis</i>	<b>CGG</b> - AATGCACGCTCATCAGACACCACAAAAGGTGTTAGTTCATCTAGACAGCCGGACGG	420
C-1	<i>C. albicans</i>	<b>CCGG</b> AATGCACGCTCATCAGACACCACAAAAGGTGTTAGTTCATCTAGACAGCCGGACGG	420
C-13	<i>C. tropicalis</i>	TGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTCACCGGCCGAATGAACTA	480
C-1	<i>C. albicans</i>	TGGCCATGGAAGTCGGAATCCGCTAAGGAGTGTGTAACAACCTCACCGGCCGAATGAACTA	480
C-13	<i>C. tropicalis</i>	GCCCTGAAAATGGATGGCGCTCAAGCGTGCTACTTATACTTCACCGTGATTGCT <b>AATTTA</b>	540
C-1	<i>C. albicans</i>	GCCCTGAAAATGGATGGCGCTCAAGCGTGCTACTTATACTTCACCGTGATTGCT <b>GTTTT -</b>	539
C-13	<i>C. tropicalis</i>	<b>T</b> GATGCTTTCACGAGTAGGCAGGCGCGTTCT	571
C-1	<i>C. albicans</i>	<b>-</b> GAC <b>G</b> GCTTTCACGAGTAGGCAGGCGCGTTCT	569

Fig. 16. The sequence of *ccgf* between C-1 *C. albicans* JCM 1538 and C-13 *C. tropicalis* JCM 1541. The similarity of two sequence is 98.2%.

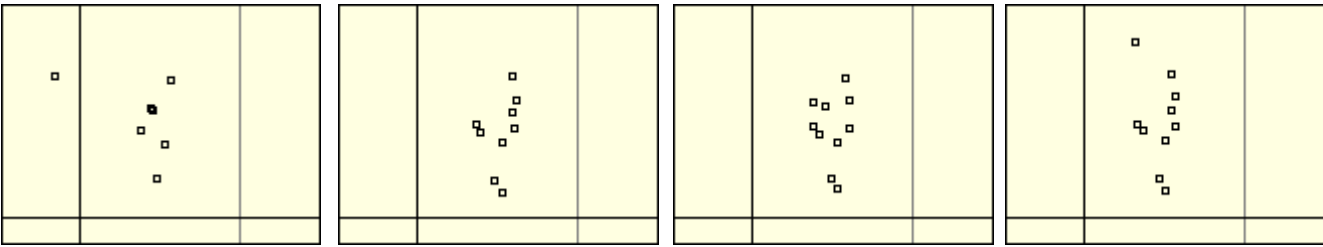


C-1 *C.albicans*

C-2 *C.albicans*

C-3 *C.albicans*

C-4 *C.albicans*

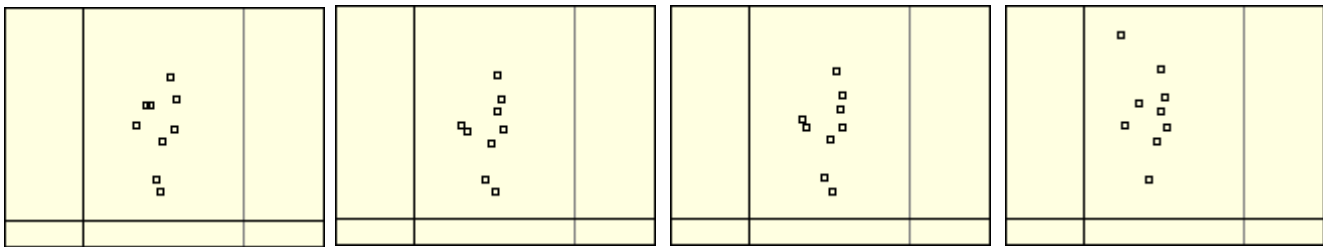


C-5 *C.albicans*

C-6 *C.albicans*

C-7 *C.albicans*

C-8 *C.albicans*



C-9 *C.albicans*

C-10 *C.albicans*

C-11 *C.albicans*

C-12 *C.albicans*

Fig.17-1 *Spiddos* of *Candida* genus GP using pfM12 probe.  
 PfM12 probe sequence is 5'dAGAACGCGCCTG.

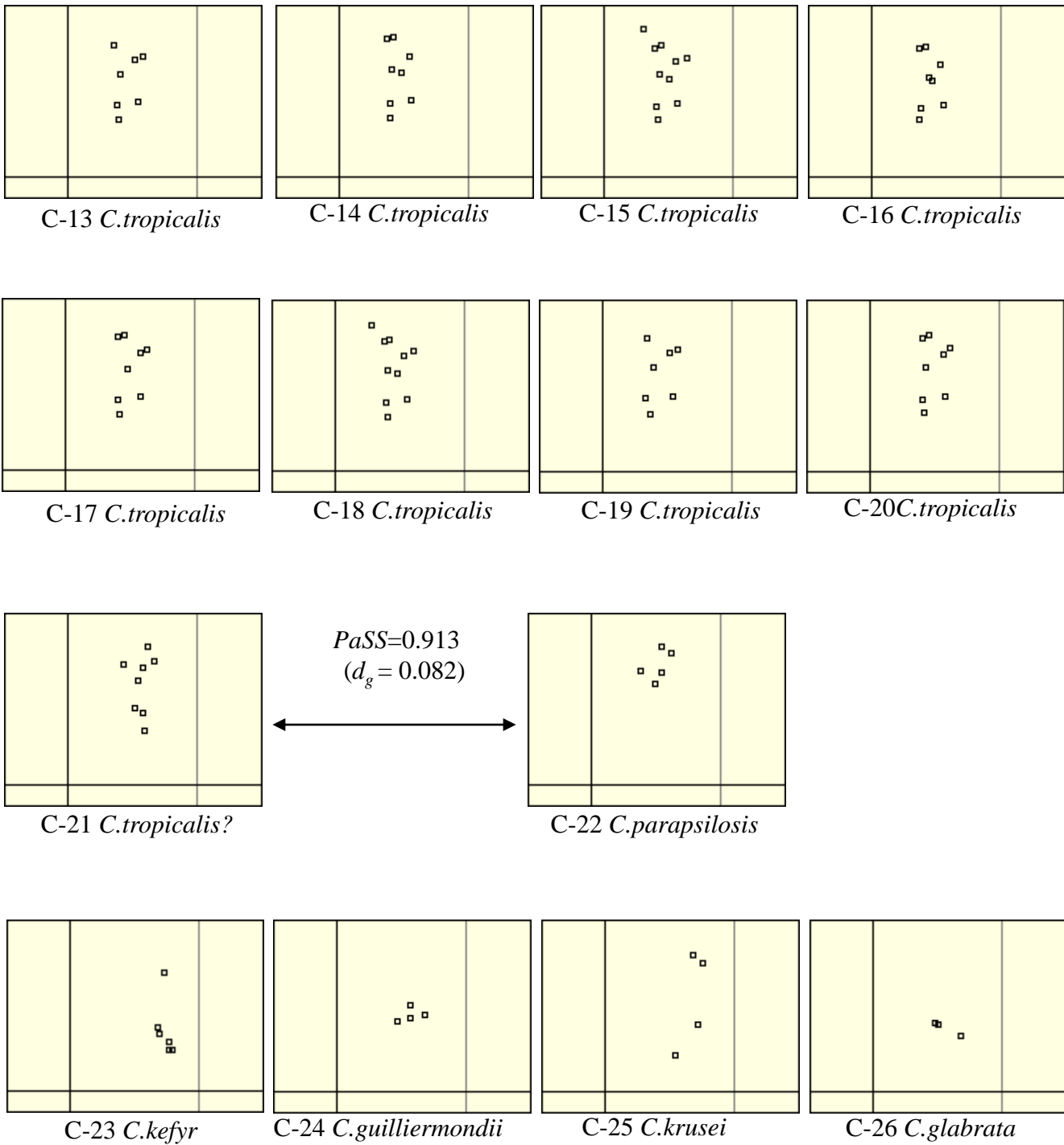


Fig.17-2 Continued. *Spiddos* of *Candida* genus GP using pFm12 probe. PFM12 probe sequence is 5' dAGAACGCGCCTG.

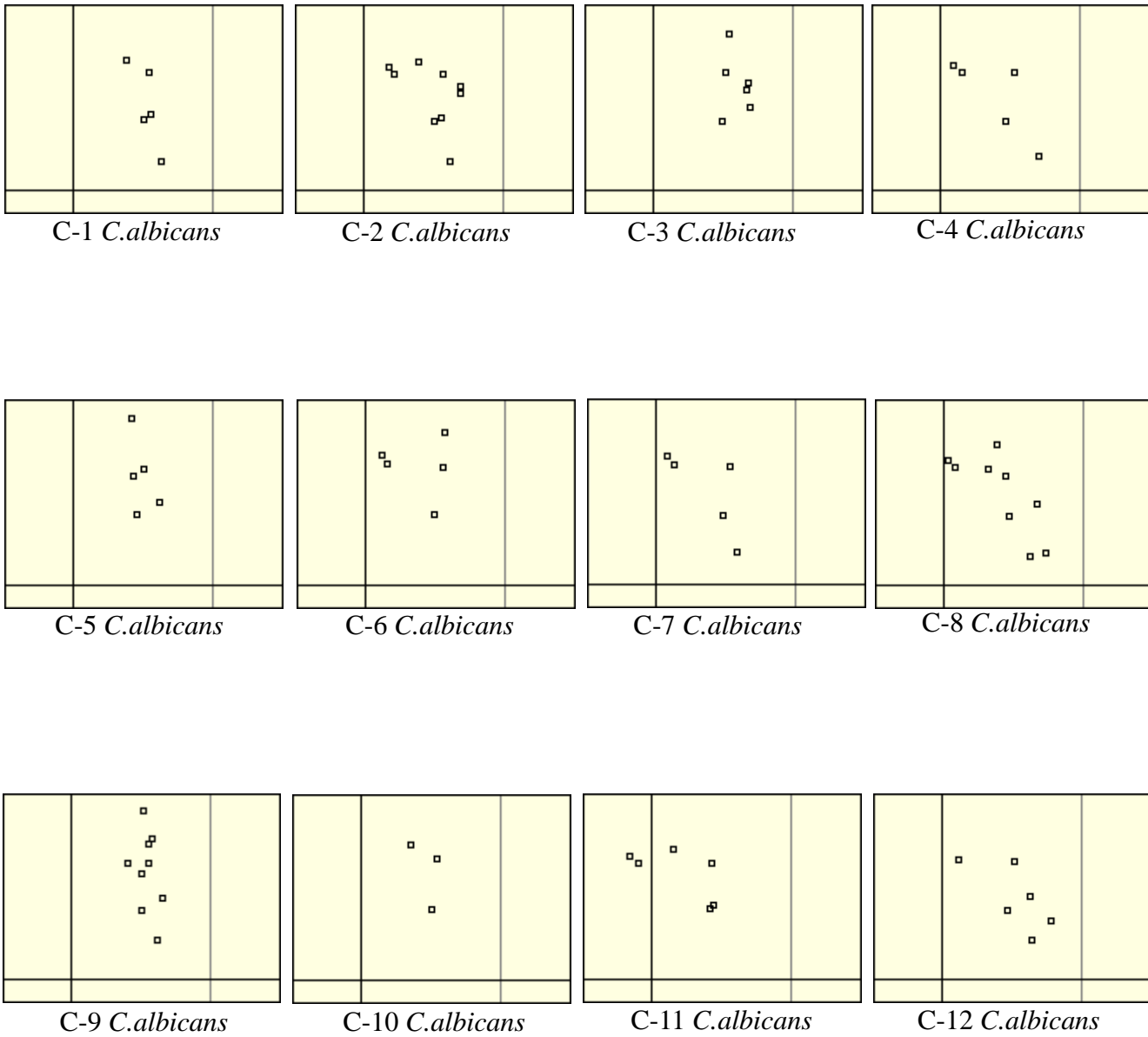


Fig.18-1 Spidpos of *Candida* genus GP using pfM19 probe.  
 PfM19 probe sequence is 5' dCAGGGCGCGTAC.

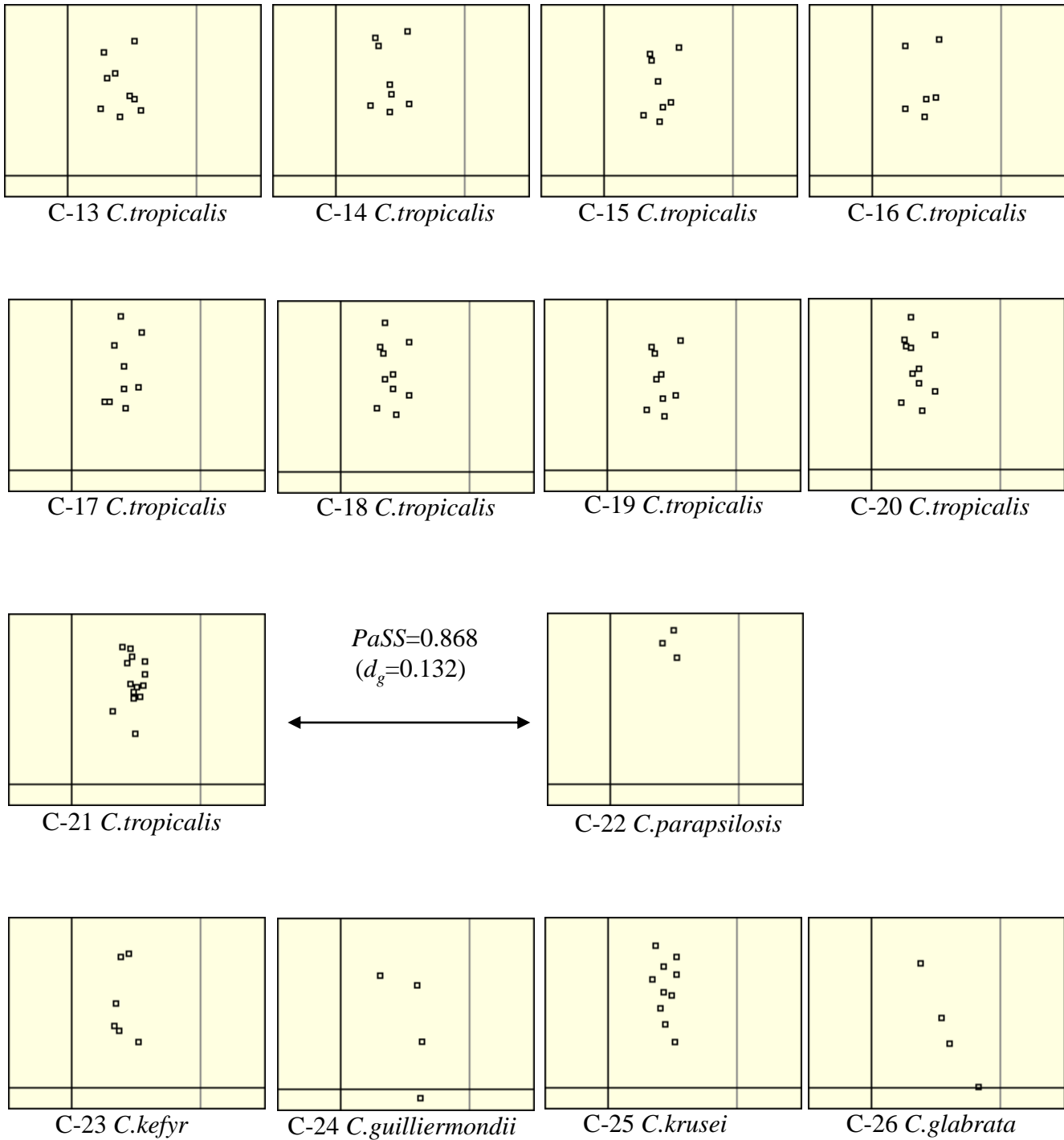


Fig.18-2 Continued. *Spiddos* of *Candida* genus GP using pfM19 probe.  
PfM19 probe sequence is 5'dCAGGGCGCGTAC



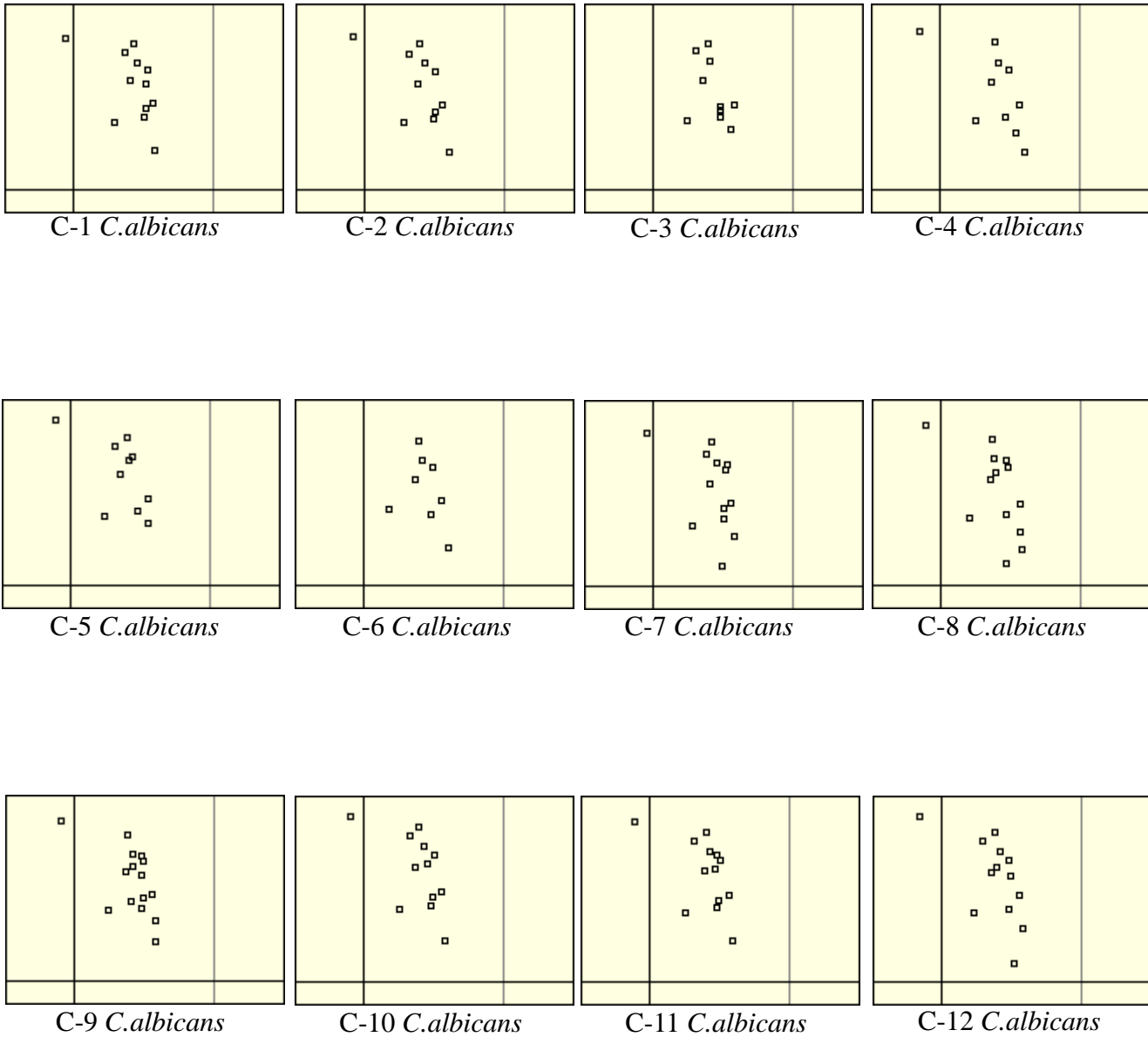


Fig.19-1 Spiddos of *Candida* genus GP using Hunt probe.  
 Hunt probe sequence is 5'd TGCTGCTGCTGC.

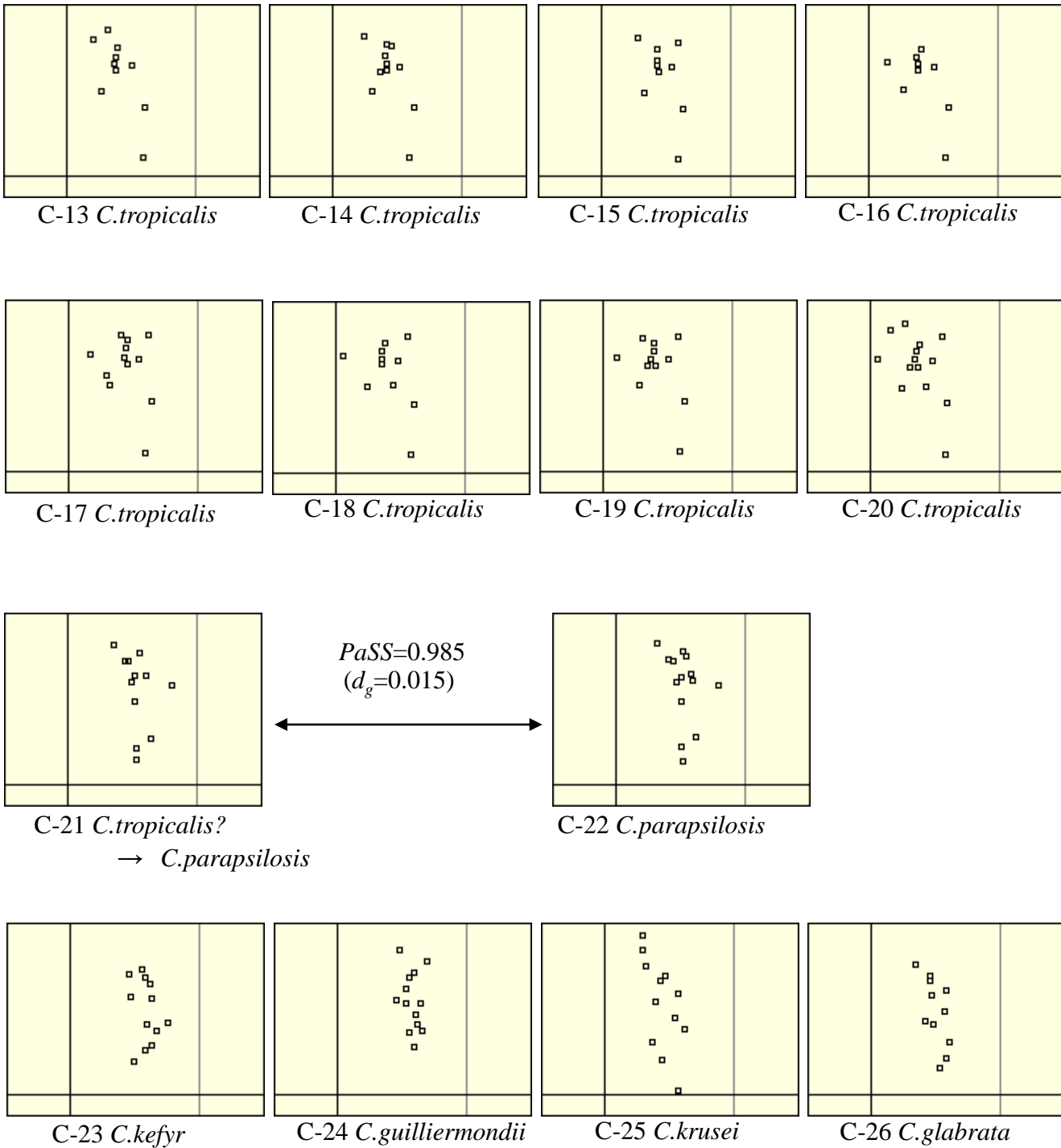
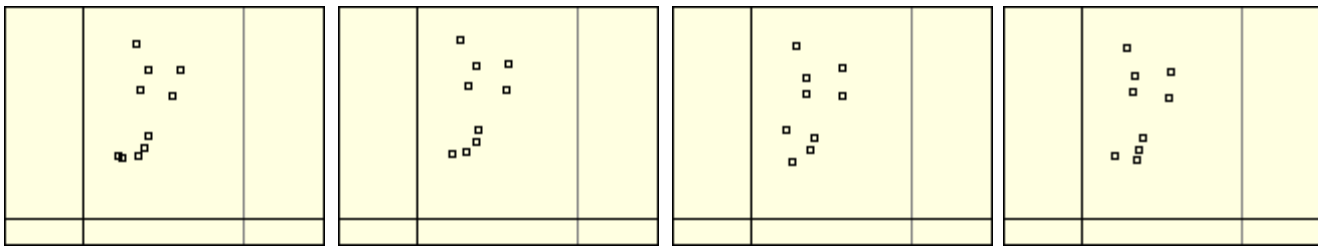


Fig.19-2 Continued. Spidpos of *Candida* genus GP using Hunt probe. Hunt probe sequence is 5'd TGCTGCTGCTGC. C-21 strain registered *C. tropicalis* is different from other *C. tropicalis*, but is similar to C-22 *C. parapsilosis*, and has very small  $d_g(=0.015)$ .

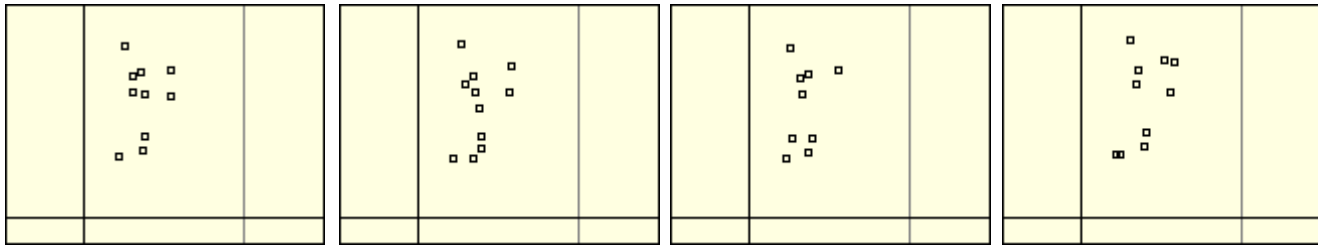


C-1 *C.albicans*

C-2 *C.albicans*

C-3 *C.albicans*

C-4 *C.albicans*

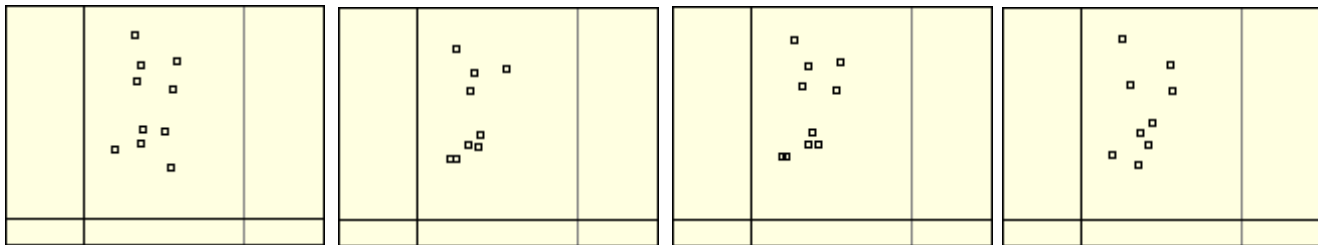


C-5 *C.albicans*

C-6 *C.albicans*

C-7 *C.albicans*

C-8 *C.albicans*



C-9 *C.albicans*

C-10 *C.albicans*

C-11 *C.albicans*

C-12 *C.albicans*

Fig. 20-1. *Spiddos* of *Candida* genus GP using TeeGee probe.  
TeeGee probe sequence is 5'dTTTGGGTTTGGG.

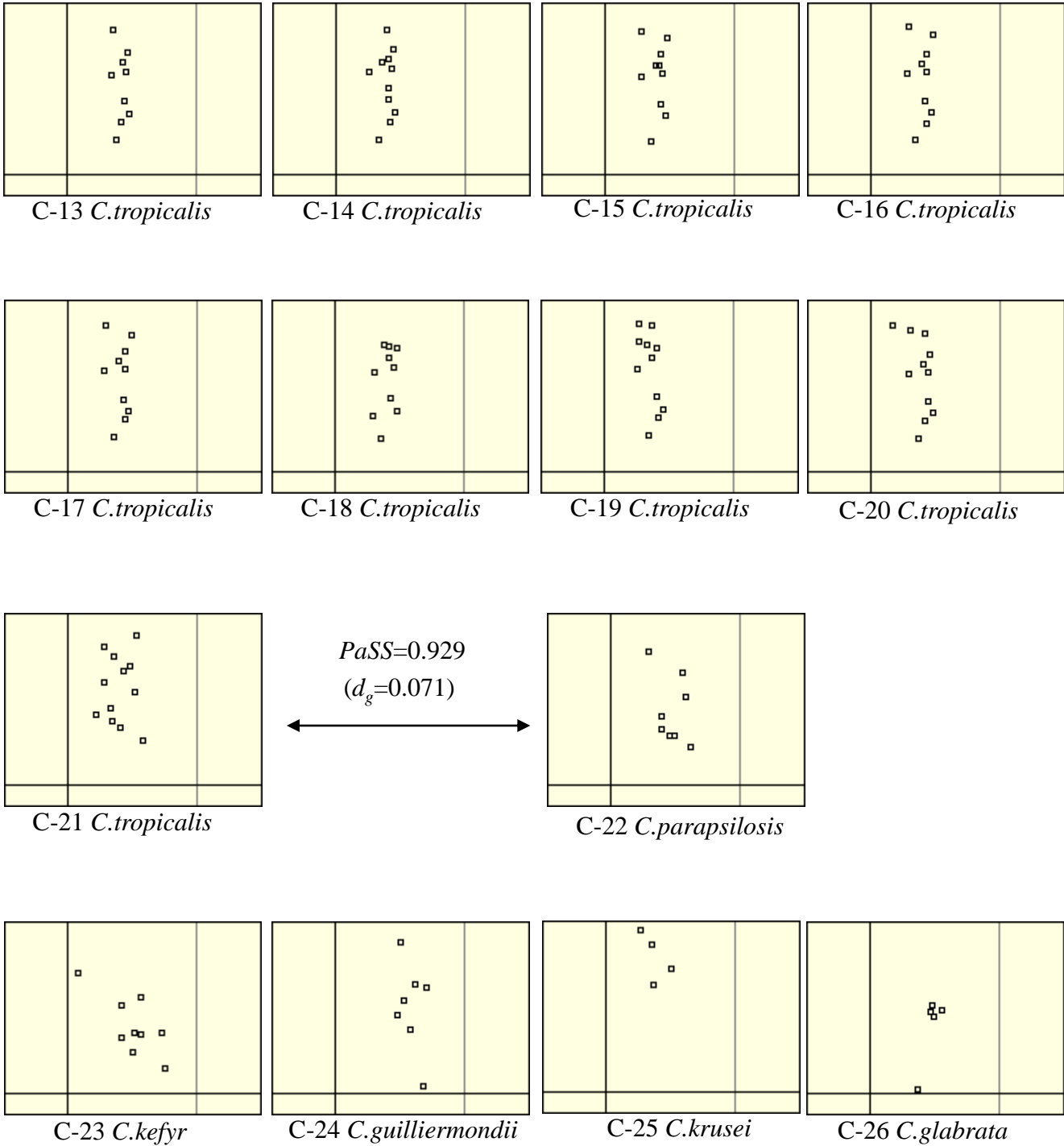


Fig. 20-2. Continued. *Spiddos* of *Candida* genus GP using TeeGee probe. TeeGee probe sequence is 5' dTTTGGGTTTGGG.

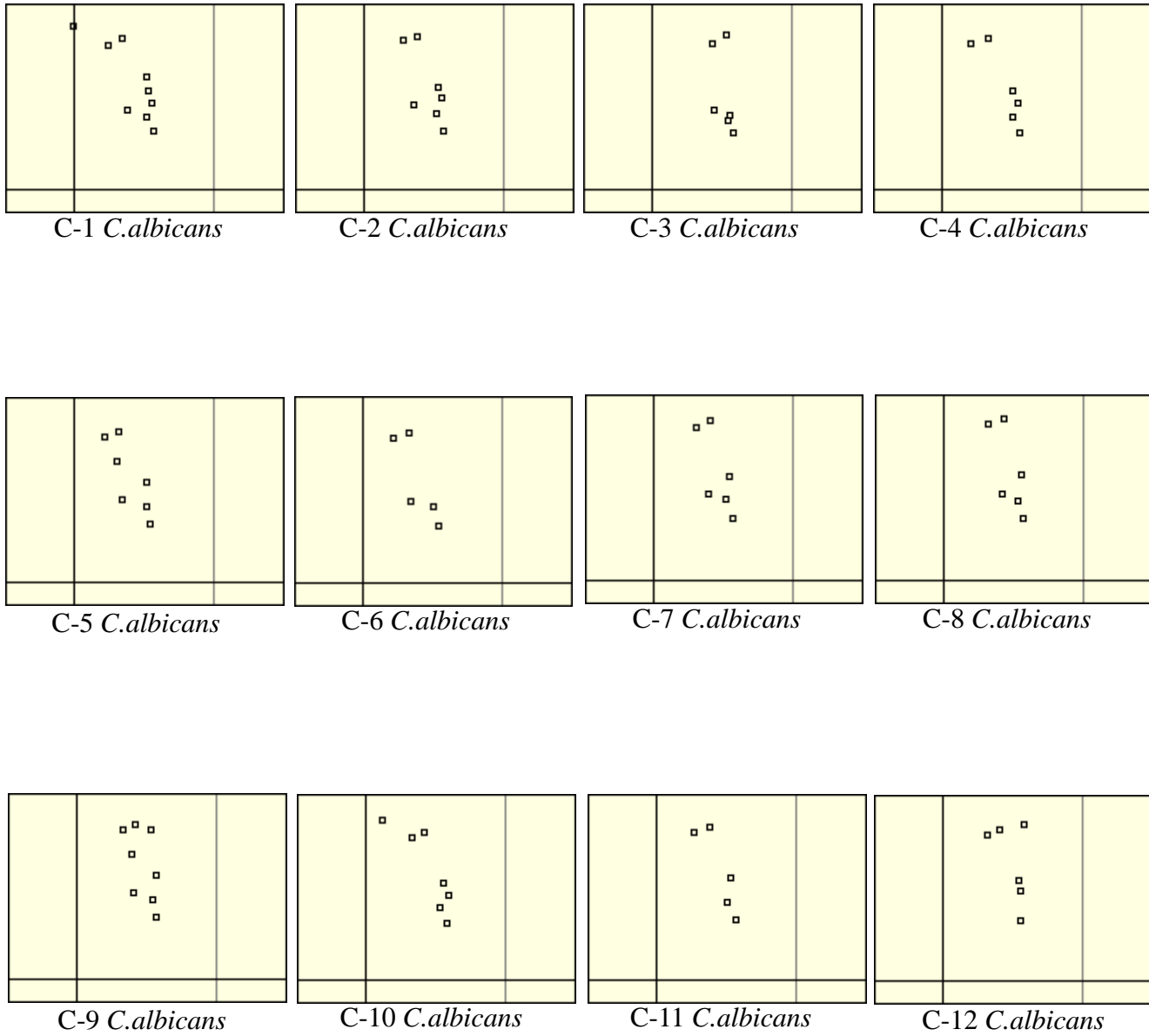


Fig. 21-1. Spidpos of *Candida* genus GP using B81 probe.  
 B81 probe sequence is 5'dGGCCGACTTGGC

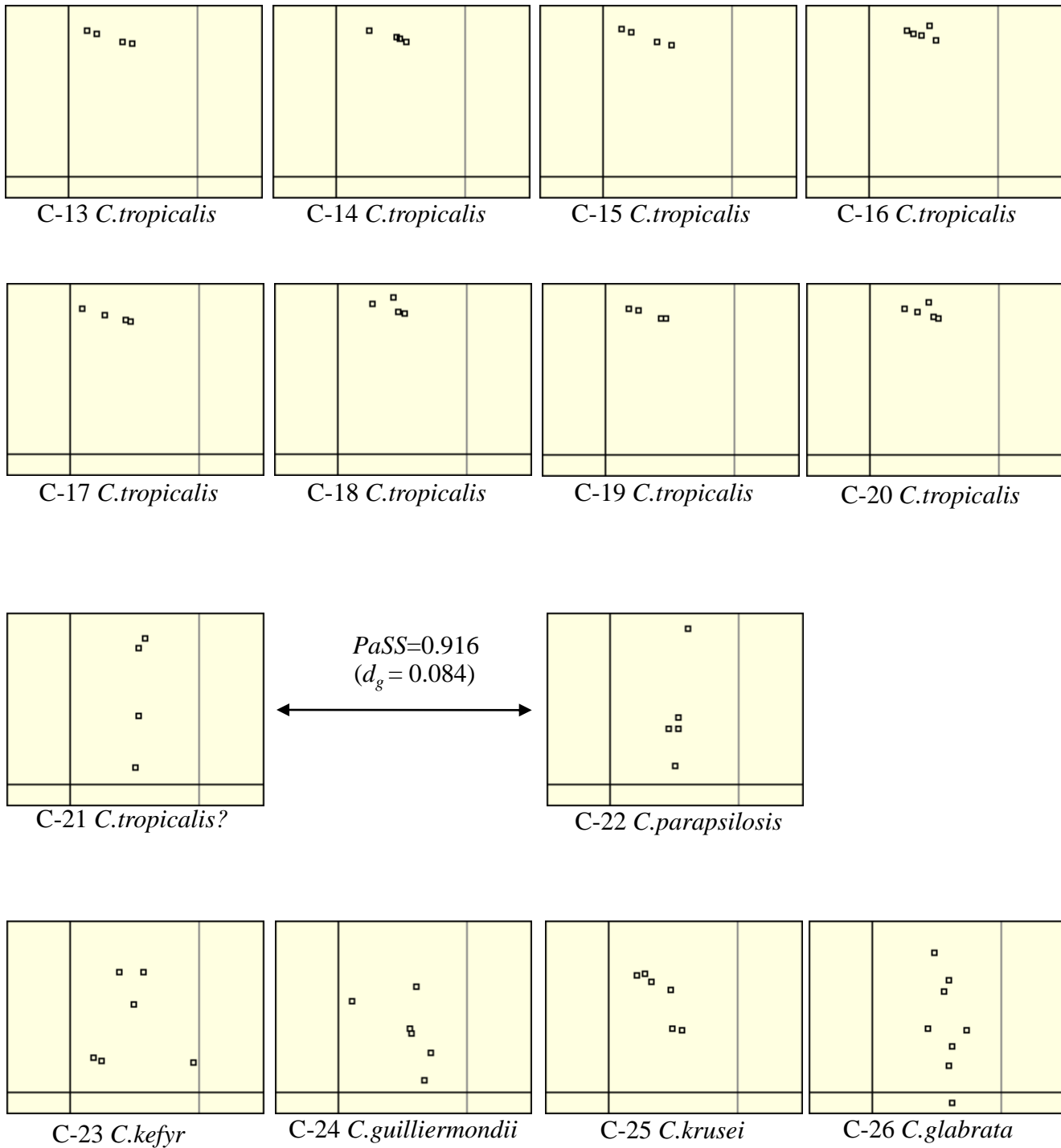


Fig. 21-2. Continued. *Spiddos* of *Candida* genus GP using B81 probe.  
B81 probe sequence is 5'dGGCCGACTTGGC

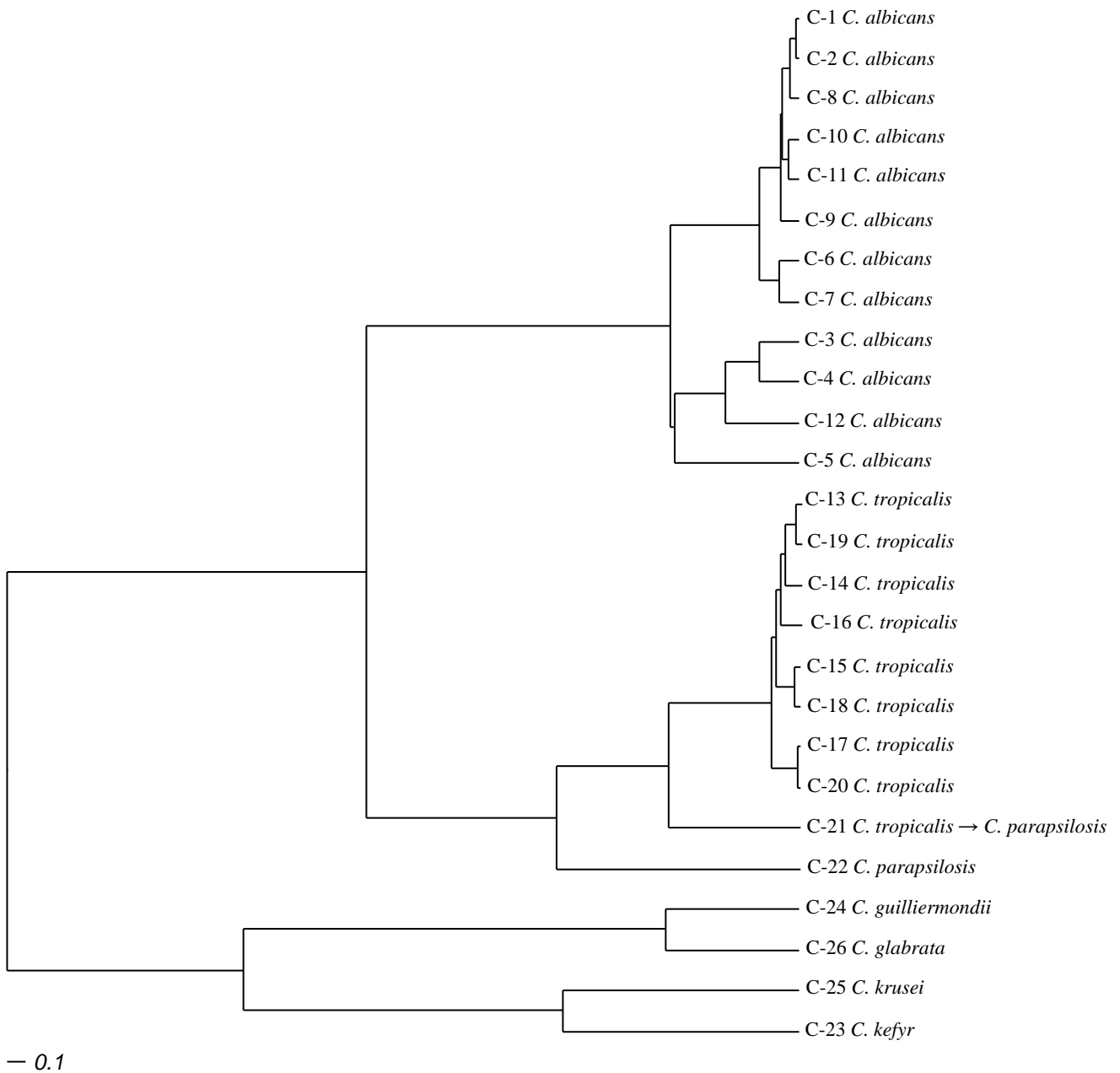


Fig. 22. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$  when pfM12 probe was used. PfM12 probe sequence is 5' dAGAACGCGCCTG.

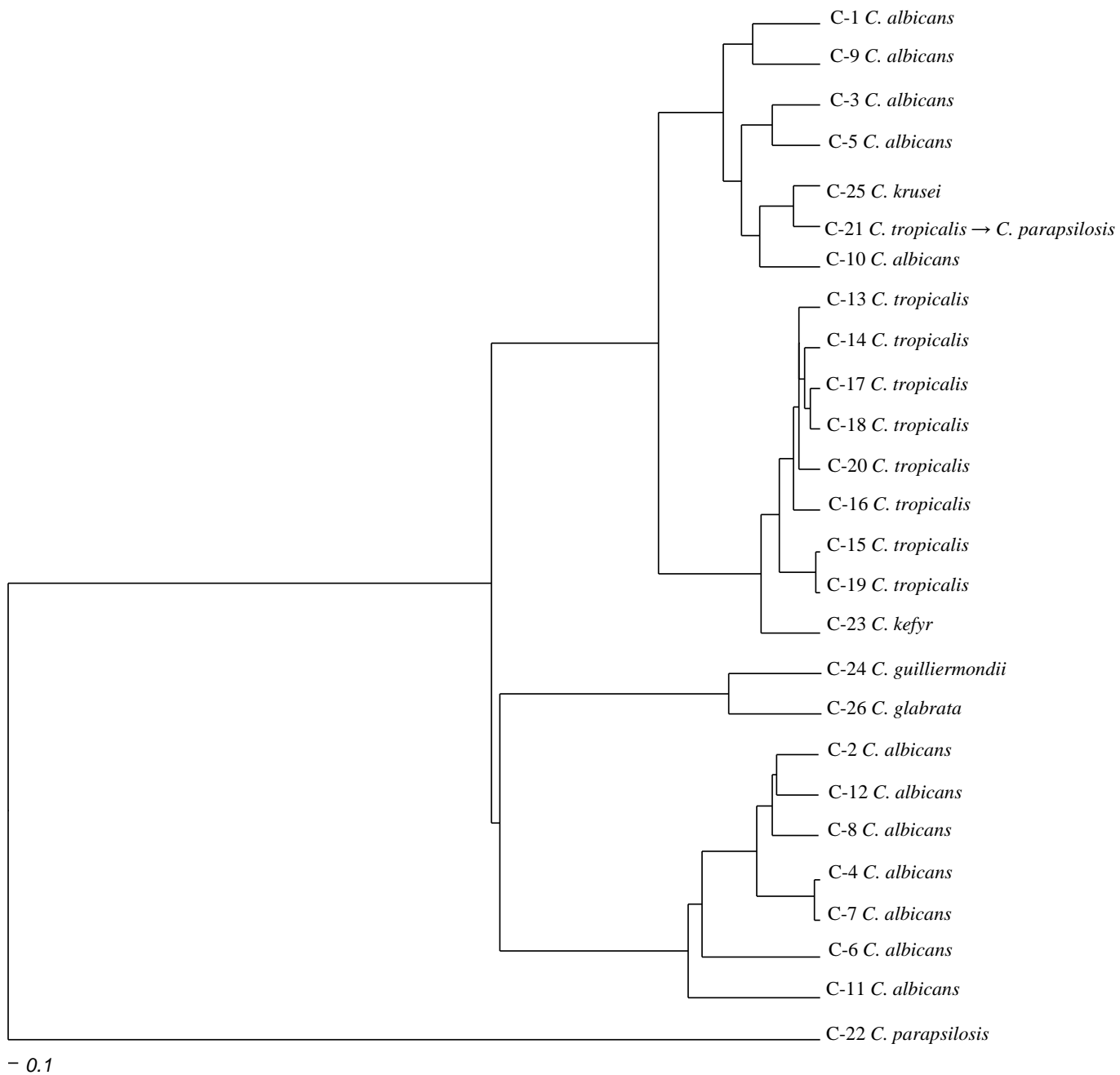


Fig. 23. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$  when pFM19 probe was used. PFM19 probe sequence is 5' dCAGGGCGCGTAC.



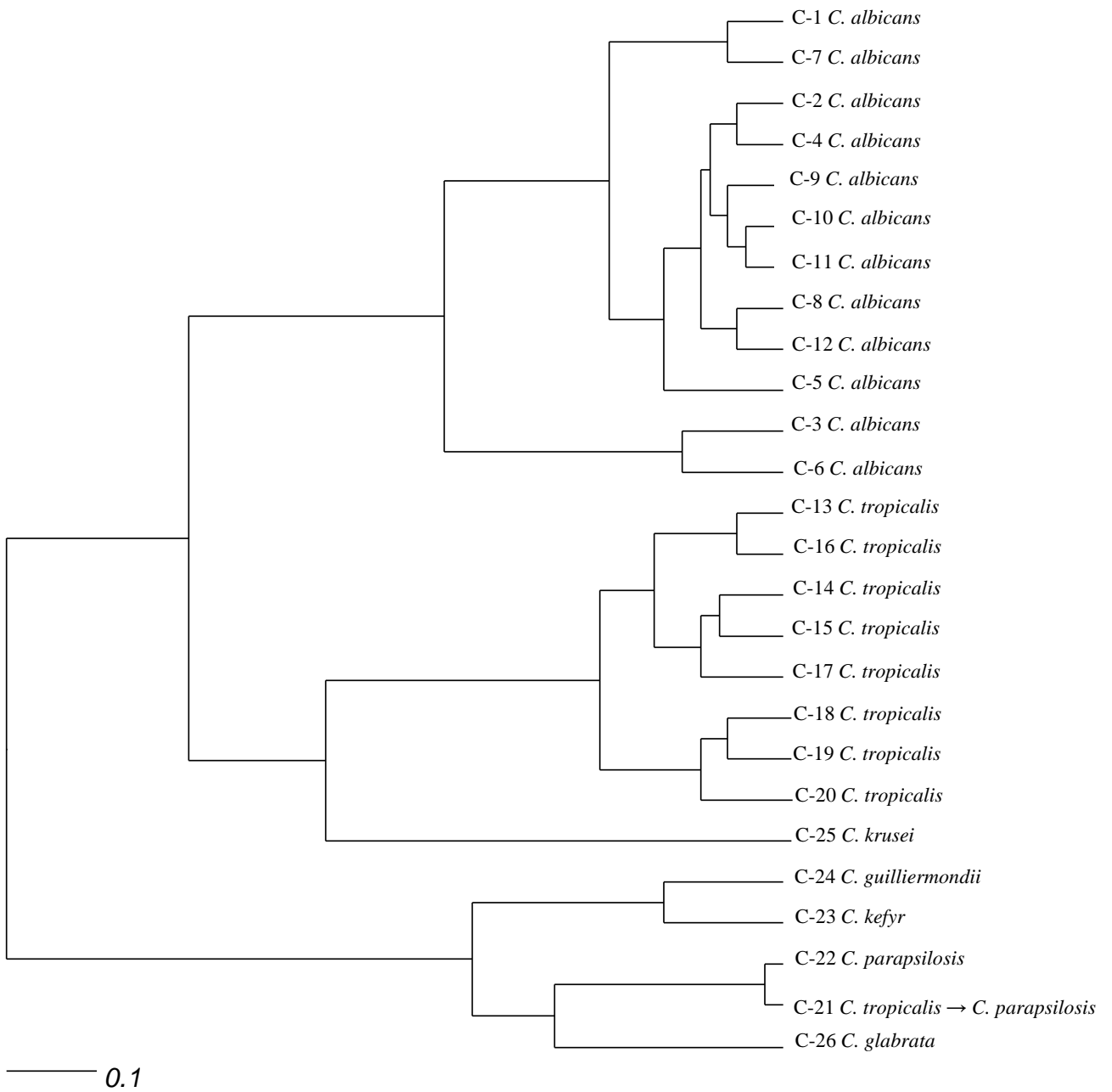


Fig. 24. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$  when Hunt probe was used. Hunt probe sequence is 5'd TGCTGCTGCTGC.

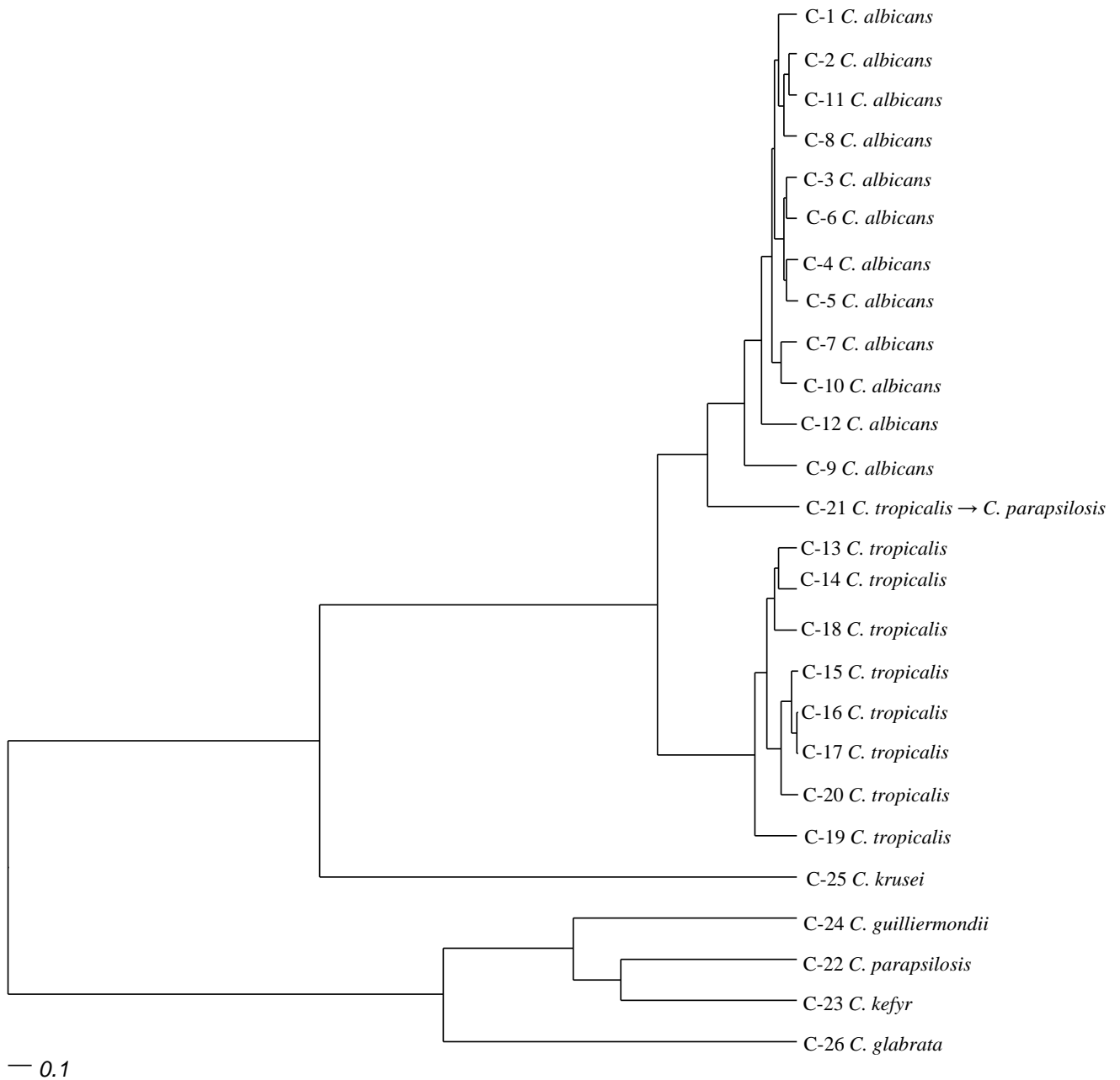


Fig. 25. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$  when TeeGee probe was used. TeeGee probe sequence is 5' dTTTGGGTTTGGG.

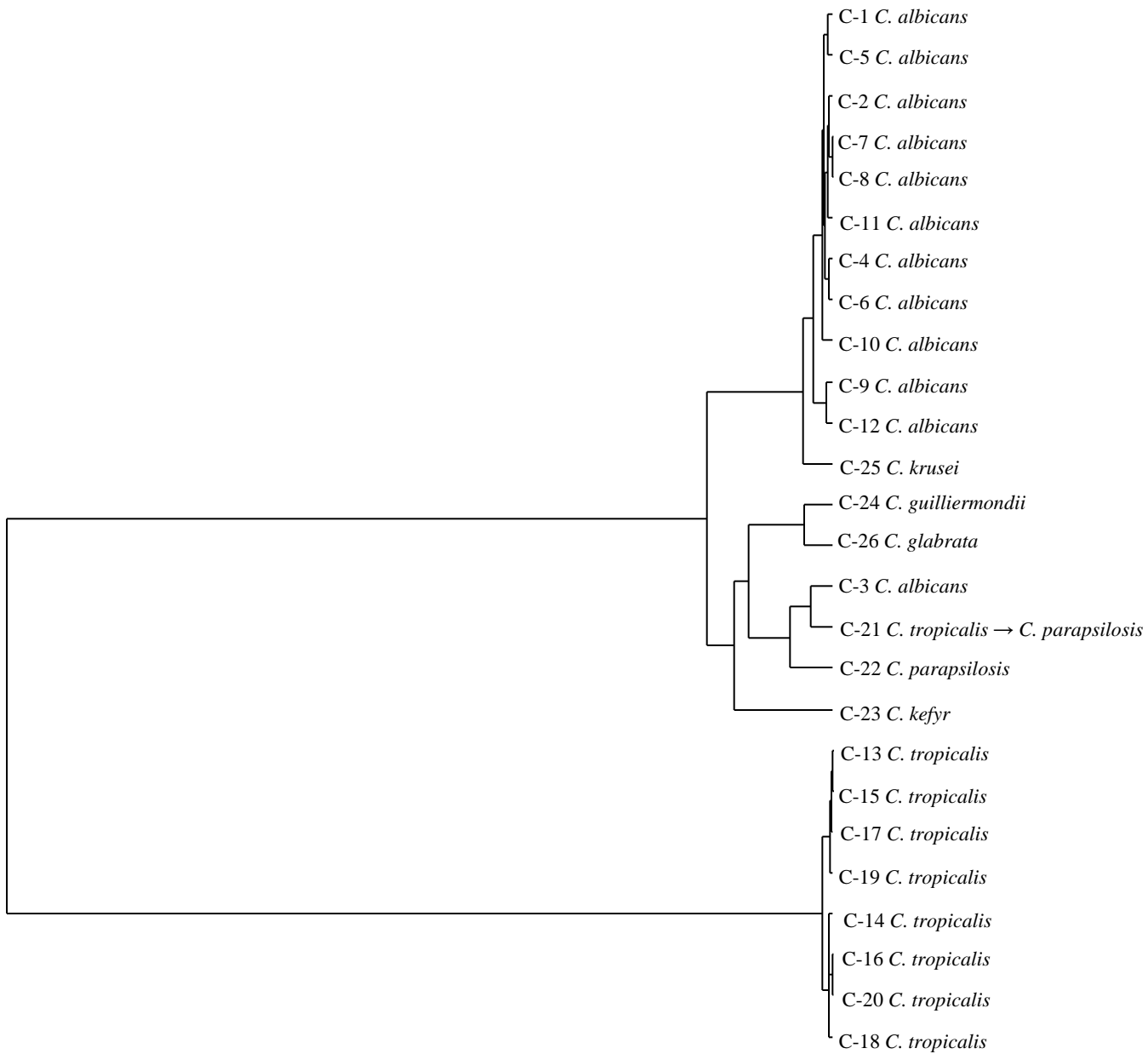


Fig. 26. Phylogenetic tree drawn based on the genome distance  $d_g (=1 - PaSS)$  when B81 probe was used. B81 probe sequence is 5'dGGCCGACTTGGC.

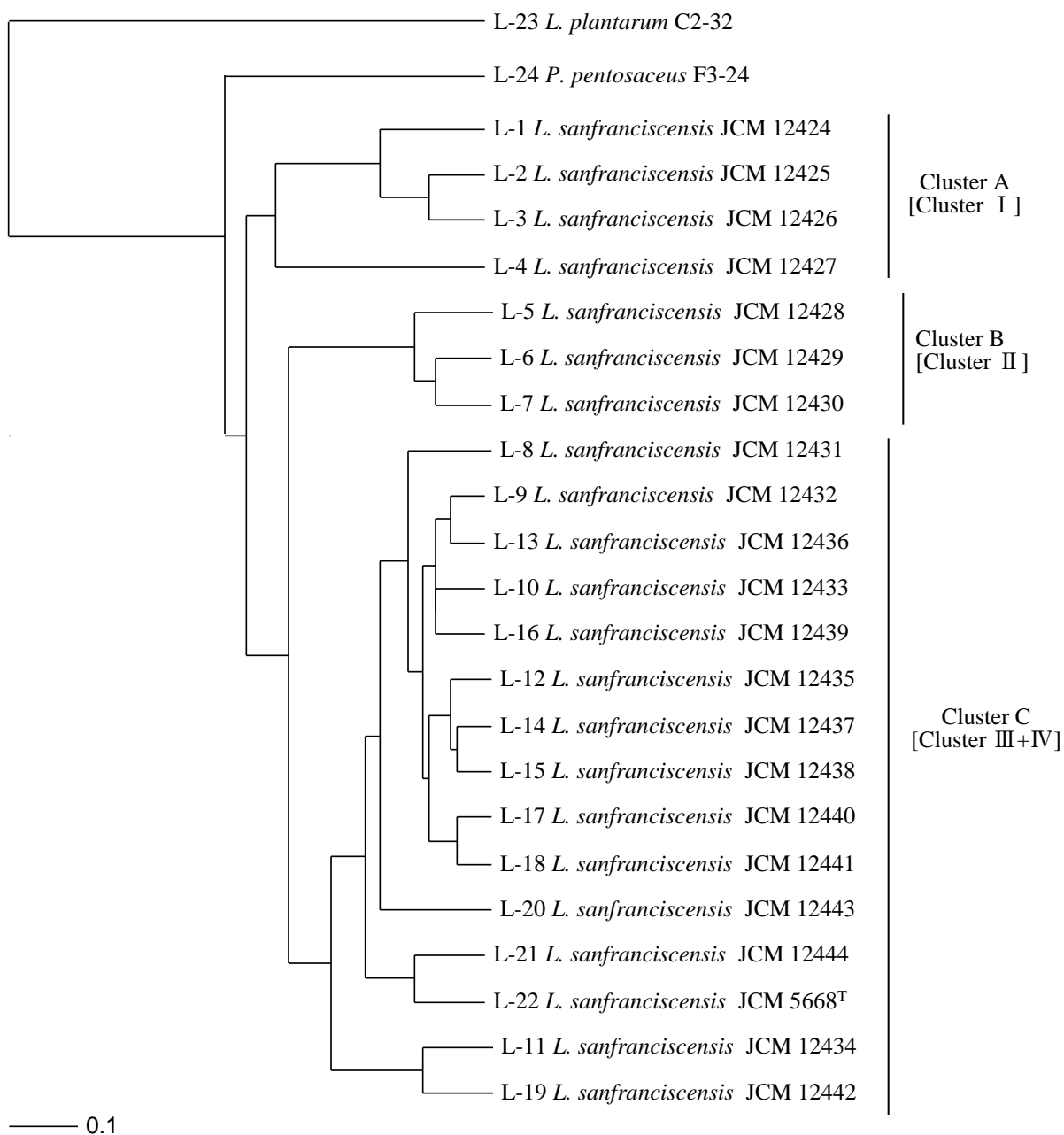


Fig. 27 Phylogenetic tree of *L. sanfranciscensis* drawn based on the genome distance  $d_g (=1 - PaSS)$ . The genome distance adopted here is the average value over the values obtained with the five different probes. UPGMA algorithm was used to make this tree. Cluster I ~ IV are used in Fig. 28.

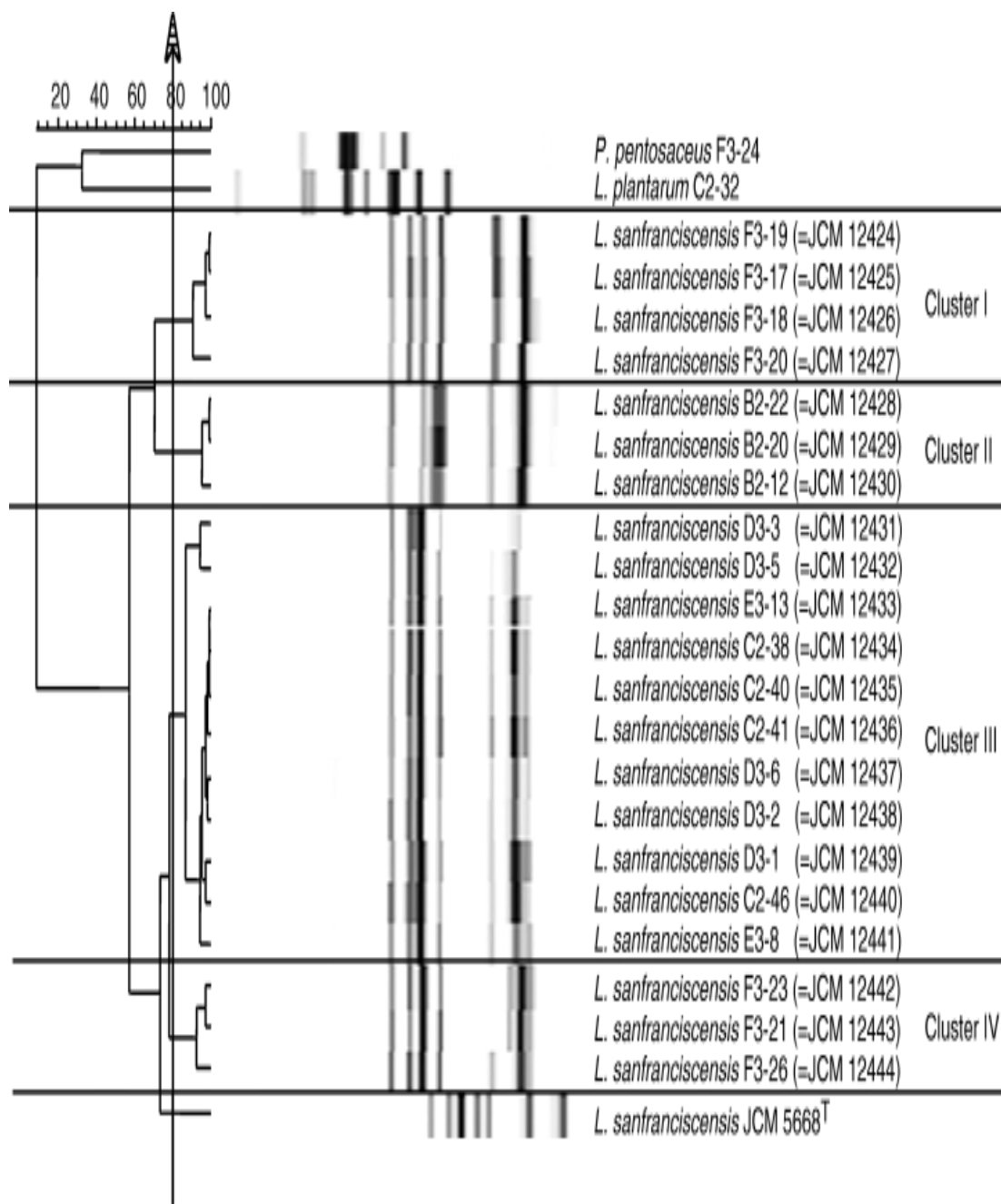


Fig. 28. Phylogenetic tree drawn based on Peason coefficient derived from ribotyping pattern by Kitahara. UPGMA algorithm was used to make this tree. This figure was taken from Letters in Applied Microbiology <sup>(49)</sup>.

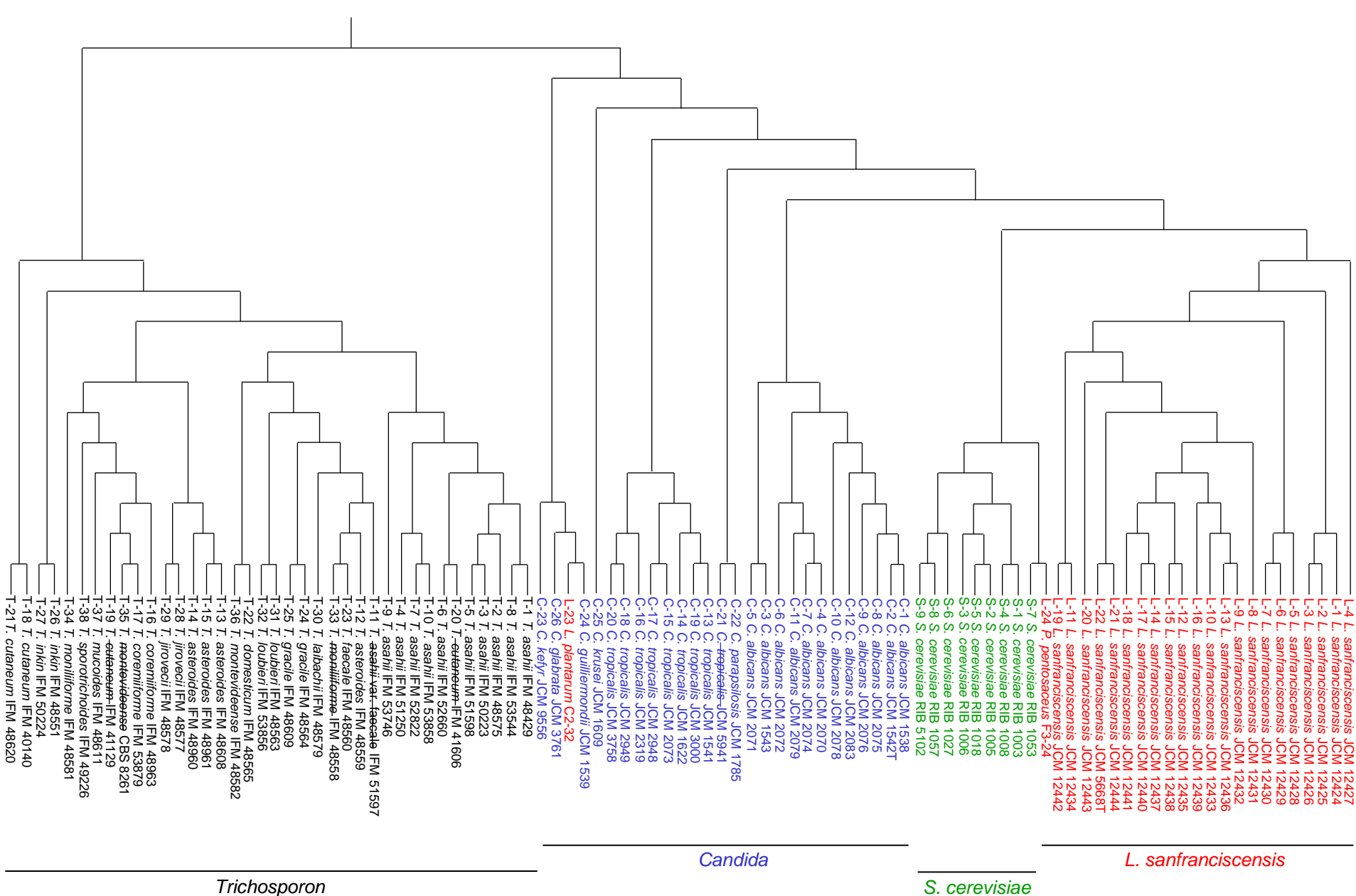


Fig. 29 Global phylogenetic tree based on the genome distance  $d_g(=1-\text{PaSS})$ . The genome distance adopted here is the average value over the values obtained with four different probes. Four different probes are pfm12, pfm19, Hunt and B81. Four mix; *Candida*, *Lactobacillus*, *Saccharomyces*, *Trichosporon* tree[rectangular cladogram].

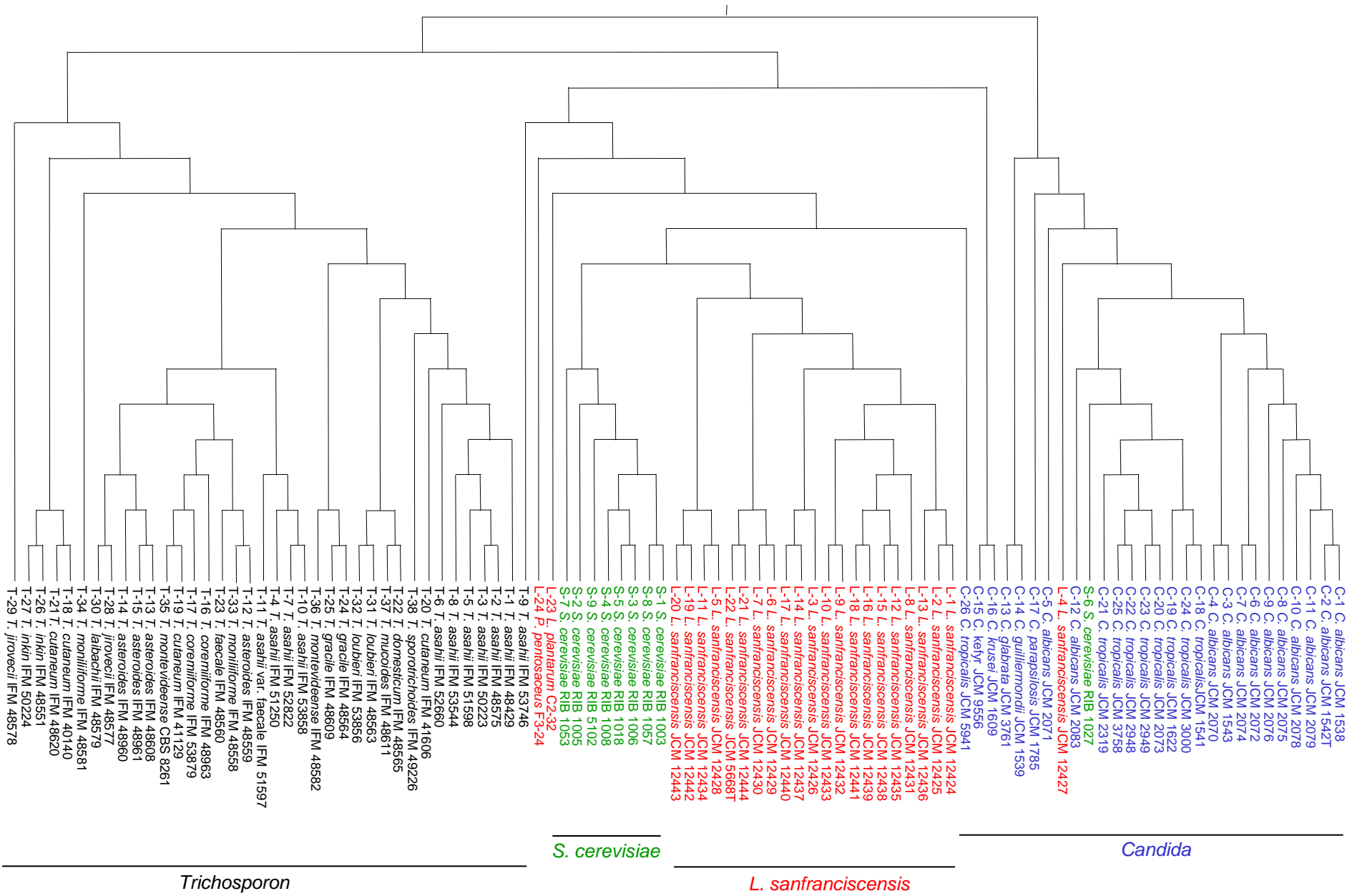
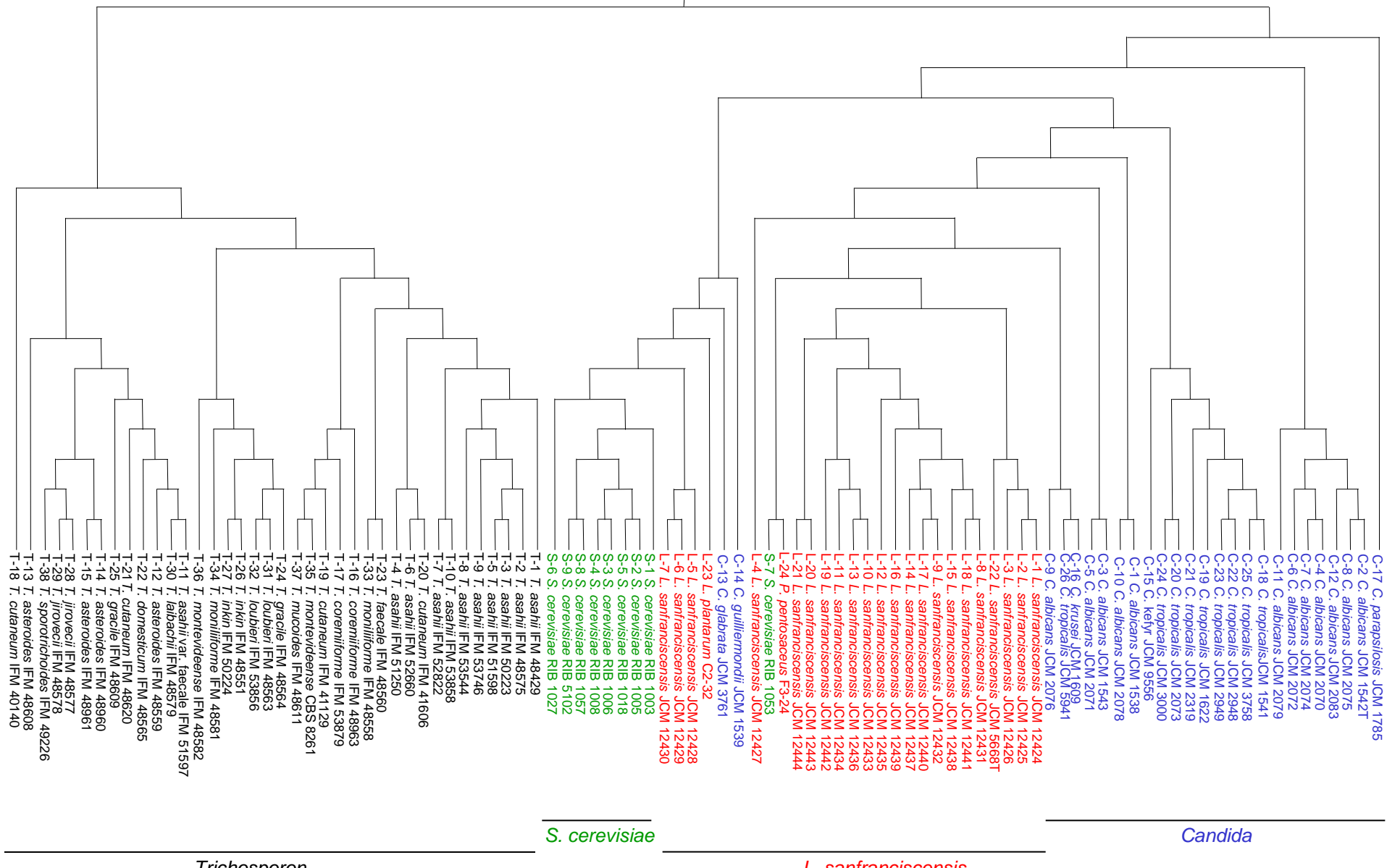


Fig. 30-1. Four mix; *Candida*, *Lactobacillus*, *Saccharomyces*, *Trichosporon* tree[rectangular cladogram]  
Using probe ; pfm12[5'dAGAACGCGCCTG]



Trichosporon

*S. cerevisiae*

*L. sanfranciscensis*

*Candida*

Fig. 30-2. Four mix; *Candida*, *Lactobacillus*, *Saccharomyces*, *Trichosporon* tree[rectangular cladogram]  
 Using probe ; pfm19[5'dCAGGGCGCGTAC]



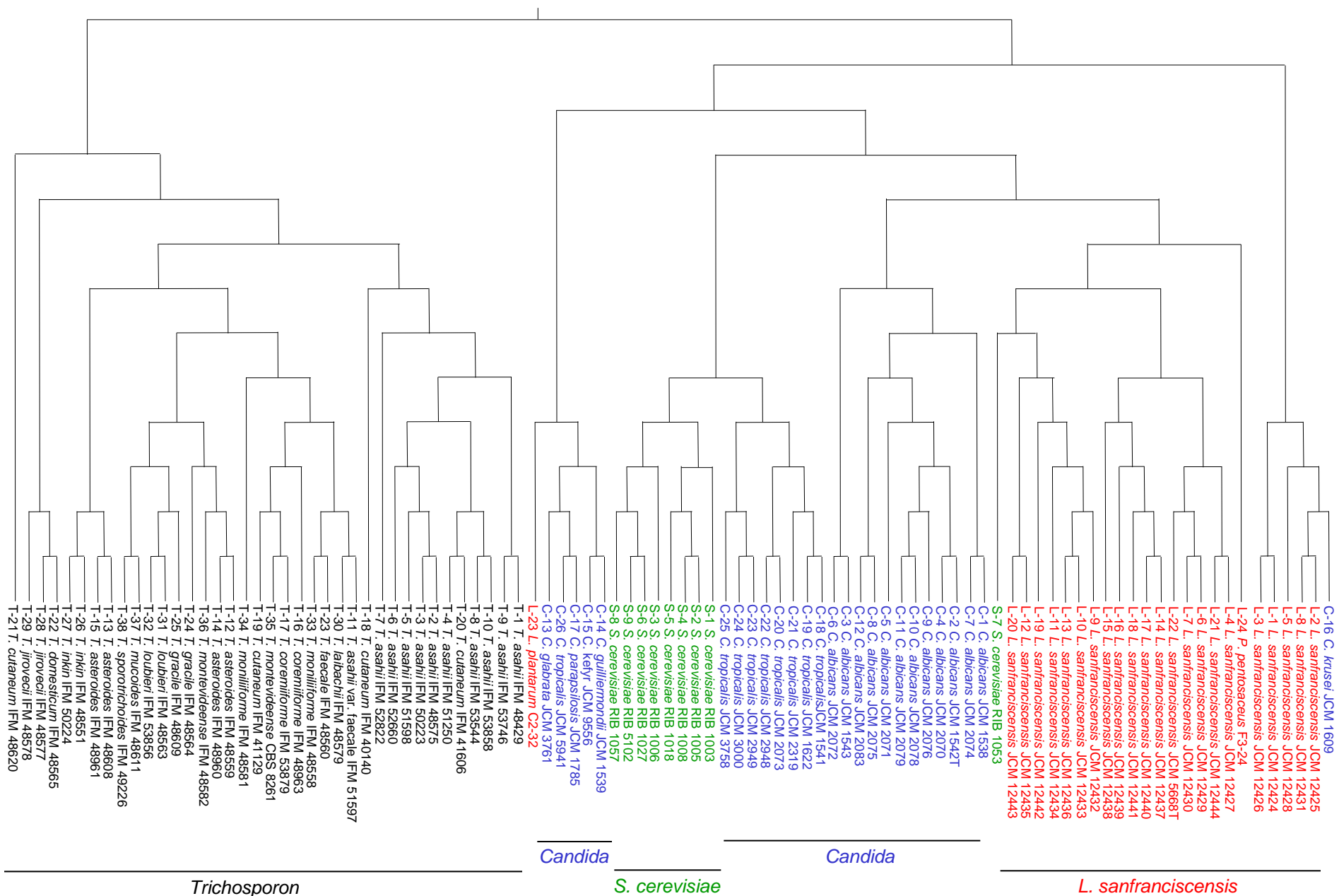


Fig. 30-3. Four mix; *Candida*, *Lactobacillus*, *Saccharomyces*, *Trichosporon* tree[rectangular cladogram]  
 Using probe ; Hunt[5' dTGCTGCTGCTGC]

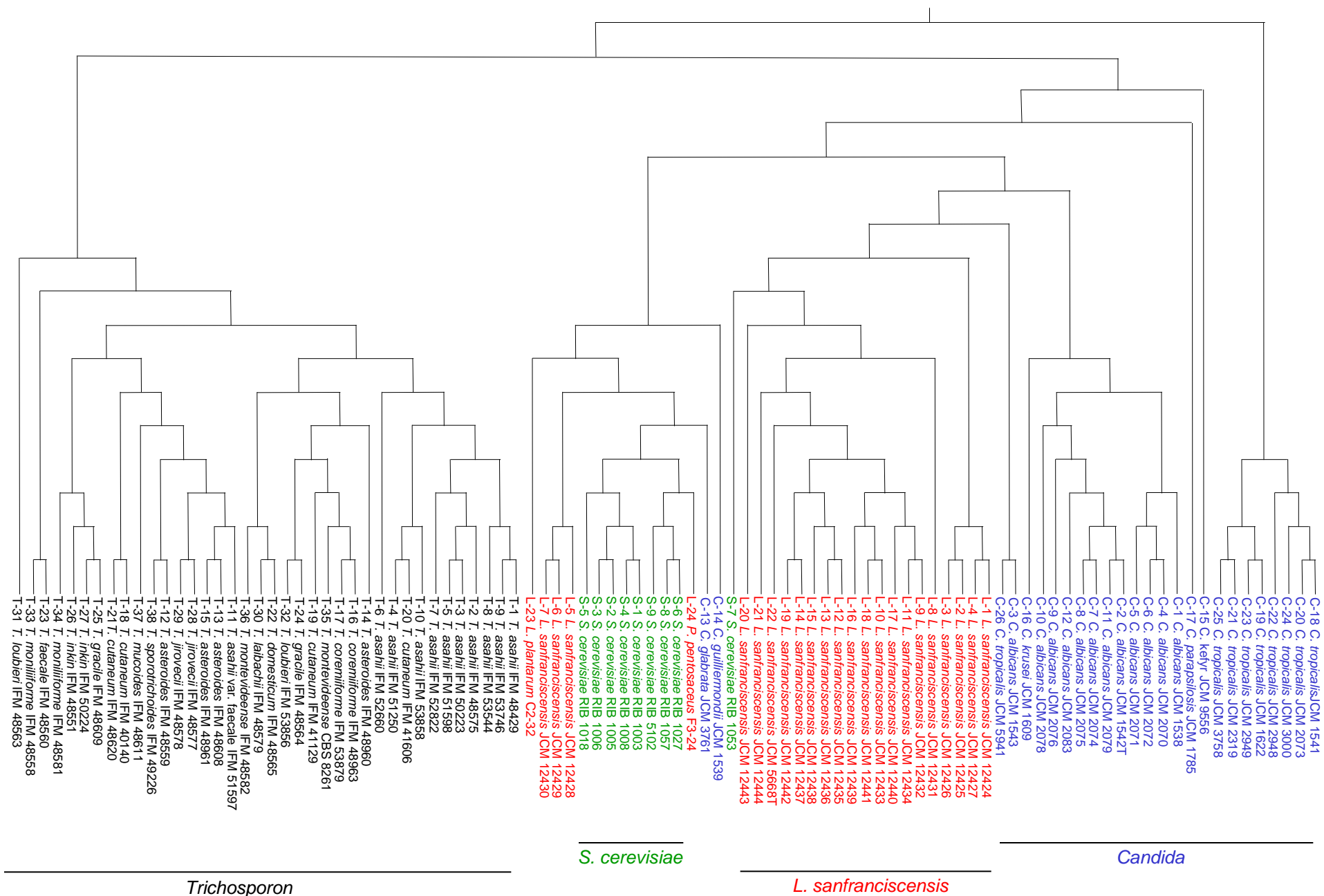


Fig. 30-4. Four mix; *Candida*, *Lactobacillus*, *Saccharomyces*, *Trichosporon* tree[rectangular cladogram]  
 Using probe ; B81[5' dGGCCGACTTGGC]

Table 1 *Trichosporon* strains used in this study

Serial No.	Strain number <sup>a</sup>	Species name (registered) <sup>b</sup>	Source <sup>a</sup>
T-1	IFM 48429	<i>Trichosporon asahii</i>	human (=CBS 2479 <sup>T</sup> )
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 <sup>T</sup> )
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 <sup>T</sup> )
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 <sup>T</sup> )
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 <sup>T</sup> )
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 <sup>T</sup> )
T-34	IFM 48581	<i>Trichosporon moniliiforme</i>	water (=CBS 5959)
T-35	CBS 8261	<i>Trichosporon montevideense</i>	human
T-36	IFM 48582	<i>Trichosporon montevideense</i>	water purification tank (=CBS 6721 <sup>T</sup> )
T-37	IFM 48611	<i>Trichosporon mucoides</i>	human (=CBS 7625 <sup>T</sup> )
T-38	IFM 49226	<i>Trichosporon sporotrichoides</i>	moist humus (=CBS 8245, JCM 9941 <sup>T</sup> )
X-1	JCM 1542 <sup>T</sup>	<i>Candida albicans</i>	skin disorder (=CBS 562, IFO 1385)
X-2	JCM 1541 <sup>T</sup>	<i>Candida tropicalis</i>	bronchitic patient (=CBS 94, IFO 1400)

<sup>a</sup>Symbols 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 culture.

<sup>b</sup>Phenotypically identified based on morphological and biophysical traits.

Table 2. *Candida* strains in this study

Serial No.	Strain number <sup>a</sup>	Species name (registered) <sup>b</sup>	History and other information <sup>a</sup>
C-1	JCM 1538	<i>Candida albicans</i>	T. Shinoda M-1012; serotype A
C-2	JCM 1542 <sup>T</sup>	<i>Candida albicans</i>	IFO 1385 =CBS 562; serotype A
C-3	JCM 1543	<i>Candida albicans</i>	IFO 1397 =CBS 1905; Type strain of <i>Candida stellatoidea</i> ; serotype B
C-4	JCM 2070	<i>Candida albicans</i>	T. Shinoda M-1447 =NIH 207; serotype A
C-5	JCM 2071	<i>Candida albicans</i>	T. Shinoda M-1445 =NIH 792; serotype B
C-6	JCM 2072	<i>Candida albicans</i>	Meiji College of Pharmacy M-2086
C-7	JCM 2074	<i>Candida albicans</i>	IFO 0197 = GRIF = M. Ota.
C-8	JCM 2075	<i>Candida albicans</i>	IFO 0579 = Y. Hamada = Natl. Hosp. Kyoto
C-9	JCM 2076	<i>Candida albicans</i>	IFO 0583 = NI 7491 = 3rd Army Area Lab. = Duke Univ. (N. F. Conant).
C-10	JCM 2078	<i>Candida albicans</i>	IFO 1060 =T. Tsuchiya J-1012; serotype A
C-11	JCM 2079	<i>Candida albicans</i>	IFO 1061 =T. Tsuchiya J-1011
C-12	JCM 2083	<i>Candida albicans</i>	IFO 1388 =CBS 2688
C-13	JCM 1541 <sup>T</sup>	<i>Candida tropicalis</i>	IFO 1400 =CBS 94
C-14	JCM 1622	<i>Candida tropicalis</i>	IAM 4965 =IFO 0006
C-15	JCM 2073	<i>Candida tropicalis</i>	T. Shinoda M-1017 = T. Tsuchiya J-1017= Harvard Univ.= ATCC 7349 = A. Castellani.
C-16	JCM 2319	<i>Candida tropicalis</i>	AJ 5962 = T. Tsuchiya J-1473 = CBS 5701 = E. K. Novák.
C-17	JCM 2948	<i>Candida tropicalis</i>	D. G. Ahearn 78-050
C-18	JCM 2949	<i>Candida tropicalis</i>	D. G. Ahearn 80-005
C-19	JCM 3000	<i>Candida tropicalis</i>	M. Suzuki T-25 = W. Daengsubha; papaya
C-20	JCM 3758	<i>Candida tropicalis</i>	CBS 8072; Type strain of <i>Candida paratropicalis</i>
C-21	JCM 5941	<i>Candida tropicalis</i>	K. Tokuoka 118-40A; lemon cake
C-22	JCM 1785 <sup>T</sup>	<i>Candida parapsilosis</i>	AJ 5970 =CBS 604
C-23	JCM 9556 <sup>T</sup>	<i>Candida kefyr</i>	IFO 10287 =CBS 834
C-24	JCM 1539 <sup>T</sup>	<i>Candida guilliermondii</i>	T. Shinoda (CBS 566)
C-25	JCM 1609 <sup>T</sup>	<i>Candida krusei</i>	IAM 12186 =CBS 573
C-26	JCM 3761 <sup>T</sup>	<i>Candida glabrata</i>	CBS 138

<sup>a</sup> Symbols and abbreviations: JCM = Japan Collection of Microorganisms, RIKEN BioResource Center, Wako, Japan; CBS = Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; IFO = Institute for Fermentation, Osaka, Japan; IAM =Institute of Molecular and Cellular Biosciences, The University of Tokyo, Japan; T = type culture.

<sup>b</sup> Phenotypically identified based on morphological and biophysical traits.

Table 3 *Lactobacillus sanfranciscensis* strains in this study.

Serial No.	Strain number <sup>a</sup>	Species name (registered) <sup>b</sup>	Source <sup>c</sup>
L-1	JCM 12424	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-2	JCM 12425	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-3	JCM 12426	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-4	JCM 12427	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-5	JCM 12428	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough B in USA,
L-6	JCM 12429	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough B in USA,
L-7	JCM 12430	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough B in USA,
L-8	JCM 12431	<i>Lactobacillus sanfranciscensis</i>	rye sourdough D in Germany
L-9	JCM 12432	<i>Lactobacillus sanfranciscensis</i>	rye sourdough D in Germany
L-10	JCM 12433	<i>Lactobacillus sanfranciscensis</i>	rye sourdough E in Germany
L-11	JCM 12434	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough C in Germany
L-12	JCM 12435	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough C in Germany
L-13	JCM 12436	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough C in Germany
L-14	JCM 12437	<i>Lactobacillus sanfranciscensis</i>	rye sourdough D in Germany
L-15	JCM 12438	<i>Lactobacillus sanfranciscensis</i>	rye sourdough D in Germany
L-16	JCM 12439	<i>Lactobacillus sanfranciscensis</i>	rye sourdough D in Germany
L-17	JCM 12440	<i>Lactobacillus sanfranciscensis</i>	wheat sourdough C in Germany
L-18	JCM 12441	<i>Lactobacillus sanfranciscensis</i>	rye sourdough E in Germany
L-19	JCM 12442	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-20	JCM 12443	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-21	JCM 12444	<i>Lactobacillus sanfranciscensis</i>	rye sourdough F in Germany
L-22	JCM 5668 <sup>T</sup>	<i>Lactobacillus sanfranciscensis</i>	San Francisco sourdough
L-23	–	<i>Lactobacillus plantarum</i> C2-32	–
L-24	–	<i>Pediococcus pentosaceus</i> F3-24	–

<sup>a</sup> Symbols and abbreviations: JCM = Japan Collection of Microorganisms, RIKEN BioResource Center, Wako, Japan; T = type culture.

<sup>b</sup> Genetically identified based on species-specific PCR.

<sup>c</sup> Written in original paper<sup>(49)</sup>: Kitahara, M., Sakata, S. and Benno, Y. Biodiversity of *Lactobacillus sanfranciscensis* strains isolated from five sourdoughs. *Lett. Appl. Microbiol.* 40, 353-357, 2005.

Table 4 *Saccharomyces cerevisiae* strains in this study.

Serial No	Strain number <sup>a</sup>	Species name (registered) <sup>b</sup>	Source <sup>a</sup>
S-1	RIB 1003	<i>Saccharomyces cerevisiae</i>	Sake yeast Kyokai No.7
S-2	RIB 1005	<i>Saccharomyces cerevisiae</i>	Sake yeast Kyokai No.8
S-3	RIB 1006	<i>Saccharomyces cerevisiae</i>	Sake yeast Kyokai No.9
S-4	RIB 1008	<i>Saccharomyces cerevisiae</i>	Sake yeast Kyokai No.10
S-5	RIB 1018	<i>Saccharomyces cerevisiae</i>	Shochu yeast S-2
S-6	RIB 1027	<i>Saccharomyces cerevisiae</i>	Wine yeast W-3, Yamanashi University
S-7	RIB 1053	<i>Saccharomyces cerevisiae</i>	Wine yeast California Champagne, Red Star
S-8	RIB 1057	<i>Saccharomyces cerevisiae</i>	Wine yeast OC-2, Geisenheim
S-9	RIB 5102	<i>Saccharomyces cerevisiae</i>	Wine yeast RIFY 7402 (=IFO 2231),Liebfrauen Milch,

<sup>a</sup> Symbols and abbreviations: RIB = National Research Institute of Brewing, Hiroshima, Japan; RIFY = Institute of Enology and Viticulture, Kofu, Japan; IFO = Institute for Fermentation, Osaka, Japan.

<sup>b</sup> Phenotypically identified based on morphological and biophysical traits.

Table 5 Species identification by GP and Sequencing. This figure was taken from Journal of Microbiological Methods <sup>(69)</sup>.

Serial No. <sup>a</sup> / Labeled name	Closest species <sup>b</sup> ( <i>PaSS</i> score)	<i>ccgf</i> <sup>c</sup>	Inference by GP <sup>d</sup>	Assignment by sequencing <sup>e</sup> (DDBJ Accession No.)	GP's Score <sup>f</sup>
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> ] <sup>h</sup> (AB492267)	○!
T-12 / <i>T. asteroides</i> IFM 48559	T-14 (0.943) <sup>g</sup>	*	- / + [=T-14]	+ (AB492252)	△
T-13 / <i>T. asteroides</i> IFM 48608	T-15 (0.952) <sup>g</sup>	*	- / + [=T-15]	- [ <i>T. japonicum</i> ] <sup>h</sup> (AB492268)	○!
T-14 / <i>T. asteroides</i> IFM 48960	T-12 (0.943) <sup>g</sup>	*	- / + [=T-12]	+ (AB492251)	△
T-15 / <i>T. asteroides</i> IFM 48961	T-13 (0.952) <sup>g</sup>	*	- / + [=T-13]	- [ <i>T. japonicum</i> ] <sup>h</sup> (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)	○

<sup>a</sup> The serial number is given in the alphabetical order of the species nomenclature.

<sup>b</sup> Inferred from *PaSS* scores (written in parentheses) for skeptic cases.

<sup>c</sup> Assigned to be positive and marked with an asterisk (\*) if there are evident *ccgf* bands that support the *PaSS*-based assignment (see Ref. 55) Naimuddin M, T. Kurazono T, Nishigaki K: Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution. *Nucleic Acids Res* 2002, 30: e42.).

<sup>d</sup> 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.

<sup>e</sup> D1/D2 region of 27S rDNA was sequenced and BLAST hit (best scored species) is shown. If the result agreed with the labeled name, then '+', else '-'.  
<sup>f</sup> 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.

<sup>g</sup> 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 ( $d_g = 0.074$ ) between pairs than the average  $d_g$  of the other *Trichosporon* species (0.050), indicating that these pairs belong to different species.

<sup>h</sup> Species not included explicitly before this analysis, which cannot be named at 'Inference by GP'.

**Table 6: Estimation of the cost and duration required for the sequencing and the GP approach (As for 2013).**

	GP experiment	r DNA sequencing
Experimental cost per run <sup>a</sup>	US\$2.3	US\$11.2
Experimental duration	4 hours	7 hours
Performance/day/person <sup>b</sup>	30 <sup>c</sup>	96
Reagents used	PCR reagents, acrylamide gel, fluorescent primer, gel cassette and buffers	PCR reagents, PCR product purification kit, cycle sequencing kit
Initial cost <sup>d</sup>	US\$21,820.0	US\$530,000.0

<sup>a</sup> The cost estimation is made based on the data shown in Table 7.

<sup>b</sup> A sequencer having the power of 96 samples processing per run assumed.

<sup>c</sup> Well-trained person assumed.

<sup>d</sup> The sequencer is, for example, Applied Biosystems 3730xl DNA Analyzer, its standard price is 530,000 dollars in japan, and  $\mu$ TGGE apparatus is the  $\mu$ -TG (TAITEC, Japan). Although the confocal laser imager was used in this research, the combinatorial photography set of CCD camera and UV transilluminator, which is used after dyeing gels with ethidium bromide or SybrGold, is enough even if researchers do not have an expensive device such a confocal laser imager. The breakdown of initial cost of GP experiment here is 11,820 dollars of  $\mu$ -TG and 10,000 dollars of combinatorial photography set of CCD camera and UV transilluminator.



**Table 7: Tentative estimation of the experimental cost for the sequencing and the GP experiments.**

Experimental Step	Reagents <sup>a</sup>	Kit size	Price	Cost for a single run	
				GP experiment <sup>b</sup>	r DNA sequencing <sup>c</sup>
Random PCR or PCR	PCR reagents	1kit (250 unit)	27000 yen	108 yen <sup>d</sup>	108 yen
	Probe (Primer)	0.2 μmol	15000 - 20000 yen	1.2 yen <sup>d</sup>	2.8 yen
Sequencing	QIAquick PCR Purification Kit	1kit (50 unit)	13000 yen	-	260 yen
	Sequencing	96 tests	67200 yen	-	700 yen
μTGGE	Gel cassette	1kit (100 unit)	9000 yen	90 yen	-
μTGGE and others	Others (Chips, acrylamide, urea, Tris, etc.)	-	-	30 yen	50 yen
Total cost				229.2 yen	1120.8 yen
(in US\$ <sup>e</sup> )				(US\$2.3)	(US\$11.2)
Ratio				1/5	1

<sup>a</sup> Standard costs available in Japan 2013 are shown.

<sup>b</sup> The running cost for the GP experiment can be attributed to random PCR and μTGGE. Here, the initial cost of implementation of the devices (μ-TG, power supply, UV transilluminator) is written on Table 6, costs for facilities like a confocal laser imager and labors are not considered. The running cost of GP becomes ¥108 + ¥1.2 + ¥90 + ¥30 = ¥229.2.

The cost calculation become  $7 \times ¥229.2 = ¥1,604$ /sample for the phylogenetic tree making in this research because every seven kind of probe was used for random PCR to genome DNA of one strain. The scale of random PCR was assumed to be 30μl. The size of μTGGE gel was 2.5cm × 2.5cm × 0.1cm.

<sup>c</sup> Here, the direct sequencing of PCR products using the QIAquick PCR Purification kit (Qiagen) is assumed.

<sup>d</sup> This is the cost for a single probe (primer). If seven probes used, multiply by 7.

<sup>e</sup> Currency conversion rate on November 15, 2013: 100 yen (Japan) = 1 USD.