International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:10, No:10, 2016

From Primer Generation to Chromosome Identification: A Primer Generation Genotyping Method for Bacterial Identification and Typing

Wisam H. Benamer, Ehab A. Elfallah, Mohamed A. Elshaari, Farag A. Elshaari

***Abstract*—**A challenge for laboratories is to provide bacterial identification and antibiotic sensitivity results within a short time. Hence, advancement in the required technology is desirable to improve timing, accuracy and quality. Even with the current advances in methods used for both phenotypic and genotypic identification of bacteria the need is there to develop method(s) that enhance the outcome of bacteriology laboratories in accuracy and time. The hypothesis introduced here is based on the assumption that the chromosome of any bacteria contains unique sequences that can be used for its identification and typing. The outcome of a pilot study designed to test this hypothesis is reported in this manuscript. Methods: The complete chromosome sequences of several bacterial species were downloaded to use as search targets for unique sequences. Visual basic and SQL server (2014) were used to generate a complete set of 18-base long primers, a process started with reverse translation of randomly chosen 6 amino acids to limit the number of the generated primers. In addition, the software used to scan the downloaded chromosomes using the generated primers for similarities was designed, and the resulting hits were classified according to the number of similar chromosomal sequences, i.e., unique or otherwise. Results: All primers that had identical/similar sequences in the selected genome sequence(s) were classified according to the number of hits in the chromosomes search. Those that were identical to a single site on a single bacterial chromosome were referred to as unique. On the other hand, most generated primers sequences were identical to multiple sites on a single or multiple chromosomes. Following scanning, the generated primers were classified based on ability to differentiate between medically important bacterial and the initial results looks promising. Conclusion: A simple strategy that started by generating primers was introduced; the primers were used to screen bacterial genomes for match. Primer(s) that were uniquely identical to specific DNA sequence on a specific bacterial chromosome were selected. The identified unique sequence can be used in different molecular diagnostic techniques, possibly to identify bacteria. In addition, a single primer that can identify multiple sites in a single chromosome can be exploited for region or genome identification. Although genomes sequences draft of isolates of organism DNA enable high throughput primer design using alignment strategy, and this enhances diagnostic performance in comparison to traditional molecular assays. In this method the generated primers can be used to identify an organism before the draft sequence is completed. In addition, the generated primers can be used to build a bank for easy access of the primers that can be used to identify bacteria.

***Keywords*—**Bacteria chromosome, bacterial identification, sequence, primer generation.

I. Introduction

ACCURACY and speed are the two principles required when testing for bacterial identification, which requires quality [1]. In addition, typing bacteria at the strain level is sometimes essential, since it exposes some features such as virulence, antibiotic resistance [2]. Hence, development in analytical technologies that reduce the time for obtaining results yet keep accuracy and quality are needed. Basically two types of methods are used in the identification of bacteria. The older method based on phenotyping in which bacterial identification is based on colony morphology, biochemical testing, serology, pathogenicity and antibiotic susceptibility, processes that could take days or weeks [2]-[4]. The second and more recent is based on genotyping in which several genetic methods have been developed. These methods have been used more frequently in bacterial identification and subtyping due to the fact that the resulting genetic profile of any bacteria by specific method can be as unique as fingerprint [2].

Currently the available bacterial genotyping methods can be mainly broken-down into three categories [5], [6].

- DNA banding pattern
- DNA sequencing
- DNA hybridization

Other techniques were introduced for identification of pathogens; these include PNA-FISH [7], in addition to PCR-ESI/MS and MALDI-TOF/MS demonstrating high accuracy of both methods for genus and species identification [8], [9].

The purpose of this project is to address the question whether the fact that for each bacteria having unique DNA sequences within its genome could be used for its identification by firstly identifying these sequences.

This paper describes the trial that exploits this fact, and the steps taken to design and trial the methods to obtain the desired results. The steps started by designing and using software that practically have the potential to identify and possibly subtype bacteria. The method starts by primer generation followed by genome scan for similarity, and

W. H. Benamer is with the Department of Information System, Faculty of Information Technology, University of Benghazi, Libya (e-mail: wisam.benamer@uob.edu.ly).

E. A. Elfallah is with the Health Informatics Department, Faculty of Information Technology, Libyan International Medical University (e-mail: ehab.elfallah@limu.edu.ly).

M. A. Elshaari is with Department of Computer Science, Faculty of Information Technology, University of Benghazi (email: mohamed.shaari@gmail.com).

F. A. Elshaari is with the Department of Biochemistry, Faculty of Medicine, University of Benghazi and Biotechnology research center (Corresponding author; e-mail: elshaari@yahoo.com, info.biomedical@uob.edu.ly).

concluded by classifying primers according to their uniqueness in having sequence matches in bacterial genomes.

In the microbiology laboratory setting traditionally used culture based methods take long time to yield results. Therefore, producing results in a matter of hours using molecular techniques holds the potential for vast improvements in laboratory services. Such techniques will yield faster and reliable results if linked to bacterial primer database.

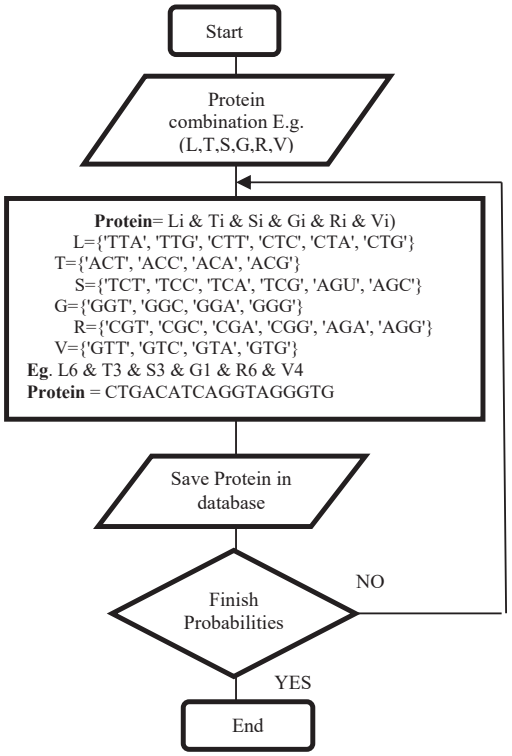

Fig. 1 Primers Generation Algorithm

II. Methods

The methods section was divided into three phases. Selection of amino acids, reverse translation and organizing the produced codons using amino acids one letter abbreviation and one digit for each codon was designated as phase I. Phase II was for bacterial genome sequence download and organization in files. Phase III was set for primer scan for the data collected and organized in phase two, and all used primers were classified according to their degree of uniqueness.

The concept of primers generation, introduced by our research team [10], was used in the phase II of the current project. Different software were developed using different tools for reverse translation, primer generation and sequence scan.

A total of six amino acids were selected, and reverse translated for all possible codons. Each amino acid is known to be represented by one to six codons. A codon is a sequence of three adjacent specific nucleotides represented as (A, C, T or G). Moreover, the selected small peptide is represented by the letters L, T, S, G, R, V. To represent each amino acid and its possible codon combinations a number is added to the letter to represent the amino acid and one of its specific codons. For instance, V (VAL) is represented by four codons (V1='GTT', V2='GTC', V3='GTA', V4='GTG'). On the other hand, L has 6 different codons, hence L1...L6. To generate the primers, the software picks V1 codon and one codon from each of the other 5 amino acids and picks a different codon each cycle (Fig. 1).

Special software has been developed to generate mixes of possible amino acid codon probabilities for the selected small peptide mentioned above. For implementing the algorithms, Visual Basic 6 and SQL Server 2014 were used to develop the software.

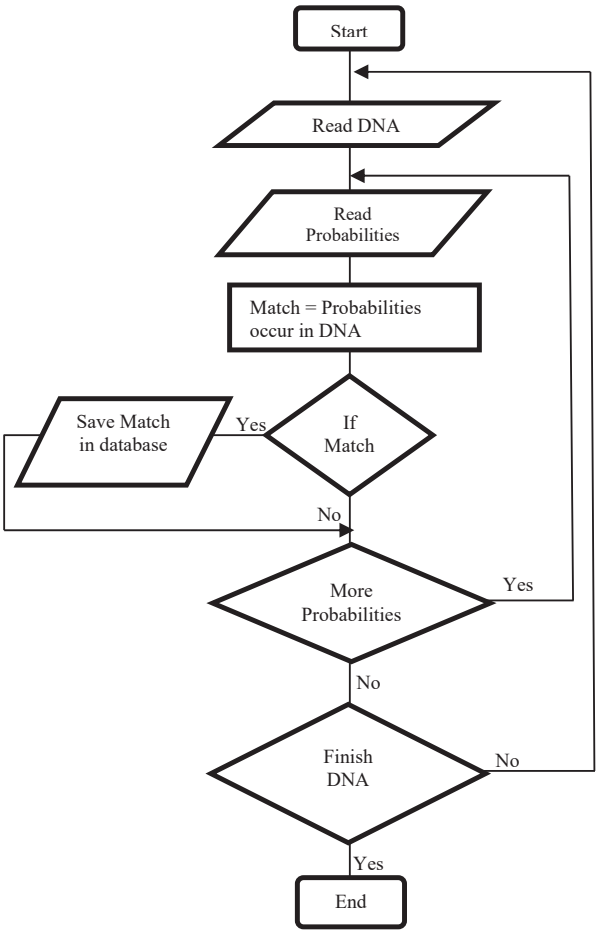

Fig. 2 Scan and Matching Algorithm

To test the hypothesis that bacterial genomes have unique sequences that enable us to identify them, software was developed to scan these genomes for primer matches. To test the developed software selected bacterial genomes were downloaded from the data available at National Center for Biotechnology Information [11]. Even though the current study can be conclusive, but it is a pilot study intended to test if the current method gives us the capability to differentiate

between different bacteria. All the chosen genomes represent classified medically important bacteria.

The six amino acids used represented by the abbreviation L, T, S, G, R, V, and having 13824 case of codon probabilities. The probabilities generated by the software were saved in the database.

In the matching stage each bacterial DNA collected in the DNA BANK were searched by the probabilities generated from the selected codon sequence. Furthermore, the matching probabilities and the DNA were saved in the database (Fig. 2).

III. Results

Both the software designed to generate the primers and to scan the genomes worked perfectly. The former enabled us to generate all possible codon combinations, whereas the latter gave us the ability to scan all downloaded sequences for primer similarity.

The majority of primers generated in the current study had identical sequences in one or more bacterial genomes. The selected primers and the identified genomes are shown (Table I).

TABLE I
Matching Results

Amino Acid codon key	Product of Reverse Translation	Species of Bacteria Having Primer Sequence Match
L6T2S2G3R2V2*	CTGACCTCCGGACGCGTC	*Pseudomonas aeruginosa*
L1T1S2G1R3V4	TTAACTTCCGGTCGAGTG	*Coxiella burnetii*
L4T1S4G1R4V1	CTCACTTCGGGTCGGGTT	*Aeromonas hydrophila* subsp. *hydrophila*
L4T2S4G2R4V2	CTCACCTCGGGCCGGGTC	*Aeromonas hydrophila* subsp. *hydrophila*
L6T2S6G2R6V2	CTGACCAGCGGCAGGGTC	*Aeromonas hydrophila* subsp. *hydrophila*
L4T2S6G3R3V1	CTCACCAGCGGACGAGTT	*Acinetobacter baumannii*
L5T3S6G1R4V1	CTAACAAGCGGTCGGGTT	*Actinobacillus pleuropneumoniae*
L6T4S3G1R4V4	CTGACGTCAGGTCGGGTG	*Citrobacter koseri*
L1T1S4G1R6V3	TTAACTTCGGGTAGGGTA	*Leptospira interrogans* serovar Lai str.
L3T1S3G2R5V1	CTTACTTCAGGCAGAGTT	*Coprococcus eutactus*

Primers' sequences were generated by reverse translation using probability mixes of amino acid codons. The primers were used to scan the downloaded medically important bacterial genomes for matching sequences. The examples of the primers shown are all unique, i.e., each primer sequence was found once in the bacterial genome shown, and was not found in any other Genome of the selected bacteria.

*The letters represent standard one letter code of selected amino acids and the numbers represents the selected codons used in standard genetic code

IV. Discussion

In molecular Identification of bacteria evidence have been produced to support the hypothesis that few general primer pairs are needed to identify several bacterial types [12]. It is even more demanding to identify, type and subtype bacteria. Earlier several molecular techniques were used to identify and classify eubacteria. They ranged from GC content [12], plasmid profiling to more recent technologies such as microarrays [13], [14].

The method described here has the potential to identify bacteria, type and subtype them. At the current phase primers were generated and tested for similarity in one or several types of bacteria. Our plan is to build a bank that can provide a set of primers that can identify and classify bacteria to a specific genus followed by or combined with the use of primer pairs that can lead us to the species and subspecies.

This project was designed to be executed in stages. The first stage, introduced in this paper was to digitally using software to test whether each bacterial genome has at least one short sequence unique in at least one single nucleotide different from all other bacteria including types of the same species. This was proven using the selected medically important bacteria in this study. In the second stage, all published bacterial genome sequences will be tested to find general primers, i.e. those that can identify a group of bacteria, and unique primers those that can identify one type of bacteria. The results of this stage will be used to build the bacteria identification primer bank. In the third stage, selected primers will be used in a suitable practical molecular biology technique especially those that are PCR based.

One of the advantages of our proposed method is that it requires minimum post reaction manipulation if any, and on the other hand, most of the software design and use is required during the first phase. In comparison in a review written by Muldrew [15], it was speculated that new sophisticated methods such as microarrays (MA) and whole genome sequencing will produce markedly growing information eventually requiring software to analyze to reach conclusive results. However, with this currently proposed method it is our understanding that the approach is realistic and the analysis of the results will be simple.

V. Conclusion

The presented technologies in this article show promise in prompt identification of microorganisms. Reaching this conclusion is based on the fact that so far we have primers that have the potential to identify several bacteria, and we have those that are highly specific for a single type. Once practically trialed, the idea could have other applications all to do with identifying a specific region of DNA sequence or a chromosome.

References

[1] Wolk DM, Kaleta EJ, and Wysocki VH. PCR-Electrospray Ionization Mass Spectrometry, The Potential to Change Infectious Disease Diagnostics in Clinical and Public Health Laboratories. Journal of Molecular Diagnostics 2012;14(4), 295-304

[2] Li W, Raoult D, Fournier P. Bacterial strain typing in the genomic era. FEMS Microbiol Rev 2009;33:892-916.

[3] Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Infect Control Hosp Epidemiol 1997;18:426-439.

[4] Jackson GW, McNichols RJ, Fox GE, Wilson RC. Bacterial genotyping by 16S rRNA mass cataloging. BMC Bioinformatics 2006;7:321 doi:10.1 186/147-2105/7/321.

[5] Taha MK, Olcen P. Molecular genetic methods in diagnosis and direct characterization of acute bacterial central nervous system infections. APMIS 2004;112:753-770.

[6] Versalovic J, Lupski JR, Molecular detection and genotyping of pathogens: more accurate and rapid answers. Trends Microbiol 2002;10:S15-S21.

[7] Forrest GN, Mehta S, Weekes E, Lincalis DP, et al. Impact of rapid in situ hybridization testing on coagulase-negative staphylococci positive blood cultures. J Antimicrob Chemother 2006;58:154–8
[8] Kaleta EJ, Clark AE, Cherkaoui A, Wysocki VH, et al. Comparative analysis of pcr-electrospray ionization/mass spectrometry (ms) and maldi-tof/ms for the identification of bacteria and yeast from positive blood culture bottles. Clin Chem 2011a;57:1057–67.
[9] Kaleta EJ, Clark AE, Johnson DR, Gamage DC, et al. Use of pcr coupled with electrospray ionization mass spectrometry for rapid identification of bacterial and yeast bloodstream pathogens from blood culture bottles. J Clin Microbiol 2011b;49:345–53.
[10] Elshaari, FA, Elshaari, FA and Elshaari, MA. New Algorithm for Selecting Degenerate Primers to Clone DNA Using a Short Peptide Sequence. ISAC Proceeding for Computational Biology, USA, 2012
[11] http://www.ncbi.nlm.nih.gov/genome/?term=bacteria: Accessed on May the 11th 2016
[12] Barghouthi S. A. A Universal Method for the Identification of Bacteria Based on General PCR Primers. Indian J Microbiol 2011; 51(4):430–444
[13] Mitterer G, Huber M, Leidinger E, Kirisits C, Lubitz W, Mueller MW, Schmidt WM Microarray-based identification of bacteria in clinical samples by solid-phase PCR amplification of 23S ribosomal DNA sequences. J Clin Microbiol 2004;42:1048–1057.
[14] Yasuoka MO A multiplex polymerase chain reaction– based diagnostic method for bacterial vaginosis. Obst Gynecol 2002; 100:759–764
[15] Muldrew K. L. Molecular diagnostics of infectious disease. Curr Opin Pediatr 2009; 21:102–111