Molecular and Cellular Probes 27 (2013) 80–85

Contents lists available at SciVerse ScienceDirect

Molecular and Cellular Probes

journal homepage: www.elsevier.com/locate/ymcpr

Massively parallel sequencing of enriched target amplicons for high-resolution genotyping of *Salmonella* serovars

Pallavi Singh [a], Steven L. Foley [b], Rajesh Nayak [b], Young Min Kwon [a,c,*]

[a] Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR 72701, USA
[b] Division of Microbiology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA
[c] Poultry Science Department, University of Arkansas, Fayetteville, AR 72701, USA

ARTICLE INFO

Article history:
Received 23 June 2012
Received in revised form
7 November 2012
Accepted 19 November 2012
Available online 29 November 2012

Keywords:
Multilocus sequence typing
Next generation sequencing
Target capturing
Strain typing
Salmonella

ABSTRACT

With next generation sequencing (NGS) technology, it is now possible to carry out in-depth, large-scale sequencing projects, such as whole genome sequencing, in a fast and inexpensive manner. However, often it is more practical and convenient to sequence and analyze multiple, smaller regions of the bacterial genome to gain valuable information about an organism. One such application is genotyping of bacterial strains by multilocus sequence typing (MLST) that involves PCR and sequencing analysis of typically 7 housekeeping genes. Recently, we described a novel MLST method, called MLST-seq that combines a PCR-based target enrichment method and NGS technology to simultaneously analyze numerous target gene sequences, thereby improving the resolution and high-throughput capacity of current MLST approaches. However, the performance of the MLST-seq method was hampered from a substantial bias in target enrichment step. In this study, we used an improved target enrichment method using hairpin selectors to amplify 21 target genes simultaneously from each of 41 *Salmonella* strains. The resulting amplicons tagged with strain-specific barcodes were pooled and sequenced *en masse* by 454 pyrosequencing. Analysis of sequence data from 38 *Salmonella* strains using combinations of 3, 7 and 14 target genes resulted in 23, 32 and 37 distinct allelic profiles, respectively. These results demonstrated that MLST-seq with an increased number of target genes is an efficient way to improve discrimination among closely-related strains of *Salmonella*. With the rapidly increasing sequencing capacity of NGS technologies combined with further improvements in target capturing methods, MLST-seq could become a promising approach to perform high-resolution strain typing of a large collection of *Salmonella*, and likely other genera in a labor- and cost-efficient manner in the future.

1. Introduction

Molecular typing methods that rely on DNA sequence differences are essential for the epidemiological studies of pathogenic bacteria. Multilocus sequence typing (MLST) is one of the widely used techniques for molecular typing and is based on the analysis of polymorphisms in defined genetic loci in the bacterial genome by PCR amplification and sequencing of the PCR products [1]. Initially most MLST schemes were focused on sequencing and analysis of a limited number of housekeeping genes [2]. Since these loci are relatively stable genetic markers, polymorphisms are typically good indicators of population diversity. While these conservative loci are valuable for long-term epidemiological studies, they often do not display adequate discrimination for short-term or local epidemiological studies [3]. This is mainly due to a lack of sufficient variations within the limited number of target genes sequenced for analysis. Bacterial housekeeping genes are typically present in the vast majority of strains of a particular species and are unlikely to be subjected to strong selective pressures that may lead to rapid sequence changes [4]. Therefore, researchers have attempted to increase the sensitivity in detecting genetic diversity of *Salmonella* by adding non-housekeeping genes, such as virulence genes [4,5] or by increasing the number of target genes sequenced. A potential drawback of the latter approach is the increase in labor and cost for the analysis of additional target loci.

Recently, we reported a prototype of a novel MLST approach to overcome the limited resolution of current MLST schemes by combining a novel PCR-based target enrichment strategy with sequencing *en masse* of the enriched targets by next generation sequencing (NGS) using 454 pyrosequencing platform (Roche 454 LifeSciences, Bradford, CT; Fig. 1) [6]. This approach, termed

* Corresponding author. Department of Poultry Science, University of Arkansas, 1260 W Maple Street, Fayetteville, AR 72701, USA. Tel.: +1 479 575 4935; fax: +1 479 575 8775.
E-mail address: ykwon@uark.edu (Y.M. Kwon).

0890-8508/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mcp.2012.11.004

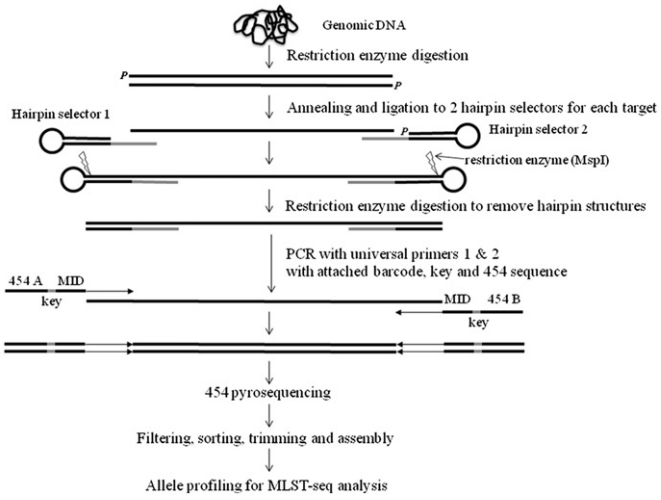

Fig. 1. Schematic diagram of the multiple target amplification protocol using hairpin selectors. Restriction enzyme-digested fragments of target genes were ligated to hairpin selectors and amplified through PCR reactions, using universal primers with attached multiplex identifiers (MID), key, and 454 sequences at 5′ ends.

MLST-seq, is an extension of existing MLST methods through the analysis of more target genes, compared to the 7 genes normally used in the current *Salmonella* MLST scheme [7]. One major shortcoming of our prototype MLST-seq was an uneven representation of all targets during PCR-enrichment step [6]. In this study, we employ an improved target enrichment method recently developed in our lab [8] to improve the performance of MLST-seq. We have evaluated the utility of this improved MLST-seq using a set of 41 *Salmonella* strains and demonstrated that this novel approach has great potential as a high-throughput strain typing method with enhanced discriminatory power.

2. Materials and methods

2.1. Bacterial strains, culture conditions and genomic DNA preparation

Forty-one *Salmonella enterica* isolates, including Heidelberg (n = 25), Montevideo (3), Senftenberg (3), Typhimurium (2), Enteritidis (2), Gaminara (2), Anatum (1), Worthington (1), Muenster (1), and an untypeable (rough) strain (1), were used in the study (Table S1). The strains were maintained at −80 °C in Brain Heart Infusion (BHI) broth with 30% glycerol. For genomic DNA isolation, the strains were cultured overnight in Luria Bertani (LB) broth at 37 °C in a shaking incubator, and genomic DNA was isolated from the overnight cultures using QIAamp DNA mini Kit (Qiagen, Valencia, CA). The purified genomic DNA samples were stored at −20 °C until used.

2.2. Target enrichment using hairpin selectors

Recently we described a target enrichment method in which target-specific hairpin selectors are used to create PCR targets with universal primer-binding sites attached on both ends [8]. The multiple universal target sequences can then be amplified using an universal primer pair (Fig. 1). To evaluate the performance of the MLST-seq based on this target enrichment method, twenty one target genes (*thrA*, *purE*, *sucA*, *hisD*, *aroC*, *hemD*, *dnaN*, *mdh*, *manB*, *fimA*, *panB*, *aceK*, *icdA*, *gyrB*, *atpD*, *fliC*, *fljB*, *pduF*, *glnA*, *glpF*, and *pgmB*) were selected from the genes that have been previously used for different *Salmonella* MLST schemes [3], including the seven genes used in the current *Salmonella* MLST scheme (Table S2) [13]. A pair of hairpin selectors were designed as described previously for each of 21 target genes ([8]; Table S2). The oligonucleotides for hairpin selectors were synthesized by Integrated DNA Technologies (Coralville, IA).

2.3. Preparation of PCR templates

PCR templates were prepared for each *Salmonella* strain using hairpin selectors as previously described [8]. First, to determine the appropriate restriction enzyme for digestion of genomic DNA samples, *in silico* restriction digestion analysis was carried out on the 21 selected target genes using NEBcutter software V2.0 (http://tools.neb.com/NEBcutter2/). Restriction enzymes were searched and analyzed for a suitable enzyme that would have digestion site in the coding region of every target gene such that lengths of produced digests were in a suitable range of ∼200–400 bp. This is important to avoid bias against long length products during multiplex amplification. Restriction enzyme MspI was chosen in this experiment because it generates fragments in the range of 191–604 bp for all 21 target genes. Then, all genomic DNA samples were digested with restriction enzyme MspI (New England Biolabs, Ipswich, MA). Ligation reaction mixtures, containing 18 nM of hairpin selectors 1 and 2, 1 μl of Ampligase Thermostable DNA ligase (5 U/μl), 2.5 μl of 10× reaction buffer (Epicentre Biotechnologies, Madison, WI), 1 μl of MspI digested genomic DNA, and ddH_2O to make up a final volume of 25 μl, were subjected to cyclic ligation reaction through 40 cycles of 95 °C for 1 min and 70 °C for 2 min using a thermal cycler. After ligation, 3 μl NEBuffer 2 and 1 μl MspI enzyme were added to the ligation mix and ddH_2O was added to make up a final volume of 30 μl for the digestion reaction. The reaction was incubated at 37 °C for 1 h, followed by enzyme inactivation at 80 °C for 20 min. The MspI digestion trims the hairpin structure at both the 5′ and 3′ends and facilitates subsequent PCR amplification with universal primers.

2.4. Amplicon preparation for 454 pyrosequencing

For each strain, PCR mixtures were prepared containing 1 μl of DNA template (100–200 ng/μl), 1 μl of cloned *pfu* DNA polymerase (Stratagene, La Jolla, CA), 5 μl of 10× cloned pfu DNA polymerase buffer, 4 μl of 2.5 mM dNTP mix, 1 μl each of 350 ng/μl universal primers 1 and 2 and ddH_2O to make up the final 50 μl volume. PCR was carried out with a hot start amplification reaction using an initial denaturation step at 94 °C for 2 min followed by 30 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min followed by an extension step at 72 °C for 10 min. The universal primer 1 (5′-CGTATCGCCTCCCTCGCGCCA-TCAG-MID-CCTGAATTCGGATGGA ATTCATG-3′) and universal primer 2 (5′-CTATGCGCCTTGCCAGCCC GC-TCAG-MID-CTGAATTCGGATGTCTAGAATG-3′) contained unique 10 bp multiplex identifiers (MID), 4 bp keys and 454 platform-specific adapter sequences 454 A and B, respectively. The MIDs (Table S1; designed and tested by Roche FLX Titanium for bi-directional sequencing) were incorporated into the universal primers to identify the individual strains with forward and reverse reads after sequencing.

A total of 21 target genes were amplified in a multiplex reaction and prepared for 454 pyrosequencing from each of 41 *Salmonella* strains. Amplicons were verified on 6–12% Tris Borate EDTA gel (Invitrogen, Carlsbad, CA) for multiplex amplification of all targets. Equal volumes of 1 μl each of amplification products from all the strains were pooled together and purified using QIAquick PCR purification kit (Qiagen). The concentrations of the purified products were measured using NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and the products were sent to the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida for pyrosequencing using 454 FLX Titanium chemistry (Roche 454 LifeSciences).