© 1991 Oxford University Press

Genomic fingerprints produced by PCR with consensus tRNA gene primers

John Welsh and Michael McClelland*
California Institute of Biological Research, 11099 North Torrey Pines Rd, La Jolla, CA 92037, USA

Received October 18, 1990; Revised and Accepted December 14, 1990

ABSTRACT

The polymerase chain reaction using only a single 'consensus' tRNA gene primer, or a pair of primers facing outward from tRNA genes, amplifies a set of DNA fragments in bacterial, plant and animal genomic DNAs. Presumably, these PCR fingerprints are mainly derived from the regions between closely linked tRNA genes. The pattern of the PCR products is determined by which genomes and which primer(s) are used. Genomic fingerprints are largely conserved within a species and, in bacteria, most products in the fingerprint are conserved between closely related species. Thus, PCR with tRNA gene consensus primers helps to identify species and genera.

INTRODUCTION

tRNA genes occur in multiple copies dispersed throughout the genome in most species (e.g. 1, 2, 3). The shared sequence motifs of tRNA genes implies that primers for the polymerase chain reaction that contain consensus tRNA sequences are likely to result in a number of characteristic PCR products. Experience with fingerprinting of genomes using PCR with arbitrary primers (AP-PCR) (4, 5) and Alu-PCR of the human genome (6, 7), indicates that complex fingerprints can be quite reproducible. Furthermore, the fingerprints obtained by consensus tRNA primers might be relatively stable over evolutionary time scales since the organization of tRNA genes is expected to change relatively slowly. This is in contrast to AP-PCR, which is very sensitive to DNA sequence changes and which is an excellent way to detect intra-specific polymorphisms.

tDNA-PCR, so called to indicate that DNA encoding tRNAs is the target for amplification, is likely to give fingerprints that vary at the species or genus level as the tRNA gene clusters evolve. Thus, the fingerprints would be a measure of DNA relatedness at this level of classification.

METHODS

Strains

See Table 1. All *Staphylococcus* strains were kindly provided by W.E. Kloos of North Carolina State Univ. except ISP-8 from Peter Pattee, Iowa State, Ful, and those from the American Type Culture Collection. DNA was prepared as in (4).

DNAs from the human pathogenic strains of *Streptococcus pyogenes*, *S. mutans* and *Enterococcus faecalis* were all kindly supplied by Susan Hollingshead (Univ. of Alabama, Birmingham, AL).

Total genomic DNA from maize and rice strains were kindly provided by Rhonda Honeycutt, (Iowa State U., Ames IA). Human DNAs from normal intestines were kindly provided by Manuel Perucho (CIBR, CA).

Primers

The primers T5A (5'AGTCCGGTGCTCTAACCAACTG-AG3'), T5B (5'AATGCTCTACCAACTGAACT3') and T3A (5'GGGGGTTCGAATTCCCGCCGGCCCCA3') were obtained from Genosys, Houston, TX.

PCR amplification

Unless otherwise specified, 50 μl reactions were prepared using 1.25 units of Taq polymerase, 1×buffer (Stratagene), 0.2 mM of each dNTP and 5 μCi alpha-[^{32}P] dCTP, 1 μM primer (or primers) and various amounts of template DNA as indicated in the figure legends. The reaction was overlaid with oil and cycled 40 times through the following temperature profile: 94°C for 30 sec. to denature, 50°C for 30 sec. for annealing of primer and 72°C for 2 min. for extension, unless otherwise specified. The samples were stored at 4°C.

The resulting products were resolved by electrophoresis in 1×TBE through 5% Acrylamide-50% Urea and visualized by autoradiography using Kodak X-Omat™ AR film with an intensifying screen at −70°C for 6 hours.

RESULTS

Rationale

We wished to develop a PCR fingerprinting method that would rapidly identify species. Candidate targets for PCR primers are the tRNA genes. These genes are repeated a large number of times in most genomes and tend to be clustered. In *E. coli*, for instance, there are estimated to be at least 100 tRNA genes of which about 30 are mapped (1). About half the mapped genes

* To whom correspondence should be addressed

Table 1. Strains and the experiments in which they appear

Species	Primer	Figure	Lane
Staphylococcus			
S. haemolyticus 29970	T5A+T3A	2	20
S. haemolyticus CC 12J2	T5A+T3A	2	21
	T5A	1	1
	T3A	1	6
	T5A+T3A	1	11
S. haemolyticus PAY 9F2	T5A+T3A	2	22
S. haemolyticus AW 263	T5A+T3A	2	23
S. haemolyticus MID 563	T5A+T3A	2	24
S. hominis 27844	T5A	1	2
	T3A	1	7
	T5A+T3A	1	12
	T5A+T3A	2	25
S. hominis 27846	T5A+T3A	2	26
S. warneri CPB10E2	T5A+T3A	2	27
	T5A	1	3
	T3A	1	8
	T5A+T3A	1	13
S. warneri GAD473	T5A+T3A	2	28
S. warneri MCY3E6	T5A+T3A	2	29
S. warneri PBNZP4D3	T5A+T3A	2	30
S. aureus ISP8	T5A+T3A	2	31
	T5A	1	5
	T3A	1	9
	T5A+T3A	1	14
	T5A+T3A	3	12
	T5B+T3A	3	14
S. aureus 8432	T5A+T3A	2	32
S. aureus 15564	T5A+T3A	2	33
S. aureus 6538	T5A+T3A	2	34
S. aureus 3A	T5A+T3A	2	35
	T5A+T3A	3	13
	T5B+T3A	3	15
S. aureus 12600	T5A+T3A	2	36
S. cohnii JL143	T5A+T3A	2	37
	T5A	1	4
	T3A	1	10
	T5A+T3A	1	15
S. cohnii CM 89	T5A+T3A	2	38
S. cohnii SS 521	T5A+T3A	2	39
S. cohnii 29974	T5A+T3A	2	40
Streptococcus (serotype)			
S. pyogenes (A) D471	T5A+T3A	2	1
	T5A+T3A	3	16
	T5B+T3A	3	18
S. pyogenes (A) Ti/195/2	T5A+T3A	2	2
	T5A+T3A	3	17
	T5B+T3A	3	19
S. pyogenes (A) 40 RS 15	T5A+T3A	2	3
S. pyogenes (A) 52 RS 15	T5A+T3A	2	4
S. pyogenes (A) 47 RS 15	T5A+T3A	2	5
S. pyogenes (A) 55 RS 15	T5A+T3A	2	6
S. pyogenes (G) T28/51/4	T5A+T3A	2	7
S. pyogenes (A?) K58 Hg	T5A+T3A	2	8
S. pyogenes (A) UAB 098	T5A+T3A	2	9
S. pyogenes (A) UAB 097	T5A+T3A	2	10
S. pyogenes (A) 14RP81	T5A+T3A	2	11
S. pyogenes (A) D471Rot	T5A+T3A	2	12
S. pyogenes (G) 1/E9	T5A+T3A	2	13
S. pyogenes (G) 040/011	T5A+T3A	2	14
S. mutans T8	T5A+T3A	2	15
S. pyogenes (B) 50316	T5A+T3A	2	16
S. pyogenes (A) UAB 092	T5A+T3A	2	17
Enterococcus			
E. faecalis, OGI X	T5A+T3A	2	18
E. faecalis, JH2-2	T5A+T3A	2	19
Maize (*Zea mays*) B73	T5A+T3A	3	1
	T5B+T3A	3	3
Maize Mo17	T5A+T3A	3	2
Human (*Homo sapiens*) 584	T5A+T3A	3	8
	T5B+T3A	3	10
Human 694	T5A+T3A	3	9
	T5B+T3A	3	11
Rice (*Oryza sativa*) G1	T5A+T3A	3	4
	T5B+T3A	3	6
Rice G2	T5A+T3A	3	5
	T5B+T3A	3	7

are in seven clusters of two to seven genes per cluster. The spacing of genes in these clusters is rather variable, ranging from 10 to 200 base pairs (e.g. 8). The genes are generally arranged in a head to tail fashion and are, at least in some cases, organized into operons. In *Bacillus subtilis* (9), *Photobacterium phosphoreum* (10) and *Spiroplasma* (11) the genes are more tightly clustered. For instance, in *B. subtilis* there are two main clusters consisting of 16 and 21 tRNA genes (9). One operon in *P. phosphoreum* has eight genes and five $tRNA^{Pro}$ pseudogenes, all in less than 1,500 base pairs (10). In the human nuclear genome there are estimated to be 1300 genes and a large number of tRNA pseudogenes (2, and references therein). At least seven clusters are on seven different chromosomes (2). However, in the few characterized cases in mammals, the tRNA genes are not in operons, being oriented in all possible directions within clusters.

Fungi, plants and animals have organelle genomes in addition to their nuclear genomes. Organelle genomes are much smaller than nuclear genomes but nevertheless encode tRNA genes for a more redundant genetic code (reviewed in 12). For example, the very small (circa 16,000 bp) animal mitochondrial DNA has 22 tRNA genes, some of which are closely spaced.

Chloroplast and mitochondrial DNA from plants are generally more complex than mitochondrial DNA in animals and often have more tRNA genes. For instance, the 121,024 base pair chloroplast genome of liverwort has 36 tRNA genes, a few of which are clustered (3). The yeast mitochondrial genome has at least 25 tRNA genes in 78,000 base pairs.

Experimental

Consensus primers were developed using tRNA sequences, of which over 500 are known. Given the variability in tRNA gene sequences between isoacceptors from different species, and the even greater difference between tRNAs for different amino acids, a substantial universal consensus does not exist. However, a reasonable match with a fraction of all tRNAs can be devised. With this in mind, it is possible to produce primers that have (a) at least a five base perfect match between the 3′ end of the primer and many tRNA genes and (b) extensive homology in the rest of the primer with a number of different tRNA genes from a wide variety of sources. For instance, the T5A primer (5′AGTCCGGTGCTCTAACCAACTGAG′) is derived from the complement of consensus sequences at the 5′ end of the tRNA genes, and thereby face out from the gene in the 5′ to 3′ direction. An example of the T5A consensus versus tRNA genes from *Bacillus* is shown:

- = match, * = missing base, IUB lettering convention.

Ala	-	-	-	-	-	-	C	-	-	-	G	*	-	-	-	-	-	G	-	-	T	-	C	-	-
Arg	-	-	-	-	-	A	-	-	-	-	A	-	-	-	-	-	-	G	T	T	T	-	-	-	-
Asn	-	-	-	-	-	-	*	-	-	-	*	*	-	-	-	-	-	T	A	T	C	-	G	-	-
Asp	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	T	G	-	-	T	-	C	-	-
Cys	-	-	C	A	-	-	*	-	-	-	*	-	-	A	G	-	-	-	G	A	-	-	T	-	-
Glu	-	-	C	A	-	-	*	C	-	-	*	-	-	A	G	-	-	-	A	-	-	-	-	-	-
Glu	T	G	C	A	-	-	*	C	-	-	-	-	-	A	G	A	-	-	-	-	-	-	C	C	-	-
Gly	-	T	-	y	-	-	*	-	-	-	*	-	-	R	-	A	-	m	y	-	A	s	C	y	-
His	-	-	G	A	-	-	*	-	-	-	-	-	-	A	C	-	-	-	-	-	A	-	-	T	-
Ile	-	-	-	-	-	-	-	y	-	-	-	-	-	-	-	-	-	r	s	A	C	-	S	-	-
Leu	T	y	G	G	-	A	-	-	-	-	C	*	-	-	-	C	R	-	r	-	W	-	-	-	-
Lys	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	T	-	T	-	-	-	-
Met	-	-	-	-	-	-	y	-	-	-	y	-	-	-	-	-	-	r	K	A	C	-	G	y	-
Met	-	A	G	-	-	-	-	-C	-	-	*	-	-	-	C	T	-	G	T	-	-	-	G	-	-
Phe	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	A	-	-	-	-	-	-
Pro	-	-	-	-	-	-	C	-T	-	-	*	-	-	-	-	-	-	-	-	A	T	-	G	T	-
Ser	Y	y	C	R	-	-	-	C-	-	-	Yy	G	-	A	G	-	G	G	-	X	C	-	x	-	-
Thr	-	-	-	-	R	y	w	-	-	-	*	-	-	-	-	-	-	-	m	Y	T	s	m	m	-
Trp	-	T	-	T	-	A	*	C	-	-	*	-	-	-	-	A	-	-	G	A	-	-	T	-	-
Tyr	-	-	G	A	-	-	*	-	-	-	C-	A	-	A	C	-	-	G	G	-	-	-	-	-	-
Val	-	-	-	-	-	-	C	-	-	-	G	*	-	-	-	-	-	-	T	-	T	-	C	-	-

Consensus:
5′ G C T C A G T T G G T T A G A G C A C C G G A C T 3′
Primer:
3′ G A G T C A A C C C A A T C T C G T G G C C T G A 5′

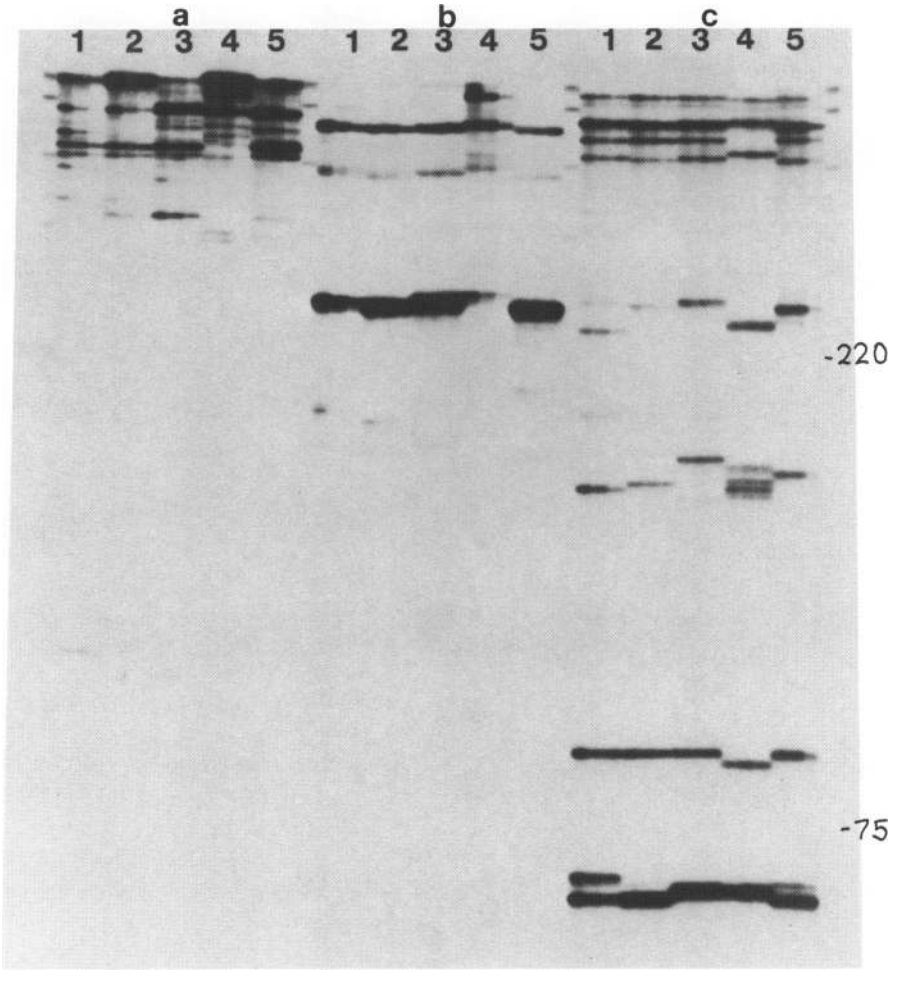

Figure 1. tDNA-PCR fingerprints of species within a genus: Strains from five species of staphylococci were chosen. PCR was performed using the primers T5A in group **a**, T3A in group **b**, or T5A plus T3A in group **c**, at 50°C with 80, 16 or 3.2 ng of template. Lane 1: *S. haemolyticus* CC 12J2. Lane 2: *S. hominis* 27844. Lane 3: *S. warneri* CPB10E2. Lane 4: *S. cohnii* JL 143. Lane 5: *S. aureus* ISP-8.

Eight out of the 21 genes shown have a five-base match at the 3′ end of the T5A primer and in these cases at least ten of the remaining 19 bases are perfectly matched.

T5B and T3A were fashioned in a similar manner. T5B, like T5A, faces 3′ at the 5′ end of the consensus, whereas, T3A is a consensus sequence at the 3′ end facing out of the genes in the 5′ to 3′ direction.

Given that there are one hundred or more divergent tRNA genes in a typical genome, there are likely to be many matches for the primers and these will vary from almost perfect to rather poor matches. PCR conditions were adapted to permit priming at imperfect matches. The primers T5A and T5B, and T3A were used alone or in pairs to amplify tRNA gene clusters in genomic DNA from various species of the bacteria *Staphylococcus* using amounts of template ranging from 100 to 3.2 ng, which are typical of PCR. The results, shown in Fig. 1, indicate that reproducible fingerprints can be obtained over a 25-fold range of template concentration at 50°C. Other experiments, not shown, indicated that the fingerprint did not vary when the low stringency annealing step was varied between 45°C and 50°C, which is suitable for partly mismatched primers. Suitable temperatures probably range from 40°C to 55°C.

There were a number of products generated for each genome in Fig. 1 whether T5A and T3A were used alone or together, indicating PCR initiated at a variety of places in the genome, as expected. With the exception of *S. cohnii*, which was already known to be the most divergent species within the genus (13, 14, 15), the tDNA-PCR patterns were very similar between the species, indicating that the tRNA gene clusters probably evolve relatively slowly. This is in contrast to AP-PCR (4) or total genome restriction digestion (16, and references therein), that give very different patterns when different species are compared.

A survey was performed on forty strains of bacteria, representing many strains from five species of *Staphylococcus*, four species of *Streptococcus* and a species of *Enterococcus*. The organization of the tRNA genes in these species has not been described, but they are presumably similar to those of other related bacteria, such as *Bacillus*. Fig. 2 shows that within a species there was generally no variation in the tDNA-PCR pattern. There were only two exceptions. A *Streptococcus pyogenes* strain K58Hg that was designated serotype A (lane 8) gave a pattern identical to serotype b (lanes 7, 13 and 14). Interestingly, AP-PCR experiments (manuscript in prep.) group this strain with serotype b and not serotype a. It is likely that K58Hg is in fact a serotype b. The other exception was a strain of *S. haemolyticus* (lane 20) that was obtained from the ATCC. We do not as yet have an explanation for this exception. However, our preliminary AP-PCR data indicates that *S.*

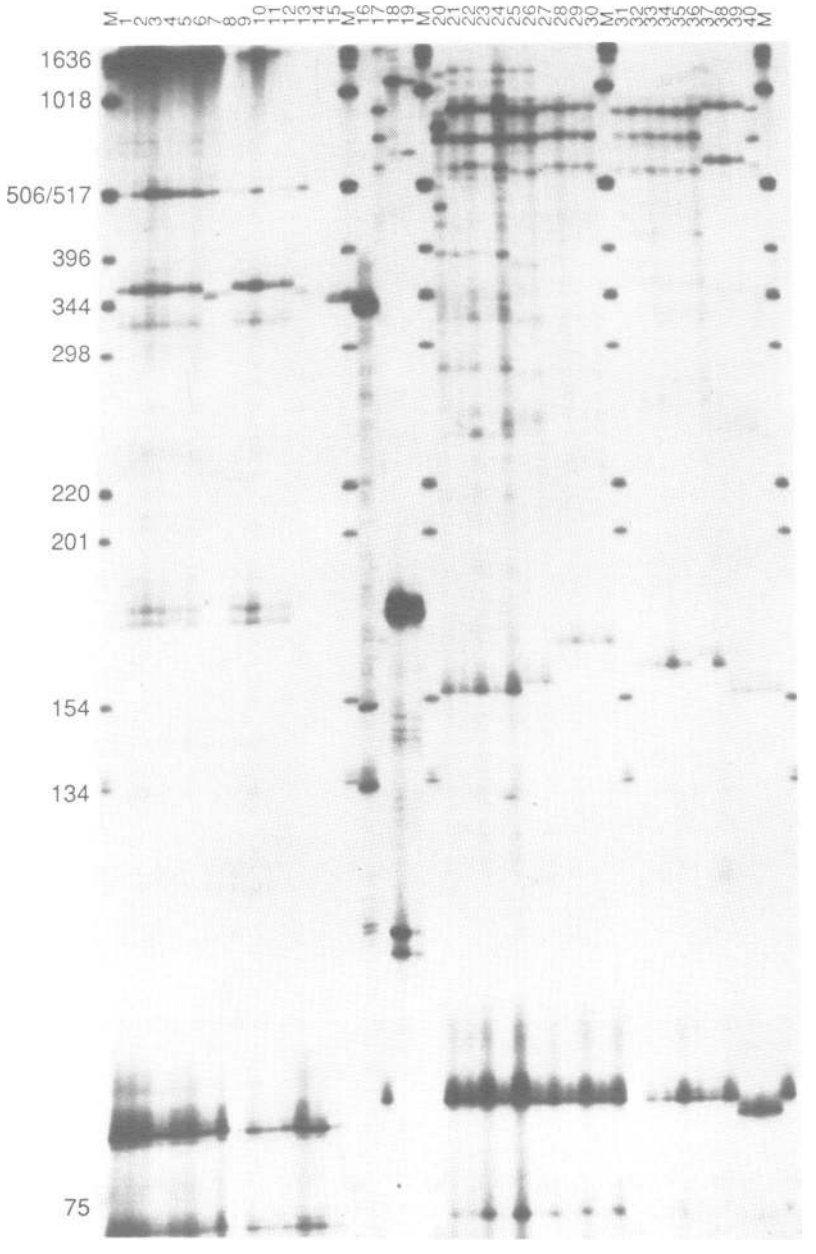

Figure 2. Fingerprints of *Streptococcus* and *Staphylococcus*: A total of 40 strains of bacteria from three genera were chosen. PCR was performed using the primers T5A plus T3A at 50°C with 100 ng of template. The templates in lanes 1 to 17 are *Streptococcus* DNAs. Lanes 18 and 19 are *Enterococcus*. Lanes 20 to 40 are *Staphylococcus*. See Table 1 for the strains used in each lane.

haemolyticus consists of at least two groups of strains that are rather divergent and may in fact be different species (4, and unpub. results).

tDNA-PCR should work for a wide variety of species because tRNA genes are highly conserved, are abundant, and are generally arranged in clusters. Fig. 3 shows tDNA-PCR reactions on the genomes of species from three kingdoms. In addition to the bacterial genomes, there were fingerprints generated for the maize, rice and human genomes with at least one of the two pairs of primers we tested. For example, the rice fingerprints are identical between strains, but the T5A/T3A pair of primers (lanes 4 and 5) gives a completely different pattern than the T5B/T3A pair (lanes 6 and 7).

In the case of eukaryotes, a typical consensus tRNA primer will prime both nuclear and organelle (mitochondrial and chloroplast) tRNA genes. Plant mitochondrial and chloroplast genomes do not seem to have a high rate of point mutation, while the evolution of animal mitochondrial genomes is fast (17 and references therein). In this latter case, it can be expected that the resulting tDNA-PCR products will vary over a shorter evolutionary time than nuclear products.

Organelle genomes, despite their small size, can contribute up to half of the DNA in the cell, due to their high copy number. Nevertheless, for tDNA-PCR the best matches with the primer are probably more important than copy number and these best matches will generally be to nuclear genes because the number of different nuclear tRNA gene sequences greatly exceeds the number of different organelle tRNA gene sequences.

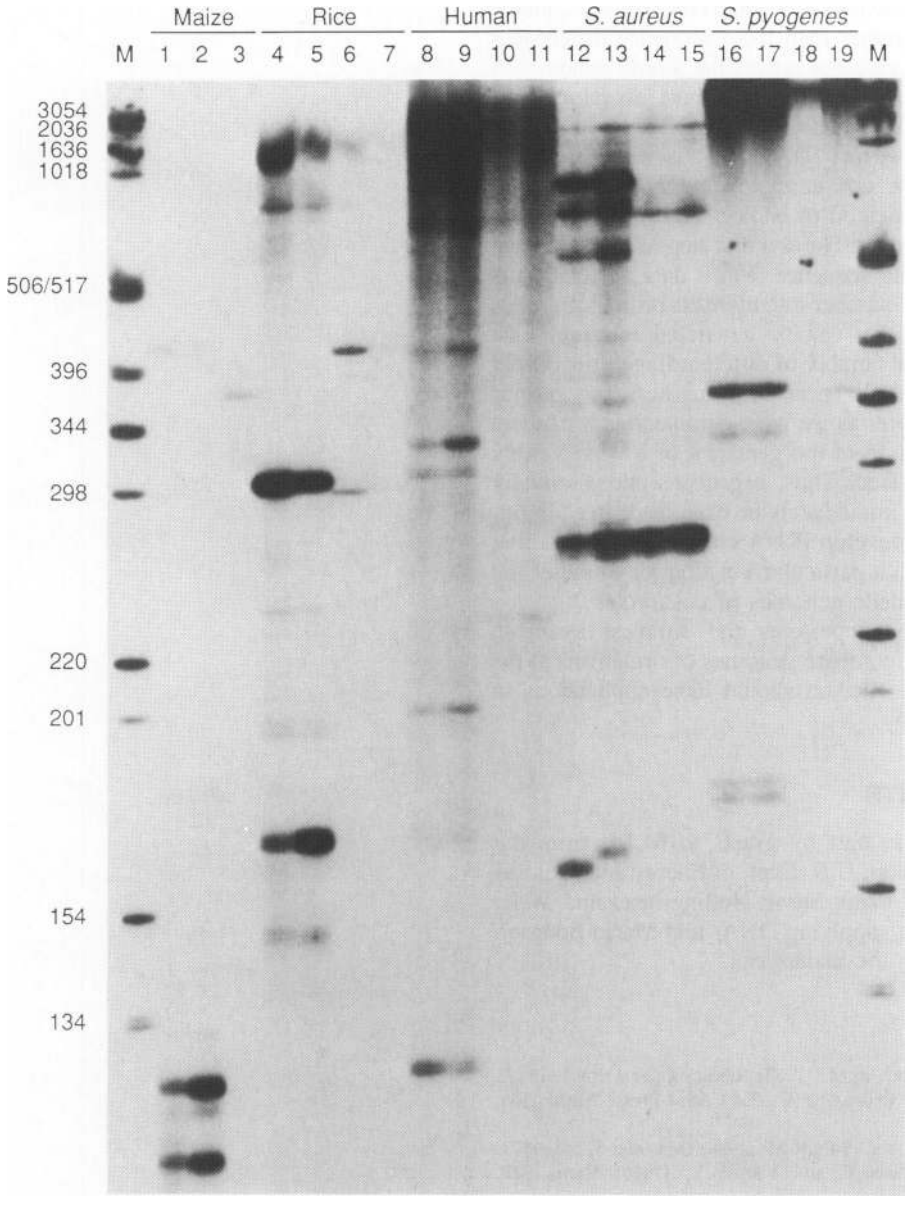

Figure 3. tDNA-PCR works on genomes from three kingdoms: The reaction was performed using 50 ng of template under the standard PCR conditions. The low temperature annealing step was 50°C. Lanes 1 to 9 used the primers T5A and T3A. Lanes 10 to 19 used T5B and T3A. See Table 1 for the strains used in each lane.

We are not sure that the tDNA-PCR products shown are, in fact, from tRNA genes nor, in the case of eukaryotes, if they are from the nuclear or organelle genomes. However, the patterns were identical between divergent individuals within each species. Regardless of the origin of the pattern, tDNA-PCR appears to be a useful method of species classification. The consistency of the patterns between bacterial species is strong evidence that the fingerprints are not generated by arbitrarily primed-PCR. Such AP- PCR fingerprints are not conserved between species (4, and unpublished data).

DISCUSSION

The experiments presented indicate that consensus tRNA gene primers that amplify the region between tRNA genes can be used to generate PCR fingerprints that are generally invariant between strains of the same species and are often substantially conserved between related species. This property may make the method applicable to the identification of organisms by a genome based method that is independent of other criteria such as morphology. The ease with which the method can be performed, independent of the genome size, sequence, concentration of genomic DNA, or the cycling parameters, means that it may be the method of choice when examining a large number of different stains for which a rapid and convenient method for categorization is desired.

While there are many ways to determine species and genus, tDNA-PCR would be a simple and fast method that complements those that already exist and has the virtue that the polymorphisms measured are not themselves likely to be selected. They are more likely to be near neutral than other characters such as nutrition and perhaps less likely to result from convergent evolution, which is a drawback of classification by morphological criteria. It may

also be an advantage that the data is collected for regions scattered throughout the genome rather than from sequence differences in a single location that may not reflect the whole genome. A single tDNA-PCR fingerprint will generally have less information than comparing the DNA sequence of a specific region in each organism (e.g. 18). However, tDNA-PCR is less technically demanding and less time consuming than DNA sequencing. tDNA-PCR could be a method of choice when large numbers of individuals are to be screened or as a first step when identifying species based on genomic sequence. Since data acquisition is trivial in tDNA-PCR, the number of consensus primers and thus the number of patterns that can be generated is large. One imagines that any required number of different fingerprints could be generated to provide the necessary markers for species classification. Furthermore, as we have demonstrated, primers that produce fingerprints from the genomes of a wide variety of organisms can be devised. Thus, organisms one essentially knows nothing about can immediately be examined. In addition, it should be possible to develop tRNA consensus primers that are targeted preferably to a particular kingdom or to either the nuclear genome or organelle genomes of eukaryotes.

The method presented represents the simplest available universal way to reliably compare genomes of organisms at the species/genus level. The method should have applications in ecology and epidemiology.

ACKNOWLEDGEMENTS

This work was funded in part by grants to M.M. from the National Science Foundation, U.S. Dept. of Energy and National Institutes of Health. We thank Susan Hollingshead and W.E. Kloos and R.M.Lipps for supplying DNAs and Mario Bourdon for helpful comments on the manuscript.

REFERENCES

1. Jinks-Robertson,S. and Nomura,M. (1987) Ribosomes and tRNA, in *'E. coli* and *S. typhimurium'*, Neidhardt,F.C. (Ed.) ASM Press, Washington, pp1358–1385.
2. McBride,O.W., Pirtle,I.L. and Pirtle,R.M. (1989) *Genomics* **5**, 561–573.
3. Ohyama,K., Kohchi,T., Sano,T. and Yamada,Y. (1986) *Nature* **322**, 572–574.
4. Welsh,J. and McClelland,M. (1990) *Nucl. Acids Res.* **18**, 7213–7218.
5. Williams,J.G.K., Kubelik,A.R., Livak,K.J., Rafalski,J.A. and Tingey,S.V. (1990) Nucl. Acids Res. **18** (in press).
6. Ledbetter,S.A., Nelson,D.L., Warren,S.T. and Ledbetter,D.H. (1990) *Genomics* **6**, 475–481.
7. Nelson,D.L., Ledbetter,S.A., Corbo,L., Victoria,M.F., Ramirez-Solis,R., Webster,T.D., Ledbetter,D.H. and Caskey,C.T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6686–6690.
8. Rossi,J.J. and Lancy,A. (1979) *Cell* **16**, 523–534.
9. Vold,B.S. (1985) *Microbiol. Rev.* **49**, 71–80.
10. Giroux,S., Beaudet,J. and Cedergen,R. (1988) *J. Bacteriol.* **170**, 5601–5606.
11. Rogers,M.J., Steinmetz,A.A. and Walker,R.T. (1984) *Israel J. Med. Sci.* **20**, 768–772.
12. Fox,T.D. (1987) *Annu. Rev. Genet.* **21**, 67–91.
13. Kloos,W.E. (1980) *Annu. Rev. Microbiol.* **34**, 559–592.
14. Kloos,W.E. and Wolfshohl,J.F. (1979) *Curr. Microbiol.* **3**, 167–172.
15. Kloos,W.E. and Wolfshohl,J.F. (1983) *Curr. Microbiol.* **8**, 115–121.
16. Cinco,M., Banfi,E., Balanzin,D., Caccio,S., Graziosi,G. and Fattorini,P. (1989) *FEMS Micobiol. Immunol.* **47**, 511–514.
17. Palmer,J.D. (1985) In 'Molecular Evolutionary Genetics'. MacIntyre,R.J. Ed., Plenum Press, New York, pp 131–240.
18. Meyer,A., Kocher,T.D., Basasibwaki,P. and Wilson,A.C. (1990) *Nature* **347**, 550–553.