

026. 7. 30.(목) 11:00
보도 시점 7. 31.(금) 조간

배포 2026. 7. 30.(목) 06:00

겉보기 같은 대장균, '유전자 지문'으로 세밀하게 가린다

- 농진청, 공개 유전체 83개 분석해 대장균 14개 유형으로 정밀 분류 기술 개발
- 기존 검사에서 놓치기 쉬운 균주 간 미세 차이까지 고해상도로 식별
- 국제학술지 「BMC Genomics」 게재...해외 식품 안전 규제 대응력 대폭 강화

겉보기에 같아도 위험도는 천차만별인 대장균을 '유전자 지문'으로 정확히 구별해내는 기술이 개발돼 농산물 안전관리에도 혁신이 기대된다.

농촌진흥청(청장 이승돈)은 대장균의 독특한 유전자 배열 패턴을 분석해 균주의 정밀한 '족보'를 찾아내는 새로운 유전형 분석기술을 개발했다고 밝혔다.

최근 미국 식품안전현대화법(FSMA), 유럽연합(EU) 위생 규정 등 세계시장의 농산물 안전관리가 강화되는 추세다. 이에 우리 농산물의 안전성을 과학적으로 뒷받침하고, 수출 전 세밀한 품질 관리로 수출 경쟁력을 키울 수 있는 기술 개발이 필요하다.

연구진은 공개된 83개의 대장균 유전체 정보를 대조·분석했다. 여기서 발견한 유전적 특이성을 바탕으로 균주를 분류하고 14개의 유형으로 정의해 대장균을 식별했다.

그동안은 병원성 유전자 검사(PCR), 표준 유전형 분석(MLST)으로 대장균을 분석했다. 그러나 병원성 유전자 검사는 병원성 유무는 확인할 수 있으나 같은 유전자를 가진 균주 간 변이는 찾을 수 없고, 표준 유전형 분석은 유전적으로 비슷한 균주들을 세밀하게 분별하기 어렵다.

이번에 개발한 분석기술은 기존 검사법이 놓치던 세균의 미세한 변이 영역까지 찾아낸다. 이를 통해 유전적으로 매우 비슷한 세균이라도 세밀하게 구분할 수 있다.

새 분석기술의 균주 변별력은 0.726으로 기존 표준 유전형 분석법 (MLST, 0.723)과 동등한 수준을 보이면서도, 기존 분석법으로는 구분하기 어려웠던 유전적 다양성까지 정밀하게 분석할 수 있음을 확인했다.

* 심슨 다양성 지수(Simpson's diversity index): 균주를 구별하는 능력을 나타내는 지표로, 1에 가까울수록 구별 능력이 뛰어나는 의미를 지닌다.

또한, 유전자 지문 분석만으로도 대장균의 병원성 및 항생제 내성 특성을 예측할 수 있음을 확인했다. 즉, 유전자 지문 패턴이 길수록 독성 유전자가 많고, 짧을수록 항생제 내성 유전자가 많은 상반된 경향을 보인 것이다.

이번 연구 결과는 국제학술지 'BMC Genomics (IF 3.9)'에 논문으로 게재됐다. 연구진은 앞으로 추가적인 현장 적용성 검증 과정을 거친 후 농산물 안전관리 기관과 관련 검사 현장에 본 기술을 보급할 예정이다.

농촌진흥청 국립농업과학원 농업미생물과 한상현 과장은 “사람마다 지문이 다르듯 똑같아 보이는 대장균도 저마다 고유한 유전자 지문을 가지고 있다.”라며, “이번 연구는 기존 분석법의 한계를 극복하고 대장균의 미세한 유전적 다양성을 명확히 가려낼 실용적인 연구 틀을 구축한 데 의미가 있다.”라고 말했다.

붙임 1. 대장균 정밀 식별 기술 개발

2. 대장균 기존 분석법과 신규 개발 기술 비교

담당 부서	국립농업과학원 농업미생물과	책임자	과 장	한상현 (063-238-3021)
		담당자	연구사	박경민 (063-238-3096)
	농촌진흥청에서 연구·개발한 농업의 모든 것  농사로			

□ 연구 배경

- (기존 유전형 분석법의 한계) 대장균(*Escherichia coli*)의 종류를 구분하고 균주 간 차이를 분석하기 위해 기존에는 단일염기다형성(SNP) 분석이나 다중유전자서열타이핑(MLST) 등을 주로 사용함.
- ⇒ 하지만 이러한 방법은 변이가 적은 유전자만 분석하기 때문에, 균주 간 숨겨진 유전적 차이를 세밀하게 구분하는 데 한계가 있었음.
- (목적) 기존 분석법의 한계를 보완하기 위해 유전자 내 단백질 반복 패턴과 주변 유전자 배열 정보를 함께 활용한 새로운 유전형 분석기술을 개발함.

□ 연구 결과

- (과정) 공개된 83개의 대장균 유전체 데이터를 대조·분석하여 다음과 같은 유전적 특이성을 발견하고 균주를 분류함.
 - RSXC 유전자의 아미노산 반복 패턴을 기준으로 7가지 유형으로 분류
 - * 83개 균주를 총 7가지 반복 유형(R7a, R7b, R6a, R6b, R5a, R5b, R4a)으로 분류함.
 - 유전자 주변 구조(syntenly)에서 변이가 많이 나타나는 영역을 분석하고, 아미노산 반복 서열 정보와 결합해 총 14개의 유전형(ST1~ST14)을 정의함.
- (결과) 새로 개발한 분석법의 균주 변별력(Simpson's diversity index)은 0.726으로 측정되었으며, 이는 기존에 널리 쓰이는 표준 방식인 MLST (0.723)와 대등하거나 능가하는 수준

□ 기대 효과

- 기존 유전학적 분석법이 간과하기 쉬웠던 대장균 균주 간의 유전적 다양성을 효과적으로 포착할 수 있는 실용적인 프레임워크를 제공
- 향후 감염병 역학 조사 및 병원체 감시 체계 시스템에서 유용한 바이오마커로 활용될 가능성이 높음.



New genotyping method for *Escherichia coli* using RSXC repeat polymorphism and gene-neighborhood architecture

Ravi Jothi¹, Kyung Min Park¹, Soobin Yim¹, Kwang-Kyo Oh¹, Injun Hwang¹, Myeong-In Jeong¹ and Dong Suk Park^{1*}

Abstract

Background Reliable genotyping tools are essential for understanding genomic diversity and tracking strain variation in *Escherichia coli*. Traditional typing approaches, including SNP and MLST based methods, focus mainly on conserved regions of the genome and may overlook potential changes in the unconserved gene content. To address this gap, this study proposed a new genotyping approach that combines diversity profiles observed from repeat-rich protein sequence and genomic neighborhood structural organization.

Results Comparative genomic analysis was conducted on 83 publicly available *E. coli* genomes. Sequence comparison identified a repeat hotspot within the RSXC protein, while gene neighborhood (synteny) analysis revealed three variable genomic regions (hotspots A–C). Differences in RSXC repeat copy number and terminal motifs allowed the strains to be classified into seven repeat types (R7a, R7b, R6a, R6b, R5a, R5b, and R4a). Structural variation within genomic hotspots A, B, and C was characterized by the presence, absence, and arrangement of intermediate genes and classified into 7, 8, and 4 structure types, respectively. Integration of RSXC repeat profiles with genomic hotspot patterns generated 14 distinct sequence types (ST1–ST14). RSXC repeat patterns and the combined genotyping scheme both showed good discriminatory power (Simpson's diversity index of 0.726), which is comparable to MLST typing (0.723) among the 83 *E. coli* strains analysed. Virulence and antimicrobial resistance (AMR) profiling of the examined strains further supported biological relevance of the proposed method. Strains with higher RSXC repeat numbers (R7 and R6) positively linked with virulence genes presence and negatively associated with AMR gene counts as supported by a significant positive correlation with virulence genes ($p = 0.323$, $p = 0.0026$) and a negative correlation with AMR genes ($p = -0.276$, $p = 0.0105$).

Conclusions This repeat–synteny based genotyping approach captures genomic diversity in the 83 *E. coli* strains that is often overlooked by conventional typing approaches and provides a practical framework for discriminating *E. coli* strains. The strong association with MLST, combined with meaningful relationships to virulence and AMR profiles, suggests that this approach may serve as a useful marker for epidemiological identifications.

Keywords *E. coli*, Amino acid repeats (AAR), Repeat hotspots, Genotyping, MLST, pathogen surveillance

*Correspondence:

Dong Suk Park
dspark@nida.go.kr

¹ Agricultural Microbiology Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea



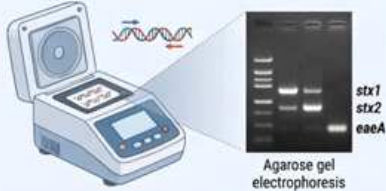
© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

'대장균 분석 기술의 정밀도 고도화: PCR, MLST 및 신규 유전형 분석 기술의 3단계 비교

[1단계] 병원성 유전자 검사 (PCR)

주요 목적: 특정 병원성 유전자 유무 진단

샘플 내에 병원성 대장균이 존재하는지 여부를 신속하게 판정하는 1차 진단 단계입니다.



분석 타겟: *stx1*, *stx2*, *eaeA* 유전자

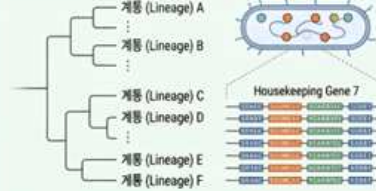
시가독소 생산 및 장 부착 능력과 관련된 핵심 유전자 시퀀스를 탐지합니다.

정보 및 한계: 병원성 여부 확인 및 개별 구분 불가
병원성 유무는 확인 가능하나, 동일한 유전자를 가진 균주 간의 변이를 구별할 수 없습니다.

[2단계] 표준 유전형 분석 (MLST)

주요 목적: 균주의 거시적 계통 분류

균주가 속한 상위 계통(Lineage)을 분류하여 역학적 관계의 기본 틀을 제공합니다.



분석 타겟: 7개의 하우스키핑 유전자 (Conserved)

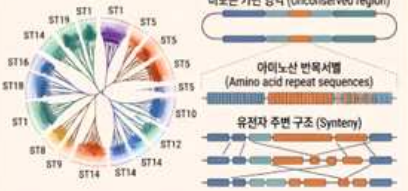
유전적으로 잘 보존된(Conserved) 영역을 분석하여 안정적인 계통 분류를 수행합니다.

기술적 한계: 유사 균주 간 변별력 부족
보존 영역 중심의 분석으로 인해 유전적으로 유사한 균주들을 세밀하게 분별하는 해상도가 낮습니다.

[3단계] 신규 유전형 분석 기술

주요 목적: 정밀 유전형 분류 및 특성 예측

고해상도 분석을 통해 균주 간의 미세한 차이를 규명하고 기능적 특성을 도출합니다.



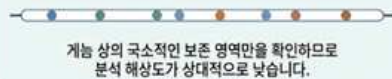
분석 타겟: 아미노산 반복서열 및 유전자 주변 구조 (Unconserved)

변이가 잦은 비보존 가변 영역(Unconserved region)과 유전자 배열 구조(Synteny)를 통합 분석합니다.

분석 성과: 14개 세부 유전형(ST1-ST14) 및 변별력 지수 0.726
기존 범으로 구분이 어려운 균주를 14개 유전형으로 세분화하며, 비보존 영역 커버리지를 극대화했습니다.

확장 정보: 병원성 및 AMR(항생제 내성) 형질 예측
유전형 분석을 넘어 해당 균주의 실제 병원성 위험도와 약물 내성 프로파일링 정보를 제공합니다.

MLST: 불연속적 '점(Points)' 기반 분석



계통 상의 국소적인 보존 영역만을 확인하므로 분석 해상도가 상대적으로 낮습니다.

계통 매핑 비교

신규 기술: 광범위 '블록(Broad Blocks)' 기반 분석



반복 영역과 주변 구조를 포함하는 넓은 범위를 커버하여 유전형 다양성을 정밀하게 포착합니다.