

DIVERSITY AND PHYLOGENETIC CHARACTERIZATION OF *RALSTONIA*-ASSOCIATED 16S rRNA SEQUENCES DETECTED IN MYOCARDIAL INFARCTION BLOOD SAMPLES FROM PAKISTAN

Areeb Khalid¹, Samyyia Abrar^{*2}, Rashid Saif³

^{1,*2}Department of Biological Sciences, The Superior University, Lahore, Pakistan

³Department of Life Sciences, Qarshi University, Lahore, Pakistan

^{*2}sammiya.abrar@superior.edu.pk

DOI: <https://doi.org/10.5281/zenodo.23152809>

Keywords

myocardial infarction; 16S rRNA; Sanger sequencing; *Ralstonia*; molecular diversity; phylogenetic analysis; BLASTn; bacterial DNA; Pakistan

Article History

Received: 26 June 2026

Accepted: 17 September 2026

Published: 05 October 2026

Copyright @Author

Corresponding Author: *

Samyyia Abrar

Abstract

Background: Bacterial DNA has been found in cardiovascular tissue and blood in several molecular studies, but little is known about the diversity of these sequences in patients with myocardial infarction (MI), and almost nothing has been reported from Pakistan. Broad-range 16S rRNA gene sequencing offers a practical way to characterize bacterial DNA detected by PCR and to place it within a phylogenetic framework.

Methods: Bacterial 16S rRNA PCR was carried out on peripheral blood from 50 clinically confirmed MI patients in Lahore, Pakistan. Fourteen samples gave usable Nested-PCR products (V3–V4 region), and the ten best amplicons were sequenced by the Sanger method. Sequences were cleaned in BioEdit and compared against GenBank with BLASTn. Alignment was done with MUSCLE in MEGA 12, and pairwise identity, ambiguous nucleotide positions and Maximum Composite Likelihood distances were calculated. A Neighbor-Joining tree with 500 bootstrap replicates was built from the ten study sequences and 20 *Ralstonia* reference sequences.

Results: The cleaned sequences were 391–420 bp long (4,083 bp in total). BLASTn placed all ten closest to *Ralstonia* or uncultured *Ralstonia*-associated sequences, with 95.82–99.75% identity and 99–100% query coverage. Identity among the study sequences ranged from 96.21% to 100.00%. Fourteen ambiguous positions (0.343% of analyzed bases) occurred in nine of the ten sequences. In the phylogenetic analysis the study sequences fell within the wider *Ralstonia*-associated group.

Conclusions: The ten sequences are highly similar to one another and are most closely related to the genus *Ralstonia*. The partial 16S rRNA fragments analyzed here may not provide sufficient resolution for definitive species-level identification, and detection of bacterial DNA does not establish bacterial viability or a causal role in MI. These sequences provide an initial molecular dataset of *Ralstonia*-associated 16S rRNA sequences detected in MI blood samples from Pakistan.

1. Introduction

Myocardial infarction (MI) arises mainly from coronary atherosclerosis followed by acute thrombosis. Atherosclerosis is now regarded as a chronic inflammatory disease, driven by

endothelial dysfunction, lipid deposition and the recruitment of immune cells [1,2]. Beyond the usual cardiovascular risk factors, microorganisms and their products have been proposed as one possible source of vascular

inflammation. Several molecular studies have reported bacterial DNA in atherosclerotic plaques, coronary thrombi and peripheral blood. The organisms named most often include *Chlamydia pneumoniae*, *Helicobacter pylori* and a number of periodontal bacteria [3-5]. Finding bacterial DNA, however, is not the same as finding living bacteria or an active infection, so these results need careful interpretation [6].

The 16S rRNA gene is among the most widely used markers for bacterial identification and phylogeny. Its conserved regions allow broad-range primers to work across taxa, while its variable regions carry the signal used to tell taxa apart [7]. A BLAST search against GenBank gives a first taxonomic assignment, and alignment [8], with a phylogenetic tree can then show how the sequences relate to each other and to reference organisms [7,9]. *Ralstonia* is a genus of Gram-negative, non-fermenting bacteria found in water, soil and plant-associated habitats. Some species, notably *R. pickettii*, *R. mannitolilytica* and *R. insidiosa*, have been isolated from clinical specimens and behave as opportunistic pathogens in certain settings. They are easily confused with other non-fermenters, which makes molecular characterization useful [10,11].

Most work linking bacteria to cardiovascular disease has concentrated on recognized cardiovascular or periodontal pathogens [3-5]. Published information on *Ralstonia*-associated sequences in blood from patients with MI appears to be limited, particularly in the Pakistani context. The present manuscript focuses exclusively on the molecular characterization of the sequenced bacterial 16S rRNA fragments and does not repeat the clinical or screening analyses conducted in the broader study. The relationship of this sequence-focused analysis to the broader study is stated explicitly to distinguish the present work from the primary screening report. The aim of the present study was to characterize the nucleotide features of the sequenced fragments, assess their similarity to database sequences and to each other, identify ambiguous nucleotide positions, and determine their phylogenetic placement among representative *Ralstonia* reference sequences. The study was not

designed to evaluate clinical associations or test relationships between bacterial DNA detection and patient characteristics.

2 Materials and Methods

2.1 Study Population and Sample Collection

This sequence-focused molecular analysis was conducted using peripheral blood samples collected from 50 clinically confirmed myocardial infarction (MI) patients admitted to tertiary care hospitals in Lahore, Pakistan. Samples were collected following ethical approval and written informed consent from participants. Patients meeting standard clinical diagnostic criteria for acute MI were included, whereas individuals with recent antibiotic exposure, ongoing acute infection, immunocompromised conditions, or other conditions that could interfere with bacterial DNA detection were excluded.

2.2 DNA Extraction and 16S rRNA Gene Amplification

Genomic DNA was extracted from peripheral blood samples under sterile conditions using a standard DNA extraction protocol. Initial bacterial screening was performed by conventional PCR targeting the 16S rRNA gene using the universal primer pair 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3'), which generated an expected amplicon of approximately 1465 bp.

Samples showing detectable or weak amplification were subsequently examined by nested PCR using the internal primer pair 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3'), targeting the V3-V4 region and generating an expected product of approximately 460-465 bp. The nested-PCR reaction was performed in a final volume of 14 µL containing 1 µL template DNA, 1 µL each of forward and reverse primers, 7 µL PCR master mix, and 4 µL nuclease-free water.

Initial amplification was performed with an initial denaturation at 95°C for 5 min, followed by 30-35 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at

74°C for 90 s, followed by a final extension at 72°C for 7 min. The annealing temperature was subsequently optimized to 56°C for subsequent amplification and re-amplification of PCR-positive samples.

2.3 Gel Electrophoresis and Selection of Sequencing Products

PCR products were separated by electrophoresis on 1% agarose gels prepared in TAE buffer and stained with ethidium bromide. Electrophoresis was performed at 100 V for 45 min, and amplicons were visualized under UV illumination. A GeneDireX DNA ladder was used for amplicon-size estimation.

Fourteen samples showing the expected nested-PCR amplification were selected for further sequence analysis. These samples were re-amplified under the optimized PCR conditions in a final reaction volume of 40 µL. Following amplification, 5 µL of each product was examined by agarose gel electrophoresis, while the remaining 35 µL was retained for sequencing. Ten samples showing the strongest and clearest amplification products and sufficient material for sequencing were selected for Sanger sequencing. PCR products selected for sequencing were purified by ethanol precipitation.

2.4 Sanger Sequencing and Sequence Processing

Purified PCR products were subjected to Sanger sequencing using the 341F primer. Resulting chromatograms were examined for sequence quality and edited using BioEdit to obtain cleaned nucleotide sequences [12]. The resulting sequences were used for downstream taxonomic and phylogenetic analyses.

2.5 BLASTn-Based Taxonomic Characterization

The cleaned nucleotide sequences were compared against sequences available in the NCBI GenBank database using BLASTn [8]. The closest database matches were evaluated based on percentage sequence identity and

query coverage. GenBank accession numbers were recorded for the sequences generated in the present study.

2.6 Multiple Sequence Alignment and Phylogenetic Analysis

The ten study sequences were aligned with 20 reference sequences representing the genus *Ralstonia* retrieved from GenBank. Multiple sequence alignment was performed using MUSCLE [13] implemented in MEGA version 12. Phylogenetic relationships were inferred using the Neighbor-Joining method based on Maximum Composite Likelihood distances [14]. Gaps and missing data were treated using pairwise deletion, and branch support was evaluated using 500 bootstrap replicates.

2.7 Sequence Characteristics and Descriptive Analysis

Descriptive analyses were performed to summarize the characteristics of the ten Sanger-derived sequences. Sequence length, GC content, BLASTn percentage identity, and query coverage were summarized using descriptive statistics. Ambiguous nucleotide positions identified in the cleaned sequences were recorded and summarized according to their frequency and distribution. The overall ambiguity rate was calculated as the number of ambiguous nucleotide positions divided by the total number of cleaned nucleotide bases.

3. Results

3.1 Sequence generation and characteristics

Fourteen of the 50 screened samples produced the expected nested-PCR amplicon, of which 10 were selected for Sanger sequencing. Only these ten sequences are analyzed here. After trimming, the sequences were 391–420 bp long (mean 408.3 bp; 4,083 bp in total) with GC contents of 52.40–54.18% (mean 53.53%) (Table 2). Length differences reflect the removal of poor-quality ends. Sequence S6, for example, went from about 445 bp raw to 417 bp cleaned.

Table 2. Characteristics of the ten cleaned 16S rRNA sequences.

Sample	GenBank accession	Length (bp)	GC (%)	Ambiguous positions
S1	PZ626114.1	417	54.11	3
S2	PZ626115.1	416	53.38	2
S3	PZ626116.1	391	53.96	0
S4	PZ626117.1	406	52.99	2
S5	PZ626118.1	403	53.48	1
S6	PZ626119.1	417	52.40	1
S7	PZ626120.1	420	54.18	1
S8	PZ626121.1	410	53.79	1
S9	PZ626122.1	405	53.71	1
S10	PZ626123.1	398	53.28	2
Total / mean	—	4,083 / 408.3	53.53 (mean)	14

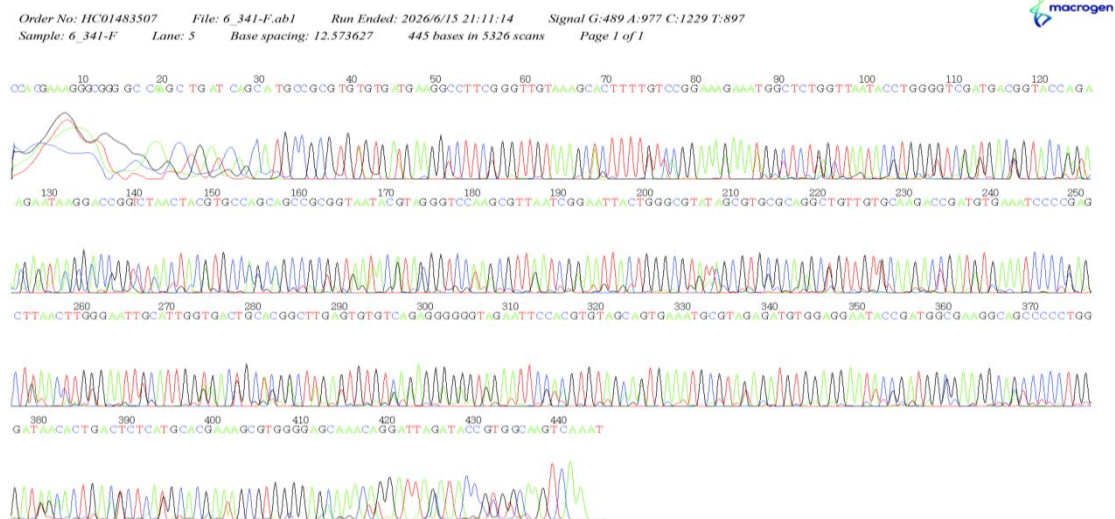


Figure 2. Representative Sanger chromatogram showing the quality-controlled region of a study sequence.

3.2 BLASTn similarity

The closest BLASTn matches for the ten study sequences were predominantly *Ralstonia*-associated or uncultured bacterial sequences. Identity ranged from 95.82% to 99.75% and query coverage from 99% to 100% (Table 3).

Several top hits were uncultured bacterial clones or environmental sequences rather than named species, and identity varied from sequence to sequence. We therefore read these results as genus-level similarity and not as species identification.

Table 3. Closest BLASTn match for each study sequence.

Sample	Closest BLAST description	Identity (%)	Query cover (%)
S1	<i>Ralstonia</i> sp. strain NEM22	97.83	99
S2	Uncultured bacterium isolate BII-2-9	96.63	100
S3	<i>Ralstonia</i> sp. strain NEM22	99.74	100

Sample	Closest BLAST description	Identity (%)	Query cover (%)
S4	<i>Ralstonia pickettii</i>	95.82	100
S5	Uncultured <i>Ralstonia</i> sp. clone 3F05	96.80	100
S6	Uncultured <i>Ralstonia</i> sp. DGGE band R56	98.06	99
S7	Uncultured bacterium isolate BII-2-9	99.27	100
S8	Uncultured <i>Ralstonia</i> sp. DGGE band R56	99.51	99
S9	Uncultured prokaryote clone 152125	99.51	100
S10	Uncultured bacterium clone EHB-P0145 #01	99.75	100

3.3 Identity and distance among study sequences

Pairwise identity among the ten sequences ranged from 96.21% (S4 vs S6) to 100.00% (S3 vs S10 and S8 vs S10) (Table 4). Pairwise nucleotide identity among the study sequences ranged from 96.21% to 100.00%, indicating substantial sequence similarity with some

nucleotide-level variation among the analyzed fragments. Maximum Composite Likelihood distances agreed with the identity values pairs with high identity had small distances and pairs with lower identity had larger ones. Distances ranged from 0.00000 (S3–S10) to 0.04276 (S1–S6); the full matrix is in Supplementary Table .

Table 4. Pairwise nucleotide identity (%) among the ten study sequences.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	—	97.53	99.48	96.73	97.49	97.32	99.02	99.25	99.00	99.49
S2	97.53	—	97.94	96.49	97.74	98.27	97.80	97.54	98.25	98.22
S3	99.48	97.94	—	97.15	97.68	98.72	99.74	99.74	98.97	100.00
S4	96.73	96.49	97.15	—	97.49	96.21	97.49	97.24	97.19	97.16
S5	97.49	97.74	97.68	97.49	—	97.49	97.75	97.74	98.22	97.94
S6	97.32	98.27	98.72	96.21	97.49	—	98.53	98.51	98.76	98.73
S7	99.02	97.80	99.74	97.49	97.75	98.53	—	99.51	99.26	99.75
S8	99.25	97.54	99.74	97.24	97.74	98.51	99.51	—	99.50	100.00
S9	99.00	98.25	98.97	97.19	98.22	98.76	99.26	99.50	—	99.75
S10	99.49	98.22	100.00	97.16	97.94	98.73	99.75	100.00	99.75	—

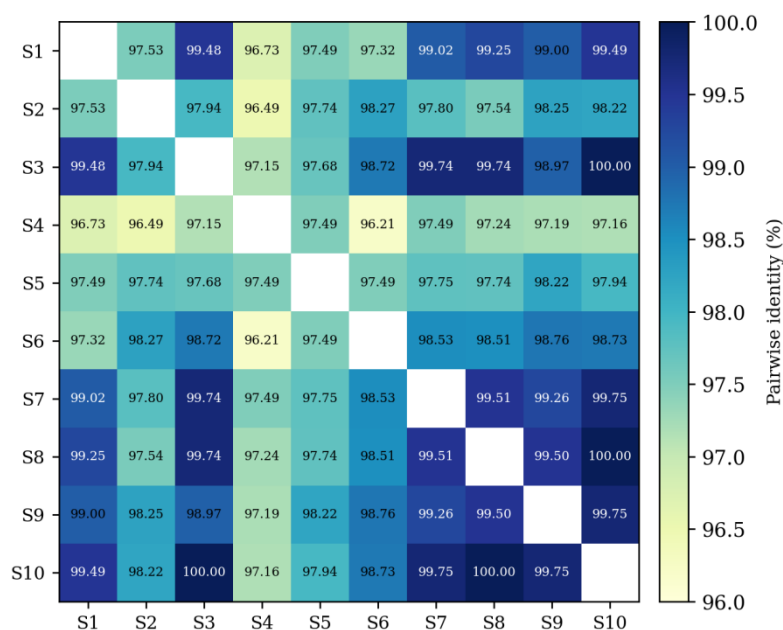


Figure 3: Heatmap showing pairwise nucleotide identity (%) among the ten study 16S rRNA sequences (S1–S10). Higher values indicate greater sequence similarity between the corresponding sequence pairs.

3.4 Ambiguous nucleotide positions

Fourteen ambiguous positions were found across the ten sequences, which is 0.343% of the 4,083 bases analyzed. Nine sequences had at least one; S3 had none. W (A/T) was the

most common code (7 of 14), followed by R (3) and K (3), with Y seen once (Table 5). The positions lay between about base 38 and base 400 (Figure 4; exact positions in Supplementary Table S1).

Table 5. IUPAC ambiguity codes found in the ten sequences.

IUPAC code	Bases	Count	Share of ambiguous positions (%)
W	A/T	7	50.00
R	A/G	3	21.43
K	G/T	3	21.43
Y	C/T	1	7.14
Total	—	14	100

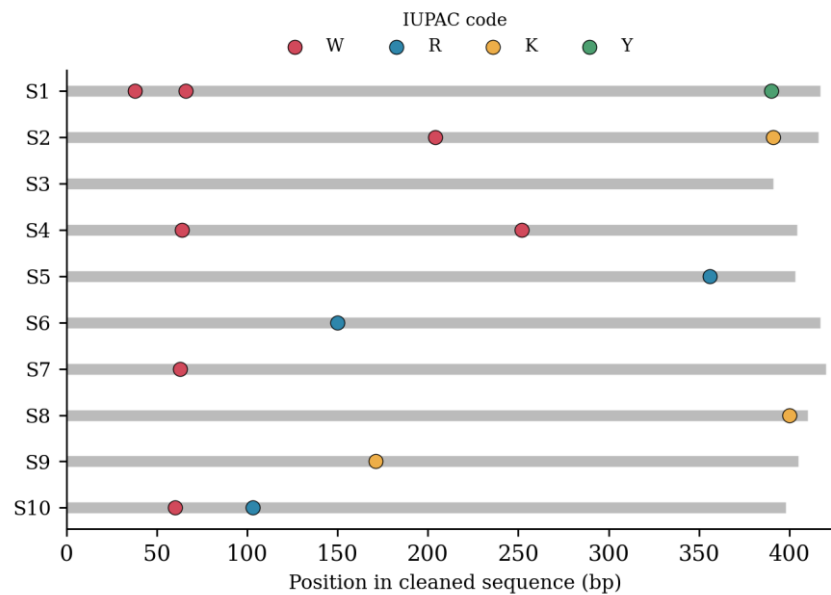


Figure 4: Distribution of ambiguous nucleotide positions (W, R, K, and Y) across the ten cleaned 16S rRNA sequences. Positions are shown relative to the respective cleaned sequence lengths.

3.5 Phylogenetic placement

The tree contained 30 taxa: the ten study sequences and 20 *Ralstonia* references (Table 6). Consistent with the BLASTn results, the study sequences grouped within the wider *Ralstonia*-associated clade (Figure 5). They were not

identical across the aligned region and sat at different distances from the reference taxa. Because the fragments are short, branch positions show sequence relatedness. They do not establish species identity, geographic origin, transmission or epidemiological linkage.

Table 6. Reference sequences used in the phylogenetic analysis.

No.	Accession	Taxon	Country
1	PZ267054.1	<i>R. solanacearum</i>	China
2	NR_025385.1	<i>R. mannitolilytica</i>	Belgium
3	NR_043152.1	<i>R. pickettii</i>	USA
4	LC868264.1	<i>R. pickettii</i>	Japan
5	AM501944.1	<i>R. pickettii</i>	Ireland
6	AM690480.1	<i>R. solanacearum</i>	Italy
7	AJ270258.1	<i>R. thomasi</i>	Belgium
8	NR_134148.1	<i>R. pseudosolanacearum</i>	Australia
9	LC772176.1	<i>R. mannitolilytica</i>	Iraq
10	MW642189.1	<i>R. mannitolilytica</i>	India
11	MW020025.1	<i>R. mannitolilytica</i>	Peru
12	EF427776.1	<i>R. mannitolilytica</i>	Australia
13	MF590120.1	<i>R. mannitolilytica</i>	Ghana
14	MH890459.1	<i>R. mannitolilytica</i>	Indonesia
15	KT933223.1	<i>R. mannitolilytica</i>	Brazil
16	NR_025975.1	<i>R. syzygii</i>	Australia

No.	Accession	Taxon	Country
17	AB021403.1	<i>R. syzygii</i>	Japan
18	PX122332.1	<i>R. syzygii</i>	China
19	NR_025242.1	<i>R. insidiosa</i>	USA
20	LT548965.1	<i>R. insidiosa</i>	Spain

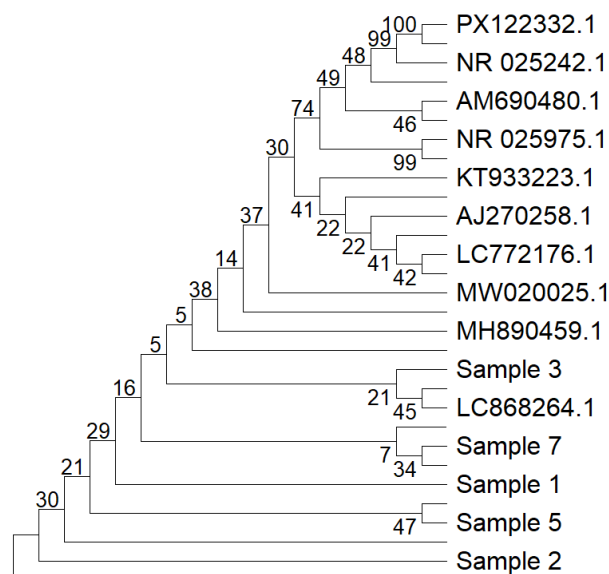


Figure 5: Neighbor-Joining phylogenetic tree showing the relationships of the ten study 16S rRNA sequences (S1–S10) with 20 reference sequences representing the genus *Ralstonia*. Bootstrap values based on 500 replicates are shown at the nodes, and the scale bar indicates substitutions per site.

4. Discussion

We characterized ten bacterial 16S rRNA fragments amplified from the blood of MI patients. All ten sequences showed closest similarity to *Ralstonia*-associated sequences, and all ten grouped with *Ralstonia* references in the phylogenetic analysis. The agreement between the two approaches supports a genus-level assignment. That assignment has limits. The fragments are about 400 bp of a single gene, and closely related *Ralstonia* species can share nearly identical 16S sequences. Several of the best BLAST hits were uncultured clones. We therefore describe the sequences as *Ralstonia*-associated and make no claim about species [7,9]. The study sequences were similar to one another (96.21–100.00% identity) but not identical, and the distance estimates matched this. Fourteen ambiguous positions appeared among 4,083 bases. Mixed or overlapping peaks

can arise from sequencing quality, template mixtures or other technical causes, so we treat them as a data-quality observation, not as evidence of mutation or of mixed populations [7].

Members of the genus have been reported from environmental sources and from clinical material, including blood [10]. Even so, our data do not show that *Ralstonia* was alive in these samples. PCR detects DNA, and low-biomass work is especially exposed to contamination from reagents, the laboratory or the environment, as well as to transient bacterial DNA [11]. The findings are preliminary and need independent confirmation. The study also cannot say that *Ralstonia* has anything to do with MI. There was no healthy control group in this sequence-level analysis, and detecting bacterial DNA does not by itself show causation. Likewise, the closeness

of our sequences to references from many countries says nothing about where the DNA came from or how it got there. The work has some strengths. Detection was confirmed by sequencing and not only by gel electrophoresis, BLASTn and phylogenetic analysis were used together, the reference set spanned several species and regions, and the sequences are public in GenBank for others to reuse.

Its limitations are equally clear. Only ten sequences were analyzed, all were partial 16S fragments, and there was no control group. Viability was not tested. The contamination risk of low-biomass samples remains, and the few ambiguous positions cannot be interpreted biologically. Later studies should use more samples, longer 16S reads or whole-genome and metagenomic sequencing, matched controls, and systematic contamination monitoring.

5. Conclusions

Ten Sanger-derived 16S rRNA sequences showed closest similarity to *Ralstonia*-associated sequences by BLASTn and phylogenetic analysis. They were highly similar to each other (96.21–100.00%) and contained 14 ambiguous positions across 4,083 bases. This is preliminary evidence of *Ralstonia*-associated bacterial DNA in these samples. It is not evidence of viable infection or of a causal link with MI, and partial 16S data cannot settle species identity. The deposited sequences can serve as a starting point for larger, controlled studies using longer sequences or genome-level methods.

Declarations

Data availability. The ten sequences are deposited in NCBI GenBank under accession numbers PZ626114.1–PZ626123.1, corresponding to samples S1–S10.

Ethics approval and consent. Ethical approval and informed consent were obtained for the broader clinical sample collection.

Author contributions. A.K: conceptualization, investigation, molecular and phylogenetic analysis, data curation, writing (original draft). S.A: supervision, methodology, writing (review and editing). R.S: co-supervision, methodology, writing (review and editing).

Funding. No external funding was received for this study.

Competing interests. The authors declare no competing interests.

REFERENCES

- Libby P. Inflammation in atherosclerosis. *Nature*. 2002;420(6917):868-874. doi:10.1038/nature01323.
- Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*. 2005;352(16):1685-1695. doi:10.1056/NEJMra043430.
- Ott SJ, El Mokhtari NE, Musfeldt M, Hellmig S, Freitag S, Rehman A, et al. Detection of diverse bacterial signatures in atherosclerotic lesions of patients with coronary heart disease. *Circulation*. 2006;113(7):929-937. doi:10.1161/CIRCULATIONAHA.105.579979.
- Koren O, Spor A, Felin J, Fåk F, Stombaugh J, Tremaroli V, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4592-4598. doi:10.1073/pnas.1011383107.
- Mitra S, Drautz-Moses DI, Alkhnbashi OS, Singh KK, Chen L, Petersen N, et al. Deciphering the microbiome in coronary atherosclerosis. *PLoS One*. 2015;10(4):e0120560. doi:10.1371/journal.pone.0120560.
- Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol*. 2019;27(2):105-117. doi:10.1016/j.tim.2018.11.003.
- Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol*. 2007;45(9):2761-2764. doi:10.1128/JCM.01228-07.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2.

- Clarridge JE 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev.* 2004;17(4):840-862.
doi:10.1128/CMR.17.4.840-862.2004.
- Ryan MP, Pembroke JT. *Ralstonia* spp.: emerging global opportunistic pathogens. *Eur J Clin Microbiol Infect Dis.* 2014;33(3):291-304.
doi:10.1007/s10096-013-1975-9.
- Coenye T, Vandamme P. Diversity and significance of *Ralstonia* species other than *R. solanacearum*. *Syst Appl Microbiol.* 2003;26(2):276-284.
doi:10.1078/072320203322346145.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;(41):95-98.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-1797.
doi:10.1093/nar/gkh340.
- Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38(7):3022-3027.
doi:10.1093/molbev/msab120.

