Published in final edited form as:

J Microbiol Methods. 2014 August ; 103: 101–103. doi:10.1016/j.mimet.2014.05.018.

Selection of Unique *Escherichia coli* Clones by Random Amplified Polymorphic DNA (RAPD): Evaluation by Whole Genome Sequencing

Karen L. Nielsen^{1,2,*}, Paul A. Godfrey³, Marc Stegger², Paal S. Andersen², Michael Feldgarden³, and Niels Frimodt-Møller¹

¹Department of Clinical Microbiology, Hvidovre Hospital, 2640 Hvidovre, Denmark

²Department of microbiology and Infection Control, Statens Serum Institut, 2300 Copenhagen S, Denmark

³Genome Sequencing and Analysis Program, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Identifying and characterizing clonal diversity is important when analysing fecal flora. We evaluated random amplified polymorphic DNA (RAPD) PCR, applied for selection of *Escherichia coli* isolates, by whole genome sequencing. RAPD was fast, and reproducible as screening method for selection of distinct *E. coli* clones in fecal swabs.

Keywords

molecular typing; whole-genome sequencing; fecal flora; *Escherichia coli*; urinary tract infection

Urinary tract infection (UTI) caused by *Escherichia coli* is one of the most common bacterial infections observed in clinical practice, and fecal flora has been shown to constitute a dominant reservoir (1–3). Whereas both high and low prevalent clones have been shown to cause UTI (2–5), identification of especially low prevalent clones in the fecal flora mandate screening of several isolates in rectal swabs to properly characterise diversity and underlying route of infection (3, 5, 6).

A large number of typing methods have been applied to characterise *E. coli* genotypes including phylogroup typing into A, B1, B2, D and non-typeables (NT) (multiplex PCR) (7, 8), pulsed-field gel electrophoresis (PFGE) (9), multilocus sequence typing (MLST) (10), and random amplified polymorphic DNA (RAPD) PCR (11). Some of these methods are

© 2014 Elsevier B.V. All rights reserved.

*Corresponding Author: Karen Leth Nielsen, Serum Institut, Artillerivej 5 (37/131), 2300 Copenhagen S, Tel: +45 32683650, KTE@ssi.dk.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

laborious (MLST, PFGE), expensive (MLST), or generally not sensitive enough to provide clone specific fingerprints (phylogrouping, MLST). The advances in whole genome sequencing (WGS) technology have provided a tool that allows highly detailed phylogenetic typing (12). However, sample preparations are still laborious and WGS is still too expensive for most laboratories to run on all available isolates. When approaching mixed samples, such as the environment found in fecal flora, the cost of WGS typically warrants a pre-selection of distinctive bacterial clones that efficiently reveal the overall population structure.

Here, we propose RAPD typing as a fast, reproducible, high-resolution and inexpensive method to identify and select distinct *E. coli* clones prior to WGS or other high-resolution typing schemes.

Initial screening of six short primers for RAPD typing (1254, 1247, 1290, 1283, 1253 and M13 (13–17)) showed that 1247 (AAGAGCCCGT) and 1283 (GCGATCCCCA) (14, 15) provided the highest resolution, i.e. number of bands on fecal *E. coli*; and two PCRs were applied to each sample, each containing one of the two primers, as described by Nielsen *et al* 2014. Briefly, Multiplex PCR Master Mix (Qiagen) was used and each 25µL reaction contained 2µM of one primer and 2.5µL of template DNA (crude lysates). The following cycling conditions were used for the 1247 and 1283 PCR, respectively: 95°C for 15 min, 35 cycles of 94°C for 1 min, 38/36°C for 1 min and 72°C for 2 min, with a final 10 min elongation step at 72°C. All isolates from each individual were analysed concurrently within same PCR run and gel (2% E-gel, Invitrogen). Reproducibility of the assay was investigated by running 11 isolates (with very diverse RAPD fingerprint) from 11 unrelated fecal samples in three unrelated analyses using both the new and identical DNA lysates.

A total of 97 rectal swabs from women aged 19–53 were plated on individual plates and 20 *E. coli* colonies were isolated whenever possible (five swabs contained no *E. coli*, three samples contained only two, four and eighteen colonies, respectively). The RAPD assay was applied to a total of 1,804 *E. coli*. In 41 swabs, all 20 isolates exhibited no band differences. *E. coli* isolates from the remaining 51 rectal swabs with apparent differences in the RAPD fingerprint (n=127) and one representative from each swab with no band differences (n=41) were subsequently whole-genome sequenced (N=168) (HiSeq 2000, Illumina). Relationship between the isolates were analysed in a phylogeny of 242 genomes in total, including other available genome sequences of *E. coli*, *E. fergusonii* and *E. albertii* (N=242). Phylogenetic reconstruction was performed using FastTree (18) on 1776 identified single copy core genes, as identified by reciprocal best hit BLAST and single linkage clustering. The phylogenetic tree was used to evaluate RAPD as an initial screening method for relatedness of unique colonies in mixed samples such as rectal swabs. Identical and different isolates were evaluated based on a criterion of >99.95% similarity based on WGS data.

The RAPD assays showed very high degree of reproducibility, as the same amplification patterns were found for each of the 11 isolates, regardless of whether a new or identical DNA crude lysates were applied. Each RAPD assay created multiple bands as illustrated in Figure 1. Of the 127 isolates with differences in RAPD, 10 isolates exhibited one band difference, but were identical in the phylogenetic analysis (Table 1). Nine isolates with 2 bands difference were identical to another isolate in the sample according to WGS (Table 1).

Isolates differing by 2 bands had $96.67\% \pm 2.62$ identity (mean $\pm$ SD) on average compared to isolates differing by 0–1 bands, which were found to have $99.99\% \pm 0.015$ similarity ($P < 0.0001$). Combined, these results demonstrate that 2 band difference in RAPD is a highly useful criterion for selection of distinct *E. coli* clones in a diverse strain collection. Only 7.7% of the isolates ($n=9$) were misclassified and assumed to be due to contamination of the DNA sample.

The presented dual-PCR RAPD method has demonstrated a high level of resolution that allows for identification of distinct *E. coli* clones. Historically, RAPD assays have exhibited low reproducibility (19). However, in the presented approach, we applied commercially available premixed PCR reagents and precast gels in order to strictly define conditions and minimize number of processing steps, which has previously proved to influence the reliability of the method (11, 19). Using these simple guidelines we were able to create a highly reproducible assay.

We only sequenced one isolate from each RAPD with identical pattern (zero band difference) according to the RAPD assays; however, as Table 1 indicates, 2 band difference was an appropriate selection criterion for identification of unique clones. We therefore find strong support to extrapolate that isolates with identical RAPD patterns share near-identical core genomes as defined by WGS analysis.

In conclusion, RAPD provides a specific and reproducible assay for highly discriminatory screening of *E. coli* sub-clones. Thus, it presents an affordable and rapid screening approach for larger *E. coli* strain collections in order to perform a substantiated pre-selection before applying more complex, laborious, and/or expensive methods such as WGS.

Acknowledgments

This work was a part of PAR, an EU FP7-Health-2009-Single-Stage project (grant agreement #241476). This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No.:HHSN272200900018C.

References

1. Foxman B. The epidemiology of urinary tract infection. *Nat Rev Urol.* 2010; 7:653–60. [PubMed: 21139641]
2. Yamamoto S, Tsukamoto T, Terai a, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. *J Urol.* 1997; 157:1127–9. [PubMed: 9072556]
3. Nielsen KL, Dynesen P, Larsen P, Frimodt-Møller N. Faecal *Escherichia coli* from patients with *E. coli* urinary tract infection and healthy controls who have never had a urinary tract infection. *J Med Microbiol.* 2014; 63:582–9. [PubMed: 24464694]
4. Moreno E, Andreu A, Pigrau C, Kuskowski Ma, Johnson JR, Prats G. Relationship between *Escherichia coli* strains causing acute cystitis in women and the fecal *E. coli* population of the host. *J Clin Microbiol.* 2008; 46:2529–34. [PubMed: 18495863]
5. Schlager TA, Hendley JO, Bell AL, Whittam TS. Clonal Diversity of *Escherichia coli* Colonizing Stools and Urinary Tracts of Young Girls. *Infect Immun.* 2002; 70:1225–1229. [PubMed: 11854204]

6. Bailey JK, Pinyon JL, Anantham S, Hall RM. Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. *J Med Microbiol.* 2010; 59:1331–9. [PubMed: 20671087]
7. Clermont O, Bonacorsi S, Bingen E. Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Appl Environ Microbiol.* 2000; 66:4555–4558. [PubMed: 11010916]
8. Gordon DM, Clermont O, Tolley H, Denamur E. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environ Microbiol.* 2008; 10:2484–96. [PubMed: 18518895]
9. Barrett TJ, Lior H, Green JH, Khakhria R, Wells JG, Bell BP, Greene KD, Lewis J, Griffin PM. Laboratory Investigation of a Multistate Food-Borne Outbreak of *Escherichia coli* O157: H7 by Using Pulsed-Field Gel Electrophoresis and Phage Typing. *J Clin Microbiol.* 1994; 32:3013–3017. [PubMed: 7883892]
10. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, 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 U S A.* 1998; 95:3140–3145. [PubMed: 9501229]
11. Berg DE, Akopyants NS, Kersulyte D. Fingerprinting microbial genomes using the RAPD or AP-PCR method. *Methods Mol Cell Biol.* 1994; 5:13–24.
12. Leimbach A, Hacker J, Dobrindt U. *E. coli* as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. 2013:3–32.
13. Aslam M, Greer GG, Nattress FM, Gill CO, McMullen LM. Genetic diversity of *Escherichia coli* recovered from the oral cavity of beef cattle and their relatedness to faecal *E. coli*. *Lett Appl Microbiol.* 2004; 39:523–7. [PubMed: 15548305]
14. Venieri D, Vantarakis A, Komninou G, Papapetropoulou M. Differentiation of faecal *Escherichia coli* from human and animal sources by random amplified polymorphic DNA-PCR (RAPD-PCR). *Water Sci Technol.* 2004; 50:193–198. [PubMed: 15318508]
15. Madico G, Akopyants NS, Berg DE. Arbitrarily primed PCR DNA fingerprinting of *Escherichia coli* O157: H7 strains by using templates from boiled cultures. These include: Arbitrarily Primed PCR DNA Fingerprinting of *Escherichia coli* O157: H7 Strains by Using Templates from Boiled Cultur. *J Clin Microbiol.* 1995; 33:1534–1536. [PubMed: 7650181]
16. Regua-Mangia AH, Andrade JRC, Gonzalez AGM, Zahner V, Cerqueira AMF, Teixeira LM. Genetic relatedness of a non-motile variant O157 enteropathogenic *Escherichia coli* (EPEC) strain and *E. coli* strains belonging to pathogenic related groups. *Microbiol Res.* 2008; 163:225–33. [PubMed: 16815695]
17. Vogel L, van Oorschot E, Maas HM, Minderhoud B, Dijkshoorn L. Epidemiologic typing of *Escherichia coli* using RAPD analysis, ribotyping and serotyping. *Clin Microbiol Infect.* 2000; 6:82–7. [PubMed: 11168077]
18. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009; 26:1641–50. [PubMed: 19377059]
19. Meunier JR, Grimont PA. Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res Microbiol.* 1993; 144:373–9. [PubMed: 8248630]

Figure 1. RAPD typing of two fecal swabs ((a)/(b) and (c)/(d), respectively). (a) and (c): Primer 1247, (b) and (d): Primer 1283. M: 1kb marker, N: Negative control, P: Positive control ATCC 25922, lane 4–23: isolates from one participant with 2 and 3 unique *E. coli* clones in flora, respectively. -: blank.

Table 1

Comparison of RAPD and WGS based analyses from all divergent RAPD patterns obtained from 20 individual *E. coli* colonies from 92 unrelated rectal swabs.

Number of RAPD band differences	Number of isolates N=127, n (%)	Identical* by WGS n (%)	Unrelated* by WGS n (%)
1	10 (7.9)	10 (100)	0 (0)
2	117 (92)	9 (7.7)	108 (92)

* Identical was defined as >99.95% similarity in phylogenetics in single copy core genes.