Award Number: W81XWH-10-1-0753

TITLE: Prospective Molecular Characterization of Burn Wound Colonization: Novel Tools and Analysis

PRINCIPAL INVESTIGATOR: Lance B. Price, Ph.D.

CONTRACTING ORGANIZATION: Translational Genomics Research Institute

REPORT DATE: February 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
February 2014	Final	23 Sep 2010 - 22 Sep 2013

4. TITLE AND SUBTITLE

Prospective Molecular Characterization of Burn Wound Colonization: Novel Tools and Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-0753

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Lance B. Price, Ph.D.

email: lprice@tgen.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Translational Genomics Research Institute
Phoenix, AZ 85004

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research
and Material Command
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Each year, 1.2 million people in the United States suffer burn injuries, resulting in 100,000 moderate to severe cases that require hospitalization, where infectious complications is a leading causes of death after the initial resuscitations phase. The current study is focused the temporal burn wound colonization, and how it is associated with patient outcomes. In the past year, we have made significant progress in developing novel molecular approaches for quantitative and qualitative molecular characterization of complex bacterial and fungal communities. Specifically, we have designed and validated a pan-bacterial and a pan-fungal quantitative real-time PCR assay for quantifying microbial burden directly from clinical specimens. We have also designed novel pan-fungal sequencing primers that are under laboratory and *in silico* validation. In addition, we have developed a high-throughput approach for fungal community sequencing.

15. SUBJECT TERMS

Burn wounds,Chronic wounds,Microbiota,Microbiome,Culture-independent microbial characterization,Quantitative polymerase chain reaction (qPCR),Fungi,Bacteria

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	UU	U	UU	58	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

I. INTRODUCTION

Work has proceeded successfully in this third and final grant year. A third qPCR assay has been developed to quantify extraintestinal pathogenic *E. coli* (ExPEC) in clinical samples, including those from wounds, and a manuscript has been drafted detailing the development of this assay. A set of *S. aureus* isolates from our collaborator at Duke University, which were isolated from patients with endocarditis and wound/soft tissue infections, have been sequenced and an initial analysis performed. Finally, enrollment in the hyperbaric oxygen study has been completed, as well as analysis of the wound swab samples. While the relatively small sample size and the large variation in inter-wound bacterial loads made it difficult to draw firm conclusions about the effect of HBOT on the wound microbiome, our results indicate that non-healing was associated with dominance by anaerobic bacteria and high bioburden.

II. BODY

II A. Summary of Work Performed

Development of quantitative real-time PCR assays for detecting, quantifying and characterizing fungi and bacteria in wounds

Bacterial community analysis**:** Development of a 16S rRNA gene-based universal bacterial quantification assay, entitled BactQuant, was validated in the first year of the grant using a combination of laboratory evaluation and *in silico* analysis. The assay was tested against 90 unique bacterial species. The *in silico* evaluation was performed using > 670,000 16S rRNA gene sequences, which showed significantly improved coverage compared with other published assays. As this assay was intended to be applied to clinical specimens, the assay was tested against human:bacteria DNA mixtures. The results showed that samples with high bacterial load (>1,000 copies/μL) were not likely to affected by the presence of human DNA; however, the lower limit of detection for low bacterial load samples could be affected by the presence of high amounts of human genomic DNA or by bacterial contaminants in *taq* DNA polymerase. The details of development of this assay were published in BMC Microbiology in April of 2012. See PubMed citation below:

Liu CM, Aziz M, Kachur S, Hsueh PR, Huang YT, Keim P, Price LB. **BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay**. BMC Microbiol. 2012 Apr 17;12:56. doi: 10.1186/1471-2180-12-56. PubMed PMID: 22510143; PubMed Central PMCID: PMC3464140.

Fungal community analysis**:** A broad-coverage 18S qPCR assay, entitled FungiQuant was developed in the first year of the grant for quantification of fungi in clinical samples. The assay, which targets a 351 bp region in the fungal 18S rRNA gene, is a *taq* qPCR assay. *In silico* analysis showed that FungiQuant is a perfect sequence match to 90.0% of the 2,617 fungal species that were analyzed. The assay was 100% sensitive and its amplification efficiencies ranged from 76.3% to 114.5% with r^2-values of >0.99 against the 69 fungal species tested. In addition, the assay's inter-and intra-run coefficients of variance ranged from <10% and <20%, respectively and had an upper limit of quantification of 25 copies and a lower limit of 5 copies. Results from human-only background DNA were compared with low-level fungal DNA and the results indicated that amplification in two of the three FungiQuant performed in triplicate was statistically significant for true positive fungal detection. A manuscript describing the development of this assay was published in BMC Microbiology in November, 2012. See PubMed citation below:

Liu CM, Kachur S, Dwan MG, Abraham AG, Aziz M, Hsueh PR, Huang YT, Busch JD, Lamit LJ, Gehring CA, Keim P, Price LB. **FungiQuant: a broad-coverage fungal quantitative real-time PCR assay**. BMC Microbiol. 2012 Nov 8;12:255. doi:10.1186/1471-2180-12-255. PubMed PMID: 23136846; PubMed Central PMCID: PMC3565980.

Identification of extraintestinal E. coli (ExPEC): An additional PCR assay was developed to specifically identify extraintestinal *Escherichia coli* (ExPEC) in clinical samples, including wounds. *E. coli* can be broadly classified into three major groups: commensal, diarrheagenic and ExPEC. In the developed world, ExPEC infections, including urinary tract infections (UTIs), sepsis, and neonatal meningitis, constitute a significant public health burden, but their epidemiology is not well understood. Rapid and reliable identification of ExPEC is important,

but current methods are time-consuming and can yield equivocal results. To address this, six real-time TaqMan® probe-based PCR assays were developed to screen for the following ExPEC hallmark virulence genes: *papA, papC, sfaE, afaC, kpsMII,* and *iutA*. All assays were sensitive and specific when run individually or multiplexed into two reactions. Most assays were slightly more sensitive when run on purified DNA (mean sensitivity 100%) as compared to unpurified, boiled DNA preparations (mean sensitivity 94.82%). Specificities were identical for all assays regardless of the extraction method (mean specificity 99.33%). The real-time PCR assays were generally more sensitive than previously published standard PCR assays for the same targets. This work has been detailed in a manuscript that is in the process of development and which will be submitted to BMC Microbiology: see below for the title and authors and a copy of the manuscript draft in Appendix B.

Design and validation of real-time PCR assays for the identification of extraintestinal pathogenic *E. coli* (ExPEC). Soldanova K, Liu CM, Contente-Cuomo T, Aziz M, Johnston B, Johnson JR, Buchhagen J, Moss O, Gauld L, Price LB.

Analysis of *S. aureus* isolates from endocarditis and skin/soft tissue infections

A strict blast search was performed on the *S. aureus* genomes (see attached spreadsheet with blast results). Using the attached blast results, an indicator gene analysis was performed to identify potential targets differentiating the cellulitis and endocarditis isolates. This was followed with a basic Pearson's Chi-squared test with Yates' continuity correction. The results of the analysis are summarized in the box below:

The following genes had statistically significant associations with cellulitis

agrB X-squared = 8.4, df = 1, p-value = 0.0037

agrC X-squared = 9.3, df = 1, p-value = 0.0023

agrD X-squared = 11.2, df = 1, p-value = 0.00080

SAAV_2632 X-squared = 4.1, df = 1, p-value = 0.02

The following genes had statistically significant associations with endocarditis

NWMN_1503 X-squared = 7.4, df = 1, p-value = 0.0066

set16 X-squared = 7.2, df = 1, p-value = 0.0072

SAMSHR1132_03962 X-squared = 5.5, df = 1, p-value = 0.019

Below is the code used to obtain the results summarized above:

```
> AgrB<-as.table(rbind(c(49,34), c(26, 49)))
> dimnames(AgrB) <- list(DiseaseGroup = c("Cellulitis", "Endocarditis"), Gene = c("Present","Absent"))
> (Xsq <- chisq.test(AgrB))

Pearson's Chi-squared test with Yates' continuity correction data:  AgrB
X-squared = 8.4313, df = 1, p-value = 0.003688
> Xsq$observed   # observed counts (same as BV_3tiers)

Gene
DiseaseGroup        Present       Absent
  Cellulitis          49            34
  Endocarditis        26            49
> Xsq$expected   # expected counts under the null
```

```
Gene
DiseaseGroup          Present         Absent
  Cellulitis          39.39873        43.60127
  Endocarditis        35.60127        39.39873
> Xsq$residuals  # Pearson residuals

Gene
DiseaseGroup          Present         Absent
  Cellulitis          1.529633         -1.454049
  Endocarditis        -1.609147        1.529633
> Xsq$stdres     # standardized residuals

Gene
DiseaseGroup          Present         Absent
  Cellulitis          3.063196         -3.063196
  Endocarditis        -3.063196        3.063196
> Xsq$p.value
[1] 0.003688103
>
```

MLST typing was performed using the SRST tool [1] on the whole genome sequencing data and MLST definitions were downloaded from http://pubmlst.org/data/profiles/saureus. Isolates were then divided into groups based on the assigned MLST types. Reference genomes were identified for each group and phylogenetic trees were generated using the maximum-parsimony method in PAUP v4.0b10. The references for each group are as follows:

N315 ST5 CC5 (also used for ST 15, ST 97) BA000018

Reference: Kuroda M et al, Whole Genome Sequencing of Methicillin-Resistant *Staphylococcus Aureus*. (2001) *Lancet*, 357, 1225-1240. DOI: 10.1016/S0140-6736(00)04403-2

Newman ST8 CC8 (also used for ST 25) AP009351

Reference: Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K, (2008), Genome Sequence of *Staphylococcus aureus* Strain Newman and Comparative Analysis of Staphylococcal Genomes: Polymorphism and Evolution of Two Major Pathogenicity Islands. *J Bacteriol*, 190(1), 300-310. DOI:10.1128/JB.01000-07

MW2 ST1 CC1 (also used for ST 12, ST 20) BA000033

Reference: Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L,Yamamoto K, Hiramatsu K (2002), Genome and Virulence Determinants of High Virulence Community-Acquired MRSA, *Lancet* 359, 1819 – 1827. DOI: 10.1016/S0140-6736(02)08713-5

TCH60 ST30 SLV CC30 (also used for ST 121) NC_017342 (McCarthy 2010)

ED133 ST133 CC133 NC_017337

Reference: Guinane CM, Ben Zakour NL, Tormo-Mas MA, Weinert LA, Lowder BV, Cartwright RA, Smyth DS, Smyth CJ, Lindsay JA, Gould KA, Witney A, Hinds J, Bollback JP, Rambaut A, Penades JR, Fitzgerald JR, (2010), Evolutionary genomics of Staphylococcus aureus reveals insights into the origin and molecular basis of ruminant host adaptation, *Genome Bio Evol*, 2, 454-466.Genome Bio Evol

CA 347 ST45 CC45 NC_021554

Reference: Stegger M, Driebe EM, Roe C, Lemmer D, Bowers JR, Engelthaler DM, Paul Keim, Andersen PS, (2013), Genome Sequence of *Staphylococcus aureus* Strain CA-347, a USA600 Methicillin-Resistant Isolate. *Genome Announc*, 1:e00517-13;doi:10.1128/genomeA.00517-1

[1] SRST2 - Short Read Sequence Typing for Bacterial Pathogens, http://katholt.github.io/srst/

The phylogenetic trees are provided as an attachment to this document. Those isolates beginning with an "A" were obtained from soft tissue infections. Those with ending with an "_I" are from endocarditis cases. For the majority of the trees that were developed, isolates from both soft tissue infections and endocarditis were represented in fairly equal measure. While the ST30 lineage did appear to be disproportionately associated with infectious endocarditis (based on previous analyses), no sub-MLST lineages were significantly associated with the disease (Figure 1 – see attached).

Prospective clinical sample collection from diabetic foot wounds treated with or without hyperbaric oxygen

Wound sample collection from diabetic foot wound patients seeking care at the Banner Good Samaritan Medical Center (BGSMC) Wound Clinic was completed on 6/26/13. In total, 7 participants undergoing hyperbaric oxygen treatment (HBOT) and 12 control participants were enrolled (see Appendix A for the IRB-approved collection protocol). Control participants were patients who were candidates for HBOT, but who either did not elect to continue with HBOT due to the significant time commitment or who did not receive insurance clearance for the procedure. All control participants underwent conventional wound care treatments at weekly visits, which included control of blood glucose, application of interventions to relieve pressure on the wound, clinical debridement of necrotic or macerated tissues and application of dressings with or without antimicrobials to prevent or treat infection. Subjects in the HBOT group received conventional wound care plus HBOT 5 days per week.

Summary of Participant Demographic and Clinical Data: Table 1 (next page) is a summary of the demographic and clinical data collected from each subject. Twelve Control subjects and seven HBOT subjects were enrolled in the study. One subject in each group completed only one study visit; thus both subjects were excluded from further analysis, as we were unable to assess wound healing. Both groups were similar in terms of age, gender distribution, the initial number of baseline abnormal baseline transcutaneous oximetry measurements (TCOM), which to determine pre-treatment lower limb perfusion and oral and topical antibiotic use. The average starting wound size for the HBOT group was larger than the Control group; however, there was also a great deal more variability in wound size in the HBOT group as one HBOT subject had the largest wound in the study at 72.9 cm^2. A greater percentage of subjects in the HBOT group had decreases in wound area at the end of the study (67% vs 54%, respectively); however, two subjects in each group were removed from the study for surgical treatment before completing the study (2 wound closure surgeries in the Control group and 2 amputations in the HBOT group). This included the HBOT subject with the largest starting wound area. Although this subject's wound decreased in area, he developed a recalcitrant infection requiring amputation.

Table 1. Summary of Demographic and Clinical Characteristics of Study Participants.

Characteristics	Control Group	HBOT Group
Age (SD)	57.3 (8.1)	65.0 (13.8)
Gender	Male: N = 9 (82%) Female: N = 2 (18%)	Male: N = 5 (83%) Female: N = 1 (17%)
DM Type	Type II: N = 10 Type I: N = 1	Type II: N = 7
Average baseline abnormal TCOM readings (out of 6)	4.5	4.5
Average starting wound area - cm^2 (SD)	4.1 (4.8)	22.1 (30.2)
Average ending wound area - cm^2 (SD)	5.0 (8.1)	9.2 (13.2)
Decreased wound area	N = 7 (64%)	N = 5 (83%)
Topical antibiotic use	N = 10 (91%)	N = 6 (100%)
Oral antibiotic use	N = 5 (45%)	N = 2 (33%)
Subjects completing study	N = 9 (75%)	N = 4 (57%)
Reasons off study	Insurance change: N = 1, Wound closure surgery: N = 2	Subject withdrew: N = 1, Amputation: N = 2

Bacterial load quantification: Swabs containing a DNA/RNA stabilizing transport media were used to obtain samples from the center, proximal and distal aspects of the wound bed during weekly debridement. While one set of samples was collected during debridement for Control subjects, HBOT subjects had two collection events: pre-debridement and post-HBOT. Samples were frozen at -80^{o}C and shipped to TGen on dry ice.

Swab samples were processed for bacterial load quantification, which was measured as the number of bacterial 16S gene copies per μl of swab eluent using the Bactquant broad coverage PCR assay described earlier in this report. For each sample, 100 μl of eluted transport medium was lysed using enzyme-free chemical and mechanical lysis. The lysate was purified using a Qiagen AllPrep DNA/RNA minikit and DNA was eluted with 100 μl of buffer. Although processed separately, the data from samples collected at different wound positions (center, distal and proximal) were pooled for all subsequent analyses to provide a more

complete picture of the entire wound microbiome. Below are plots of bacterial load calculations over time for each subject (Figures 2 – 5). The plots are arranged by group (control vs HBOT) and by percent change in wound size over the course of the study, as measured by wound area in cm^2 (wound increase vs wound decrease). For the HBOT subjects, bacterial loads for both pre- and post-HBOT samples are plotted.

The bacterial loads showed a significant variation between subjects. Table 2 shows a summary of the bacterial load measurements, in which the average load and standard deviation have been calculated for the baseline (first study visit) and final sampling (final study visit) for each study group, as well as the pre- and post-HBOT samplings. The average difference in load between the baseline and final samplings was also calculated. While there was a great deal of variability overall between samples, the HBOT group, both pre- and post-treatment, had higher average bacterial loads than Control samples. This difference can in large part be attributed to the two HBOT subjects who received amputations, as both subjects had extremely high baseline (2,899.9 and 227.8 16S rRNA gene copies/ mL, respectively) and final (4,667.7 and 1,065.5 copies/ mL) bacterial loads.

Bacterial loads of Control and HBOT subjects who required surgical intervention during the study were compared with those subjects in each group who did not require surgery. Not surprisingly, the two HBOT subjects requiring amputation had the highest loads of the study – both at baseline and at the final sampling. Interestingly, the two Control subjects requiring surgical wound closure did not have loads that were appreciably different than non-surgical subjects in either the Control or HBOT groups.

Figure 2. Bacterial loads over time for Control subjects with **decreases** in wound area

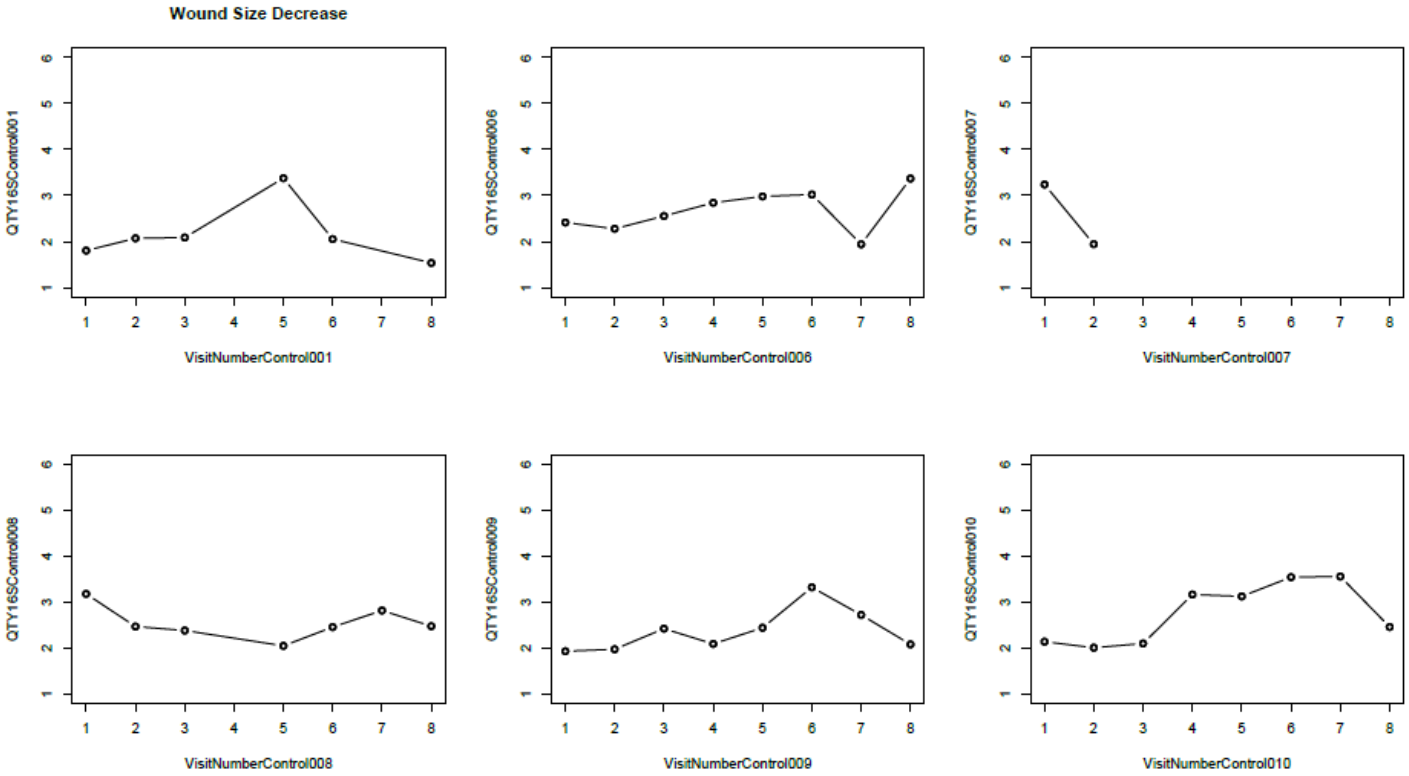


Figure 3 Bacterial loads over time for HBOT subjects with **decreases** in wound area

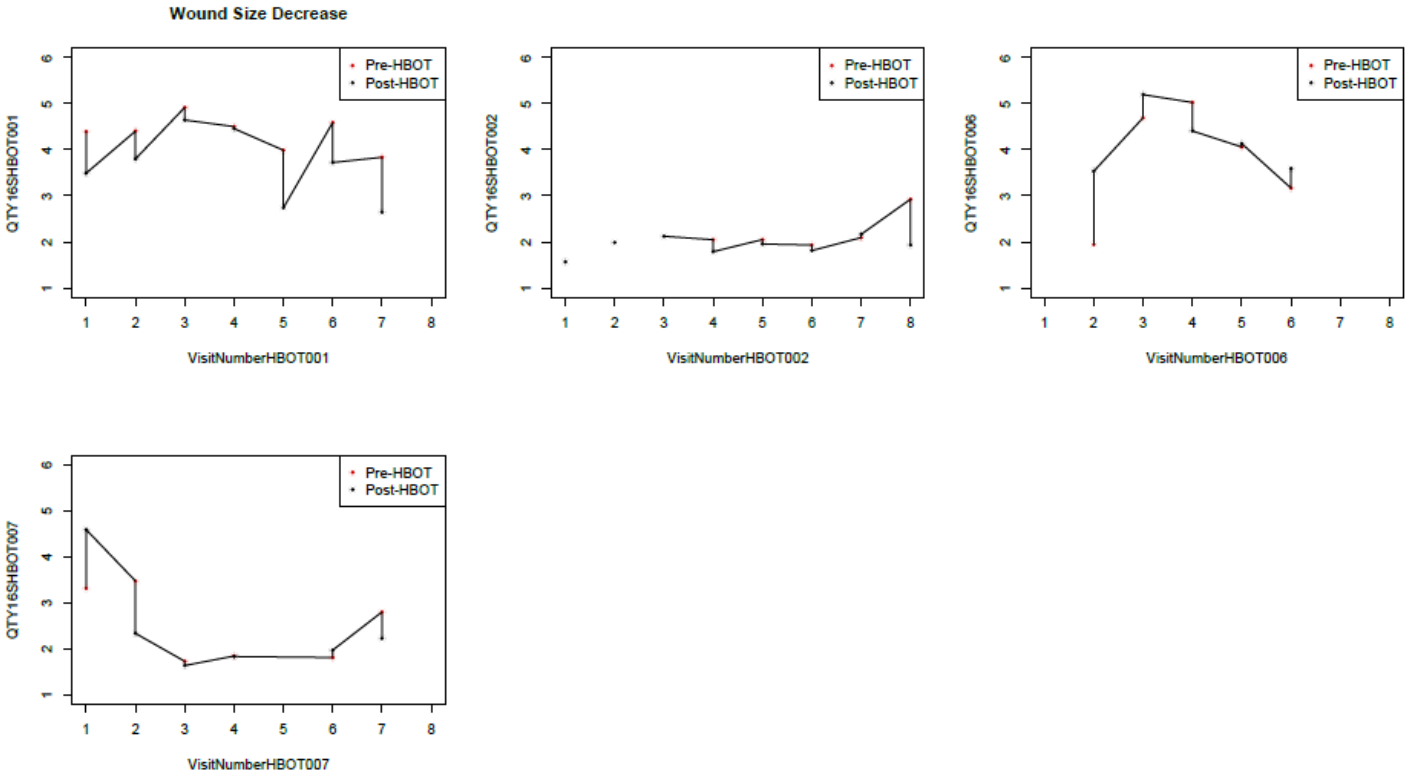


Figure 4. Bacterial loads over time for Control subjects with **increases** in wound area

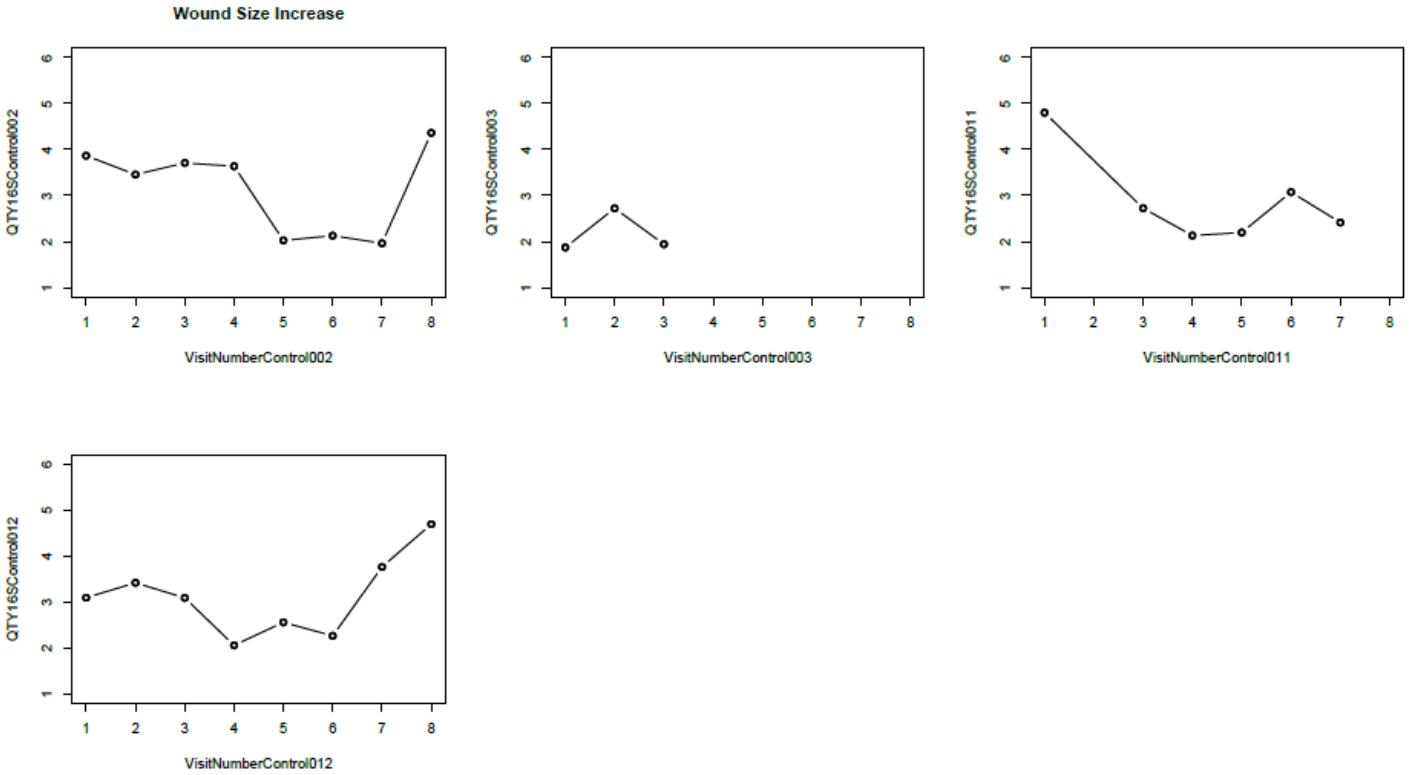


Figure 5. Bacterial loads over time for HBOT subjects with **increases** in wound area

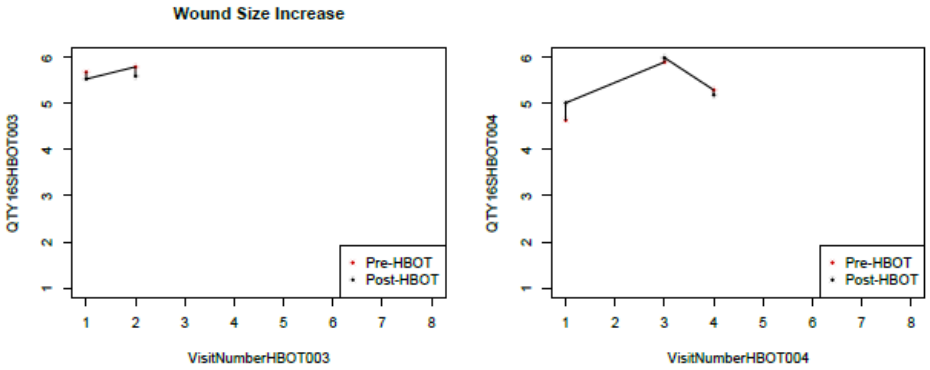


Table 2: Summary of Bacterial Load Calculations				
	Baseline load (BL)		Final load (FL)	Load difference: (FL-BL)
	N	Average (SD)	Average (SD)	Average (SD)
Control	11	28.1 (55.0)	13.0 (9.6)	-15.1 (57.2)
HBOT pre-treatment	5	637.3 (1268.1)	1202.4 (1984.9)	565.1 (754.6)
HBOT post-treatment	6	617.9 (1277.7)	933.0 (1453.6)	418.1 (1249.9)
Control non-surgical	9	32.4 (60.5)	11.8 (9.2)	-20.6 (62.4)
Control surgical	2	8.7 (1.6)	18.3 (13.6)	9.5 (12.0)
HBOT non-surgical	3	19.6 (18.6)	93.0 (140.5)	73.4 (-125.4)
HBOT surgical	2	1563.9 (1889.4)	2866 (2574.1)	1302.7 (657.7)

16S rRNA gene-based pyrosequencing analysis: Bar-coded V3-V6 amplicons were generated using broad-coverage fusion PCR primers, which were pooled and sequenced on the Roche 454 platform. Pyrosequence data was processed using an in-house pipeline (PyroDB) and taxonomic classification was performed using the Ribosomal Database Project Naïve Baysian Classifier. Dufrene & Lefendre indicator analysis was used to distinguish the bacterial genera associated with wound increase or decrease.

Indicator analysis: Table 3 includes a list of indicator bacteria associated with either wound area increase (Group 1) or wound decrease (Group 2). Wounds that increased in size (Group 1) were distinguished by anaerobic genera such as *Peptoniphilus*, *Pavimonas*, *Peptostreptococcus* and *Fastidiosiphila* and *Fusobacterium* (p = 0.0068). Conversely, wounds that decreased in area were associated with aerobic or facultative anaerobic genera, such as *Staphylococcus* (p = 0.0068), *Enterobacter* (p = 0.0135), *Shimwellia* (p = 0.0162), and *Cornebacterium* (p = 0.0231). Species of *Staphylococcus*, such as *S. epidermidis* are ubiquitous

on normal skin and *Enterobacter* members can frequently be found as fecal contaminants on skin, suggesting that a wound microbiome that is closer to normal skin is associated with wound healing/decrease in area.

Table 3. Indicator Analyses for Genera Significantly Associated with Wound Increase and Decrease

Genus	Group (1 = wound increase, 2 = wound decrease)	Indicator Value	p-value	False-discovery rate adjusted p-value
Peptoniphilus	1	0.54	0.0010	0.0068
Parvimonas	1	0.47	0.0010	0.0068
Peptostreptococcus	1	0.38	0.0010	0.0068
Fastidiosipila	1	0.37	0.0010	0.0068
Fusobacterium	1	0.37	0.0010	0.0068
Gemella	1	0.36	0.0010	0.0068
Dialister	1	0.27	0.0010	0.0068
Catonella	1	0.25	0.0010	0.0068
Prevotellaceae.C.0.8	1	0.22	0.0010	0.0068
Gallicola.C.0.8	1	0.22	0.0010	0.0068
Proteus	1	0.19	0.0010	0.0068
Clostridiales_Incertae. Sedis.XI.C.0.8	1	0.41	0.0020	0.0108
Clostridiales.Incertae. Sedis.XI.C.0.8	1	0.37	0.0020	0.0108
Solobacterium	1	0.22	0.0020	0.0108
Peptoniphilus.C.0.8	1	0.36	0.0030	0.0135
Porphyromonas	1	0.32	0.0030	0.0135
Salmonella	1	0.29	0.0040	0.0162
Staphylococcus	2	0.71	0.0010	0.0068
Enterobacter	2	0.39	0.0030	0.0135
Shimwellia.C.0.8	2	0.41	0.0040	0.0162
Corynebacterium	2	0.60	0.0060	0.0231
Staphylococcus.1	2	0.67	0.0070	0.0258
Klebsiella.C.0.8	2	0.41	0.0100	0.0352

Heat Map Analysis. Figures 6 and 7 are heat maps that represent the amount of bacteria present in each of the Control (Figure 6) and HBOT (Figure 7) samples. The shaded legend on the right of each figure indicates the color scheme for the number of 16S rRNA gene pyrosequences for each genera present in each sample. The samples are organized by wound increase (samples on the left – see top black bar) and wound decrease (top green bar). For the HBOT samples, the identifier “DSW” signifies a pre-treatment sample, while “PSW” is

post-treatment. While there was a great deal of diversity in bacterial genera between subjects in both groups, the pattern of genera present in each wound did not appreciably change over time. In addition, pre- and post-treatment swabs for HBOT subjects were quite similar in composition, which could be explained by the fact that DNA from both live and dead bacteria would be sequenced.

Figure 6: Heat Map – Control Group

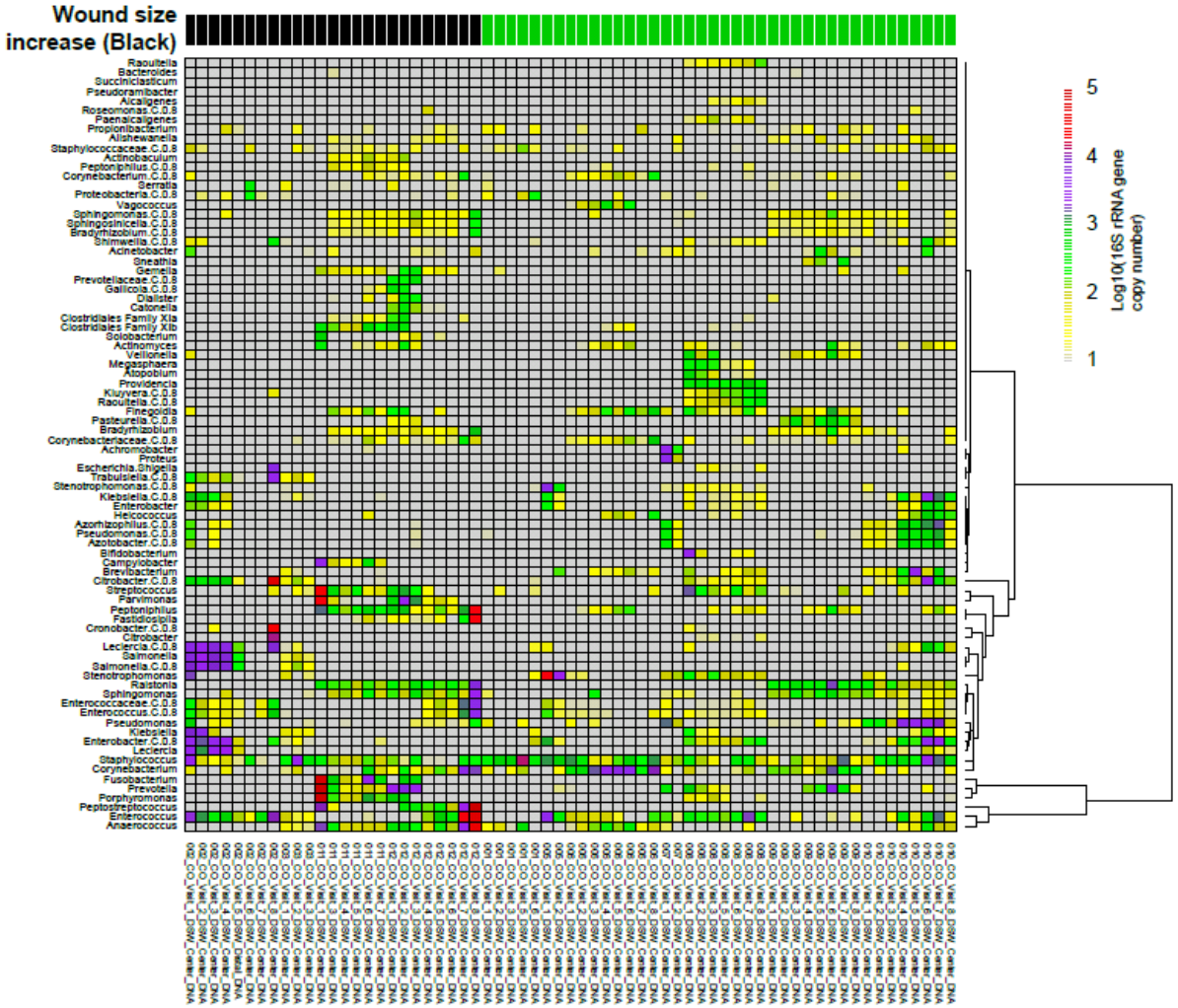
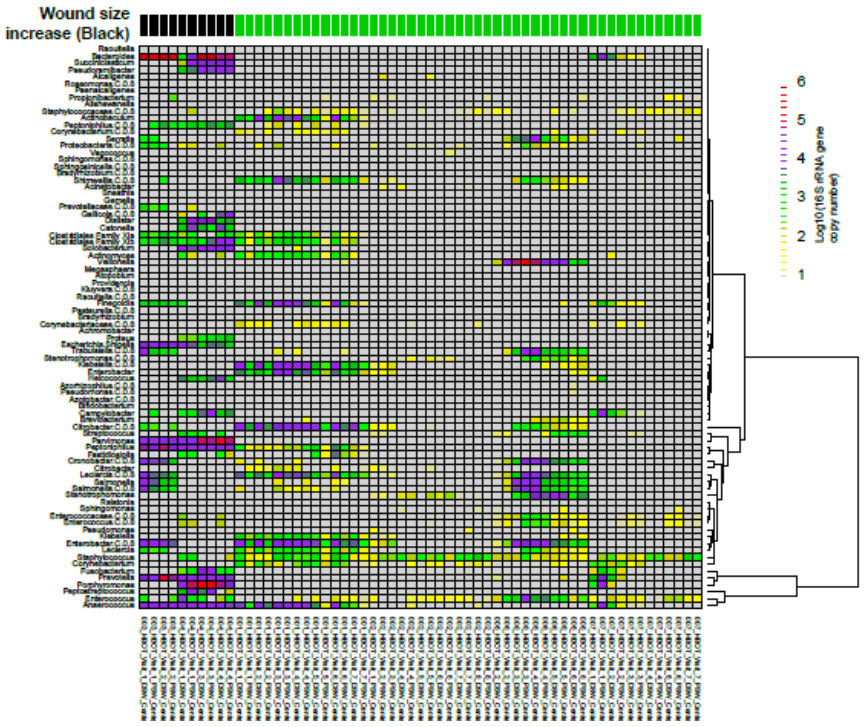


Figure 7: Heat Map – HBOT Group



III. KEY RESEARCH ACCOMPLISHMENTS

The development of the BacQuant and FungiQuant qPCR assays has provided us with an invaluable resource for quantification of bacterial and fungal loads in a wide variety of clinical samples. Thus, these assays may be useful in future wound studies as well as in a variety of other clinical specimens. A third qPCR assay has been developed which focuses on the discrimination ExPEC in clinical specimens, by assaying for six hallmark ExPEC virulence genes. Similarly, the ExPEC Hallmark assay has the potential to be used in a broad range of clinical specimen types, as well as in wound samples. The development of this multiplex assay has been detailed in a manuscript, which will be submitted to BMC Microbiology.

Finally, *S. aureus* samples collected from endocarditis and soft tissue/wound infections have been subjected to whole genome sequencing and four genes have been identified that are significantly associated with cellulitis (agrB, agrC, agrD and SAAV_2632) and three different genes that are associated with endocarditis (NWMN_1503, set16, SAMSHR1132_03962). These data suggest that there are identifiable differences in *S. aureus* strains found in wounds vs. those found in other infections, such as endocarditis.

The prospective enrollment of subjects in the Hyperbaric Oxygen study at Banner Good Samaritan Medical Center Wound Clinic proceeded smoothly, with 7 subjects undergoing HBOT enrolled and 12 control subjects enrolled. Although enrollment was less than the expected 25 HBOT patients and 25 controls, the lower enrollment numbers were a function of a decreased number of eligible patients seen in the clinic during the enrollment period, rather than ineffective screening or consent processes. Despite the low enrollment, the our results showed that non-healing was associated with a dominance by anaerobic bacterial species and high bioburden, suggesting a potential mechanism for the effectiveness of HBOT as an adjuvant therapy in diabetic wound treatment.

IV. REPORTABLE OUTCOMES

Liu CM, Aziz M, Kachur S, Hsueh PR, Huang YT, Keim P, Price LB. **BactQuant: an enhanced broad-coverage bacterial quantitative real-time PCR assay**. BMC Microbiol. 2012 Apr 17;12:56. doi: 10.1186/1471-2180-12-56. PubMed PMID: 22510143; PubMed Central PMCID: PMC3464140.

Liu CM, Kachur S, Dwan MG, Abraham AG, Aziz M, Hsueh PR, Huang YT, Busch JD, Lamit LJ, Gehring CA, Keim P, Price LB. **FungiQuant: a broad-coverage fungal quantitative real-time PCR assay**. BMC Microbiol. 2012 Nov 8;12:255. doi:10.1186/1471-2180-12-255. PubMed PMID: 23136846; PubMed Central PMCID: PMC3565980.

Manuscript in preparation: to be submitted to BMC Microbiology:

Design and validation of real-time PCR assays for the identification of extraintestinal pathogenic *E. coli* (ExPEC). Soldanova K, Liu CM, Contente-Cuomo T, Aziz M, Johnston B, Johnson JR, Buchhagen J, Moss O, Gauld L, Price LB.

A manuscript including the HBOT study data is also in preparation.

V. CONCLUSION

This project has been extremely productive in terms of developing highly specific assay tools for bacterial and fungal load analysis that may be used in a wide range of other projects (both wound-related and unrelated) for the assessment of bioburden. While our sequencing analyses of the *S. aureus* isolates did not provide a clear delineation between isolates from soft tissue infections and endocarditis, a more in-depth analysis of the data is planned. Finally, we successfully recruited 19 subjects into our prospective HBOT study. While subject numbers were less than expected, due to a low number of HBOT candidates presenting to the clinic, the data showed that non-healing of diabetic foot wounds was associated with the presence of anaerobic species and high bacterial load.

VI. APPENDICES

APPENDIX A. IRB-Approved Clinical Sample Collection Protocol

TRANSLATIONAL GENOMICS RESEARCH INSTITUTE

HUMAN SUBJECT PROTOCOL

1. General Protocol Information

Investigators:

Lance Price, PhD
Cindy Liu, MD, MPH
Translational Genomics Research Institute (TGen) – Awardee institution
Assurance #: FWA00003918
Assurance expiration date: 8/4/2012
IRB of record for TGen study analysis: Western IRB (WIRB)
Pathogen Genomics Division
3051 West Shamrell Boulevard
Suite 106
Flagstaff, AZ 86001
Phone (Dr. Price): 928-226-6371
Fax: 928-226-6360
E-mail: lprice@tgen.org

Yadwinder Dhillon, MD
Glenn Silverstein, DPM
Banner Good Samaritan Medical Center (BGSMC)
Wound Care and Hyperbaric Medicine Center
Assurance #: FWA00002630
Assurance expiration date: 5/19/2013
IRB of record for subject recruitment and clinical procedures: BGSMC IRB
1012 East Willetta Street
Phoenix, AZ 85006

Clinical Research Coordinator:
Lora Nordstrom, PhD, BSN, CCRC (TGen)

Sponsor: Department of Defense (DOD) Telemedicine and Advanced Technology Research Center (TATRC)

Protocol Title: The Impact of Hyperbaric Oxygen Therapy on the Wound Microbiome, Wound Host-Microbe Interactions, and Patient Outcomes

Protocol Number: lprice11-023

2. Subject Issues.

(a) Number and brief description of subjects to be enrolled at TGen: No subjects will be enrolled at TGen. All subjects will be enrolled by clinical collaborators at Banner Good Samaritan Medical

Center (BGSMC).

(b) Number and brief description of subjects to be enrolled elsewhere: Patients with diabetic foot ulcers who are receiving care for their wounds as outpatients at the Wound Care Program at BGSMC will be screened, consented, and enrolled in this study. Up to 100 subjects will be enrolled in this study. It is expected that a certain percentage of participants will not complete the study due to missed study visits; thus, up to 100 participants will be enrolled in order to accrue 25 subjects actively receiving **both** standard-of-care wound care and adjunctive hyperbaric oxygen therapy and 25 diagnosis- and sex-matched control subjects, who are receiving standard-of-care wound care, but not adjunctive hyperbaric oxygen therapy.

The sample size is based on typical enrollment rates and the number of patients that would meet enrollment criteria within the Wound Care Program over the time frame of the study. Since this is the first study of its kind, a true statistical sample justification could not be determined.

☒ **Adults** ❑ **Children** ☒ **Males** ☒ **Females** **Age Range:** 18-88 years

❑ **Yes** ☒ **No** **Do TGen Personnel Interact with Living Individuals?**

☒ **Yes** ❑ **No** **Do Collaborating Investigators Interact with Living Individuals?**

☒ **Yes** ❑ **No** **Do Collaborating Investigators Provide Human Biospecimens?**

❑ **Yes** ☒ **No** **Do Collaborating Investigators Provide Information about Individuals?**

❑ **Yes** ☒ **No** **Is any genetic testing for clinical proposes involved?**

Please explain any "YES" responses. All wound biosamples and annotated clinical information will be collected by clinical collaborators at BGSMC. All biosamples and clinical data will be completely de-identified and assigned a study-specific code number. The key linking the study code to identifiable subject information will be kept by the clinical PI at BGSMC on a password-protected database. TGen investigators will be prohibited from accessing or attempting to gain access to the code.

3. Radiation or Bio-Safety Considerations

DESCRIBE ANY RADIATION USE OR BIO-SAFETY ISSUES SUCH AS USE OF INFECTIOUS AGENTS, REGULATED TOXINS, RECOMBINANT DNA, HUMAN GENE TRANSFER, ETC

No radiation will be used in this study. Standard Universal Bio-safety precautions will be taken when handling the human samples.

4. Regulatory Criteria for IRB Approval of Research.

FEDERAL REGULATIONS AT 21 CFR 56.111 AND 45 CFR 46.111 REQUIRE THE IRB TO DETERMINE THAT THE FOLLOWING CRITERIA ARE SATISFIED.

4.1 Risks are Minimized Through Sound Research Design

(a) Briefly state the hypothesis or the objectives of the proposed research.
(b) Briefly describe the background of the research, citing scientific literature as appropriate.
(c) Fully describe the research design and procedures (including statistical design).
(d) Briefly describe the research team's qualifications for performing the proposed research.

(a) Objectives.

Our overall objectives are to:
1) Elucidate the impact of hyperbaric therapy on the microbial ecology, host immunity, and host-microbe interactions in diabetic foot ulcers;
2) Determine the host immunological and microbial predictors of foot ulcer outcomes in diabetic patients.

Hypothesis: Hyperbaric oxygen therapy results in decreased absolute and relative abundance of anaerobic bacterial species in diabetic foot ulcers in comparison to standard of care diabetic foot ulcer management.

(b) Background.

Chronic wounds are wounds that require a prolonged time to heal, do not heal, or recur; and it is a major public health in the United States and globally. It has been estimated that 1% of the industrialized countries will experience a leg ulcer at some time. Broadly, chronic wounds encompass four major categories, which are diabetic foot ulcers, venous ulcers (i.e., associated with venous insufficiency), arterial ulcers (i.e., associated with limb ischemia), and pressure ulcers. In diabetes mellitus, the development of foot ulcers can usually be ascribed to peripheral neuropathy and/or peripheral vascular disease, which can be venous or arterial. The annual incidence of foot ulcers among people with diabetes have been estimated at 2.5-10.7%, with an annual incidence of amputation at 0.25-1.8%. It has also been estimated that 50% of diabetic foot ulcers become infected at some point, with 25% of the infected foot ulcers resulting in lower limb amputation, making wound infection the most important risk factor for amputation in

To address this challenging clinical group of disease, we need to better understand the pathophysiology underlying the chronicity and recurrence of chronic wounds. One of the major hypotheses of delayed wound healing in chronic wounds is hypoxia, which has led to the investigation of adjunctive hyperbaric oxygen therapy for chronic wound patients. Hyperbaric oxygen therapy is a non-invasive therapy that uses 100% oxygen under increased atmospheric pressure in a controlled body chamber. Another important knowledge gap is the role of the wound microbiome the resultant host-microbe interactions. While there have been several studies that have shown the potential benefit of adjunctive hyperbaric oxygen therapy in chronic wound patients, particularly in diabetic foot ulcers for reducing major limb amputations, there has been no study that investigates the impact of hyperbaric oxygen therapy on the wound microbiome and the wound host-microbe interactions in patients with diabetic foot ulcers. Therefore, in this collaborative study, we will perform a prospective, observational study using cutting-edge molecular genomic, transcriptomic, and proteomic techniques to study the molecular impact of hyperbaric oxygen therapy in patients with diabetic foot ulcers.

(c) Design and Procedures

Identification of Participants: Two groups participants will be identified:

- **Treatment Group:** Diabetic foot ulcer patients actively receiving both standard-of-care and adjunctive hyperbaric oxygen therapy for their wounds.

- **Control Group:** Diabetic foot ulcer patients receiving the standard-of-care wound treatment but who are **not** receiving adjunctive hyperbaric oxygen therapy.

Collection of Biospecimens: Biospecimens from wounds will be collected via swab and/or curette during routinely scheduled clinical visits to minimize any discomfort for the subjects. The most difficult-to-heal site will be selected and will be recorded and photographed at each sample collection. The site will be cleansed to remove surface dirt using a gauze pad saturated with sterile normal saline before sample collection. All swab and curette samples will be collected from the apposing leading edges of each wound (proximal and distal), as well as the center of each wound, if the wound is of sufficient size to obtain discrete samples from each area. With smaller wounds (those less than 2 cm in diameter) only a single curette and swab sample will be collected from the center of the wound. Swab samples will be collected using a swab pre-moistened in sterile saline. NOTE: swab samples should be taken from the wound base after curette debridement for the weekly debridement visits.

Curette samples will be collected from wounds until the treating physician makes a determination that collection of a curette sample would be detrimental to the wound healing process. Swab samples will be collected until the wound has healed completely.

Routine Sample Collection

- **Treatment Group**: Patients receiving adjunctive hyperbaric oxygen therapy have regularly scheduled clinic visits 5 days per week for approximately 6 weeks. As the standard of care for these wounds involves weekly debridement, samples will be collected at weekly debridement visits. After normal saline cleansing, both swab and curette samples will be collected during the debridement procedure, as well as after the adjunctive hyperbaric oxygen therapy. Thus, two sets of swab and curette samples will be collected during the debridement visits. The swab and curette samples will be collected at the same location on the wound (distal, proximal and center or center only, if the wound is small), with curette samples collected first, followed by the swab samples. All swab and curette samples will be immediately flash frozen on dry ice or in liquid nitrogen and stored at -80°C at BGSMC, until they can be shipped to TGen for analysis.

- **Control Group**: Patients who do not receive adjunctive hyperbaric oxygen therapy are scheduled for weekly clinic visits, at which time their wounds are debrided. Patients not receiving hyperbaric treatment usually have 7 – 8 weekly clinic visits. Matching swab and curette samples will be collected at the same locations (distal, proximal, center or center only, if the wound is small) during the weekly clinic visits from subjects in this group. The curette samples will be collected first, followed by the swab samples. All swab and curette samples will be immediately flash frozen on dry ice or liquid nitrogen and stored at -80°C at BGSMC, until they can be shipped to TGen for analysis.

  If a participant in the control group misses more than 2 regularly scheduled clinic appointments in a row, he or she will be withdrawn from the study, as his/her wound healing will likely be affected by the cessation of treatment.

Schedule of Sample Collection:

- **Treatment Group**

  Study visit 1:
  - Participant will be enrolled in the study after a thorough informed consent discussion
  - Initial clinical data will be collected
  - A photograph will be taken of the wound from which samples will be collected
  - The wound site will be cleansed with normal saline
  - Three curette samples will be collected during the debridement procedure, from the center, proximal and distal edges of the wound. **Note:** three samples will be collected for wounds greater than or equal to 2 cm in diameter. For wounds less than 2 cm in

diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center, proximal and distal edges of the wound.
    - The participant will undergo standard of care hyperbaric therapy
    - Three curette samples will be collected after the hyperbaric treatment, from the center and proximal and distal edges of the wound. **Note:** three samples collected for wounds greater than or equal to 2 cm in diameter. For wound less than 2 cm in diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center and proximal and distal edges of the wound.

  Subsequent weekly study visits:
    - Follow-up clinical data will be collected
    - A photograph will be taken of the wound from which samples will be collected
    - The wound site will be cleansed with normal saline
    - Three curette samples will be collected during the debridement procedure, from the center, proximal and distal edges of the wound. **Note:** three samples will be collected for wounds greater than or equal to 2 cm in diameter. For wounds less than 2 cm in diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center and proximal and distal edges of the wound.
    - The participant will undergo standard of care hyperbaric therapy
    - The wound site will be cleansed with normal saline
    - Three curette samples will be collected after the hyperbaric treatment, from the center, proximal and distal edges of the wound. **Note:** three samples will be collected for wounds greater than or equal to 2 cm in diameter. For wounds less than 2 cm in diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center, proximal and distal edges of the wound.

- **Control Group**
  Study visit 1:
    - Participant will be enrolled in the study
    - Initial clinical data will be collected
    - A photograph will be taken of the wound from which samples will be collected
    - The wound site will be cleansed with normal saline
    - Three curette samples will be collected during the debridement procedure, from the center, proximal and distal edges of the wound. **Note:** three samples will be collected for wounds greater than or equal to 2 cm in diameter. For wounds less than 2 cm in diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center, proximal and distal edges of the wound (see Note above).

  Subsequent weekly study visits:
    - Follow-up clinical data will be collected
    - A photograph will be taken of the wound from which samples will be collected
    - The wound site will be cleansed with normal saline
    - Three curette samples will be collected during the debridement procedure, from the center, proximal and distal edges of the wound. **Note:** three samples will be collected for wounds greater than or equal to 2 cm in diameter. For wounds less than 2 cm in diameter, single curette and swab samples will be collected.
    - Collection of curette samples will be followed by collection of three swab samples from the center, proximal and distal edges of the wound (see Note above).

Collection of Clinical Data: In addition to biospecimens, annotated clinical information will be

obtained from each study participant (see attached data collection forms). Briefly, basic demographic data will be collected as well as medical history of the wound, including: history of wound care, medications and treatments received, and co-morbid conditions. In addition, data from transcutaneous oximetry measurements (TCOM), which will be provided as part of standard of care therapy for all study participants, will be collected.

All diabetic foot ulcer patients who receive adjunctive hyperbaric oxygen therapy are tested for blood glucose levels before and after treatment. This data will be collected for study purposes. Diabetic patients not receiving hyperbaric oxygen therapy do not routinely blood glucose testing performed at their clinic visits; thus, a research-specific blood glucose reading, collected via peripheral finger stick, will be performed at weekly sample collection visits, using a glucose monitor at the clinic (not the patient's personal glucose monitor). Hemoglobin A1C (HgbA1C) values will be collected from participants when performed as part of their standard of care treatment.

No PHI will be collected for this study. All clinical data will receive a bar-coded study number, which matches the collected biosamples from each subject. The key linking the study code to identifiable subject information will be kept by the clinical PI or delegee on a password-protected database. No TGen employees may access or attempt to gain access to this key.

Laboratory Methods:

Once samples are transferred to the Pathogen Genomics Division at TGen, they will be processed and analyzed as follows:

1) Isolate and purify total DNA, RNA, and Protein from each sample
2) Analyze the microbial community composition by sequencing analysis
    a. 16S targeted approach for community-level bacterial characterization
        i. Amplify the 16S rRNA gene or cDNA generated from16S rRNA using fusion PCR primers.
        ii. Pool the barcoded amplicon for sequencing
        iii. Sequence the amplified 16S rRNA gene region.
        iv. Process the resultant sequences using an in-house data processing pipeline to assay each sequence to its sample source.
        v. Assign taxonomic classification to each processed sequence.
        vi. Generate community matrix data for comparative ecological analyses including: rarefaction, species accumulation curve, diversity calculations, and other ecological community visualization and analyses.
        vii. Sequences that could not be classified taxonomically will be analyzed phylogenetically and full-length 16S rRNA gene sequence analysis may also be performed.
    b. Non-targeted metagenomic approach
        i. Shear total DNA or RNA content
        ii. Separate human DNA or eukaryotic RNA from the microbial contingent by subtraction technique
        iii. Ligate barcoded sequencing adapter to sheared DNA or RNA
        iv. Sequence the resultant DNA or RNA shotgun library using next-generation sequencing technology.
        v. The resultant sequences will be analyzed using a combination of BLAST, MEGAN, and other computational tools to elucidate the DNA and RNA metagenome.
    c. Non-targeted RNA-based transcriptome approach
        i. Separate human RNA from bacterial RNA
        ii. Isolate RNA from each component
        iii. Create human and microbial cDNA mRNA libraries
        iv. Sequence transcriptome on next-generation sequencing platforms.

d. Non-targeted Proteomics approach
    i. Perform lysis and extraction for proteins from each sample.
    ii. Perform additional enzymatic digestions to isolate desired protein contingents
    iii. Proteome sequencing and analysis

3) Quantify specific target organisms using species/strain specific quantitative PCR assays.
    a. Measure associations between quantities of specific pathogens and wound healing.

4) Measure diabetes-related inflammatory and wound healing proteins and gene expression among the subjects
    a. Measure associations among proteins, gene expression, wound microflora and wound healing.
    b. Expression will be assessed by microarray analysis, mass spectrometer, sequencing, qPCR and/or ELISA.
    c. Host genotyping

4.2 Risks Are Reasonable in Relation to Anticipated Benefits

(a) Fully describe the reasonably foreseeable (physical, psychological, social) risks, side effects, and discomforts to the subject of the proposed research.
(b) Describe steps taken to minimize these risks (e.g., instituting specific protections, choosing specific techniques, or relying on procedures already being performed for other purposes).
(c) Describe the reasonably anticipated benefits of the research to subjects and the importance of the knowledge that may be reasonably expected to result.
(d) Demonstrate that the risks are reasonable in relation to these benefits and/or resultant knowledge.
(e) Justify the use of a placebo control group if one is included in the proposed research.

(a) Risks. There is a slight risk of pain, infection and/or bleeding involved in taking samples from open wounds. Aseptic technique will be used at all times during the collection process and only study staff who are trained in collection of samples from wounds will be involved in collection procedures. All sample collections from wounds will be done at the same time as normal clinical procedures/assessments. Analgesis will be administered as needed during sample collection. The Wound Care Clinic utilizes the following options for analgesia:

- Topical 4% Lidocaine solution
- Topical 5% Lidocaine gel/ointment
- Injected 1% and 2% Lidocaine plain
- Injected 0.5% Marcaine plain

Possible side effects of these medicines include: skin irritation, fatigue, weakness, dizziness, blurred vision, numbness, and or tingling.

The decision to use hyperbaric therapy, in addition to standard of care treatment, for diabetic foot ulcers is entirely dependent on the study participants' decisions of whether or not to have hyperbaric treatment. This decision is made in consultation with the participant's physician. There are some known risks involved in this treatment. The most common side effect is barotrauma to the ears and sinuses due to the change in pressure. To minimize this effect, the hyperbaric technicians will work with the patients to show them techniques for ear-clearing. A rare side effect is oxygen toxicity. All patients will be continuously monitored during the procedure by wound care staff and are given "air breaks" at scheduled times during the procedure to prevent oxygen toxicity.

In addition to the physical risks outlined above, there may be some emotional anxiety for subjects, regarding the fact that subjects will not directly benefit from this study, nor will they receive any

results from the analysis of there samples.

(b) Risks Minimized. As described above, every effort will be made to make certain that sample collection uses aseptic technique and occurs at the same time as clinical procedures, to decrease any discomfort associated with the procedure. Hemostasis will be ensured after each sample collection. No additional samples will be collected from subjects exhibiting a coagulopathy. If any subject requires immediate medical care related to a study-related procedure, emergency medical care will be provided at the hospital. No subjects will receive compensation for a research-related injury.

To minimize any emotional discomfort, the voluntary nature of the study will be stressed at all times and measures to keep study data confidential will be communicated to study subjects.

(c) Benefits. There will be no direct benefit to subjects participating in this study. The subjects may, however: feel good about participating in research that may help wound patients in the future.

(d) Risks Reasonable. The physical risks outlined above may be lessened to a great degree with meticulous sample collection techniques. Given these facts, the risks are minimal compared with the potential to benefit subjects in the future with wounds in terms of improved prognostics and treatments.

(e) Placebo. There will be no placebo group for this study.

4.3 Selection of Subjects is Equitable

(a) Describe the procedures for recruitment of subjects, including any advertising for subjects.
(b) Describe the inclusion and exclusion criteria for each subject population. Provide justification for excluding any subject populations or groups from the research.
(c) Describe how subjects will be assigned to experimental and control groups.
(d) Describe any compensation or other inducements that will be offered to subjects for participating in the research, and the schedule/procedure for conveying such inducements.
(e) List any costs to the subject associated with participation in the research.
(f) Describe any compensation or other inducements that will be offered to <u>investigators</u> for conducting the research, and the schedule/procedure for conveying such inducements

(a) Recruitment. Potential subjects will be recruited for this study when they present for care at the BGSMC Wound Care Clinic. Subjects will be approached by the clinical PI or sub-investigator and a detailed informed consent discussion will take place using the informed consent form as a guide for the discussion. No advertising will be used to recruit subjects for this study.

(b) Inclusion/Exclusion Criteria.

<u>Treatment Group</u>

Inclusion Criteria:

- Eighteen years of age or older
- Has been diagnosed with diabetic foot ulcer
- Has diabetes mellitus
- Patient is a candidate for adjunctive hyperbaric oxygen therapy and will receive adjunctive hyperbaric oxygen therapy
- Patient has received TCOM as part of standard of care treatment.

Exclusion Criteria:

- Prisoners
- Pregnant women

Control Group

Inclusion Criteria:

- Eighteen years of age or older
- Has been diagnosed with diabetic foot ulcer
- Has diabetes mellitus
- Patient is a candidate for adjunctive hyperbaric oxygen therapy but will NOT receive adjunctive hyperbaric oxygen therapy
- Patient has received TCOM as part of standard of care treatment.

Exclusion Criteria:

- Prisoners
- Pregnant women

(c) Assignment to Groups. Subjects will be assigned to groups with different collection procedures based on their personal decisions for their clinical treatment course (i.e., the addition of hyperbaric oxygen therapy to their normal diabetic foot ulcer care). As this study uses convenience sampling, we will have knowledge of whether each potential subject meets the inclusion criteria for either the treatment or the control group. Therefore, during consent and enrollment, all potential study subjects will be assigned and enrolled into the appropriate study group, as they will have already consulted with their physician regarding treatment options.

(d) Compensation. Subjects will not be compensated in any way for participation in this study.

(e) Costs. Subjects will not incur any additional costs for participation in this study.

(f) Investigator Inducements. Study investigators will not receive any inducements for participation in this study.

4.4 **Informed Consent Process and Documentation**

(a) Describe the informed consent process. If the consent process will occur in a context that might involve lowered comprehension (e.g., emergency room setting, crisis state, under sedation), explain how comprehension will be assured. If children are involved in the research, describe the process for obtaining the permission of parents and the assent of the child-subjects.

(b) If Informed Consent will not be obtained, or if an informed consent document will not be used, explain why. Show how obtaining consent would not pose greater than minimal risks to subjects, would not adversely affect their rights and welfare, and would not be practicable.

(a) Consent. A complete informed consent discussion will take place for all subjects, lead by the clinical P, a sub-investigator or clinical research coordinator. The discussion will include the risks and benefits of study participation, as well as the voluntary nature of the study. Subjects will be given ample time to ask questions and to determine whether they would like to participate in the study.

(b) Waiver of Consent. All subjects will be required to provide written, informed consent in order to participate in this study. A waiver of consent will not be sought.

4.5 Monitoring of Data Adequate to Ensure Safety of Subjects

(a) Describe any local or study-wide data monitoring procedures.

(b) Explain how the IRB will be kept informed of the results of such monitoring.

(a) **Monitoring.**

MEDICAL MONITOR STATEMENT

The medical monitor may be assigned to assess one or more of the following phases of research project: volunteer recruitment, volunteer enrollment, data collection, or data storage and analysis. The medical monitor will provide an independent evaluation of serious adverse events and unanticipated problems involving risk to subjects or others to the IRB and the HRPO. The medical monitor may be assigned to discuss research progress with the principal investigator, interview volunteers, consult on individual cases, or evaluate adverse event reports. Medical monitor, Dr. Michael Berman, DO will promptly report discrepancies or problems to the IRB and the HRPO. He shall have the authority to stop a research study in progress, remove individual volunteers from a study, and take whatever steps are necessary to protect the safety and well-being of research volunteers until the IRB can assess the medical monitor's report

Although this study is slightly higher than minimal risk, the sample collection procedures are similar to standard of care procedures performed daily at the Banner Wound Care and Hyperbaric Medicine Center. The research personnel obtaining the samples have extension experience in collecting samples from diabetic foot wounds. At the start of each sample collection procedure and throughout the procedure, the hemostasis and pain control of the subject will be assessed. Any Serious Adverse Events (SAEs) will be reported promptly to the medical monitor, the BGSMC IRB and the HRPO. The medical monitor for the study will be Dr. Michael Berman, DO, who is a physician in the clinic. Although he is not involved in this research study, Dr. Berman has extensive experience in treating wounds.

All unanticipated problems involving risk to subjects or others, serious adverse events related to participation in the study and subject deaths related to participation in the study should be promptly reported by phone (301-619-2165), by email (hsrrb@det.amedd.army.mil), or by facsimile (301-619-7803) to the USAMRMC, Office of Research Protections, Human Research Protection Office. A complete written report will follow the initial notification. In addition to the methods above, the complete report will be sent to the U.S. Army Medical Research and Materiel Command, ATTN: MCMR-ZB-PH, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

(c) Communication with IRB. A copy of the approved continuing review report and local IRB approval notification will be submitted to HRPO as soon as these documents become available. A cop of the approved final study report and local IRB approval notification will be submitted to the HRPO as soon as these documents become available. If any issues arise with study subjects or with biospecimen or data collection, the IRB will be promptly notified.

(d) Protocol Modifications. Major modifications to the research protocol and any modifications that could potentially increase risk to subjects must be submitted to the ORP HRPO for approval prior to implementation. All other amendments will be submitted with the continuing review report (if applicable) to the ORP HRPO for acceptance.

(e) Review of Research Records. Accurate and complete study records will be maintained and made available to representatives of the U.S. Army Medical Research and Materiel Command. These representatives are authorized to review research records as part of their responsibility to protect human research volunteers. Research records will be stored in a confidential manner so as to protect the confidentiality of subject information.

(f) ORP HRPO Final Approval. The protocol will be conducted in accordance with the protocol submitted to and approved by the ORP HRPO and will not be initiated until written notification of approval of the research project is issued by the ORP HRPO.

(g) Compliance: The knowledge of any pending compliance inspection/visit by the FDA, DHHS-OHRP, or other government agency concerning clinical investigation or research, the issuance of Inspection Reports, FDA Form 483, warning letters or actions taken by any Regulatory Agencies including legal or medical actions and any instances of serious or continuing noncompliance with the regulations or requirements will be reported immediately to ORP HRPO.

(h) Protocol Deviations. Any deviation to the protocol that may have an effect on the safety or rights of the subject or the integrity of the study must be reported to the ORP HRPO as soon as the deviation is identified.

Accurate and complete study records will be maintained and made available to representatives of the U.S. Army Medical Research and Materiel Command. These representatives are authorized to review research records as part of their responsibility to protect human research volunteers. Research records will be stored in a confidential manner so as to protect the confidentiality of subject information.

4.6 Privacy of Subjects and Confidentiality of Data are Adequately Protected

(a) Describe procedures for ensuring the privacy of subjects and the confidentiality of data, including procedures for protecting electronically stored data.

(b) If identifiable private information will be obtained from anyone other than the target subject (eg, subjects' family members, classmates, friends), please explain and justify.

(a) Protections. All biospecimens and clinical data collected for this study will be completed de-identified. All data will be identified by a study-specific code. The key, linking identifiable subject information to the code, will be kept by the clinical PI or delegee on a password-protected database. Only de-identified data, containing the study code, will be sent to TGen for analysis. The TGen PI and any study staff at TGen will be prohibited from accessing or attempting to access the key to the code. Clinical data will be entered into a password-protected electronic database at TGen. As described earlier, no identifiers will be present in the database. Clinical data and biosamples will be stored at TGen indefinitely.

(b) Non-Target Subjects. No non-target subjects will be approached for information about the subject.

4.7 Safeguards for Vulnerable Subjects.

Describe safeguards to protect the rights and welfare of any subjects who may be vulnerable to coercion or undue influence (e.g., children; pregnant women; persons with cognitive, mental, economical, educational, or social disadvantages).

No vulnerable subjects will be approached for study participation. All subjects must be able to provide consent to participation in this study.

6. Principal Investigator's Statement of Commitment

The proposed investigation involves human subjects. I certify that I am knowledgeable about and will follow applicable federal regulations, TGen requirements, and IRB determination s for the conduct of human subject research. I agree to:

- ***a. Obtain the voluntary informed consent of subjects (or of subjects' legally authorized representatives) to the extent required by federal regulations and by the determinations of the IRB.***
- ***b. Report to the IRB any serious or unexpected on-site or off-site adverse events or unanticipated problems involving risks to subjects or other within the appropriate reporting period.***
- ***c. Cooperate fully with the IRB in the timely continuing review of this project.***
- ***d. Obtain prior approval from the IRB before amending or altering this research project or implementing changes in the approved informed consent document.***
- ***e. Maintain informed consent documents and progress reports as required by institutional policies, IRB requirements, and federal regulations).***
- ***f. Accept responsibility for the conduct and supervision of this research and the protection of human subjects as required by state and federal law and regulation, and as documented in the TGen Federalwide Assurance, guidelines of the TGen-designated IRB(s), and TGen policies and procedures.***

Signature of Principal Investigator **Date**

Appendix B: ExPEC Hallmark Assay Manuscript

TITLE: Design and validation of real-time PCR assays for the identification of extraintestinal pathogenic *E. coli* (ExPEC)

AUTHORS: Katerina Soldanova[1], Cindy M. Liu[1,2,3], Tania Contente-Cuomo[1], Maliha Aziz[1, 4], Brian Johnston[5], James R. Johnson[5], Jordan Buchhagen[1], Owain Moss[1], Lori Gauld[6], and Lance B. Price[1,4]§

[1] Division of Pathogen Genomics, Translational Genomics Research Institute (TGen), Flagstaff AZ 86001, USA

[2] Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ 86001, USA

[3] Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA

[4] Department of Environmental and Occupational Health, George Washington University, Washington, DC 20037, USA

[5] Veterans Affairs Medical Center and University of Minnesota, Minneapolis, MN, USA

[6] Flagstaff Medical Center, Flagstaff, AZ 86001, USA

§ Corresponding author

E-mail addresses:

KS: ks458@nau.edu

CML: cindy.liu@nau.edu

TC: tcontente-cuomo@tgen.org

MA: maziz2@tgen.org

BJ: bdj26miles@yahoo.com

JRJ: johns007@umn.edu.

JB: jbuchhagen@tgen.org

OM: OMoss@tgen.org

LG: Lori.Gauld@nahealth.com

LBP: lprice@tgen.org

KEYWORDS:

Detection, *Escherichia coli*, ExPEC, Real-time PCR, Urinary tract infection, and Virulence.

ABSTRACT

Background. *Escherichia coli* can be broadly classified into three major groups: commensal, diarrheagenic and extraintestinal pathogenic *E. coli* (ExPEC). In the developed world, ExPEC infections, including urinary tract infections (UTIs), sepsis, and neonatal meningitis, constitute a significant public health burden, but their epidemiology is not well understood. Rapid and reliable identification of ExPEC is important, but current methods are time-consuming and can yield equivocal results. To address this, we designed and validated six real-time PCR assays to screen *E. coli* for the hallmark ExPEC virulence genes.

Results. TaqMan® probe-based real-time PCR assays were designed to target the following ExPEC virulence genes: *papA, papC, sfaE, afaC, kpsMII,* and *iutA*. All assays were sensitive and specific when run individually or multiplexed into two reactions. Most assays were slightly more sensitive when run on purified DNA (mean sensitivity 100%) as compared to unpurified, boiled DNA preparations (mean sensitivity 94.82%). Specificities were identical for all assays regardless of the extraction method (mean specificity 99.33%). The real-time PCR assays were generally more sensitive than previously published standard PCR assays for the same targets.

Conclusion. The real-time PCR assays presented here can be applied to clinical, environmental, and foodborne isolates to reliably differentiate ExPEC from other types of *E. coli.*

BACKGROUND

Escherichia coli is a diverse species that can be broadly classified into three major groups: commensal, diarrheagenic and extraintestinal pathogenic *E. coli* (ExPEC) [1-3]. Commensal *E. coli* strains are found in the normal human gastrointestinal tract and typically lack the virulence genes necessary to cause serious human disease [3]. The diarrheagenic *E. coli* are responsible for gastrointestinal infections and are frequently associated with foodborne diarrhea outbreaks [3, 4]. In contrast, ExPEC strains do not cause gastrointestinal illness, but can cause infections outside of the gastrointestinal tract, including cystitis, pyelonephritis, meningitis, and sepsis [5-8]. In the United States, ExPEC urinary tract infections (UTIs) present a serious public health burden: ExPEC is responsible for 70-95% of community-acquired UTIs and 50% of hospital-acquired UTIs each year [9-12]. Treatment of these prevalent infections in the US is estimated to cost between $1.6 and $2.5 billion [13-15]. The epidemiology of community-acquired ExPEC infections is not well understood, but several studies have linked *E. coli* from retail meat—particularly poultry—to extraintestinal human infections [16, 17].

Distinguishing ExPEC from non-ExPEC *E. coli* strains is an essential part of elucidating the reservoirs and transmission routes of ExPEC. A widely accepted molecular definition for ExPEC is the possession of at least two of the following virulence genes: *papA, papC, sfa/foc, afa/dra, kpsM II* and *iutA* [15, 18]. These hallmark ExPEC genes encode virulence factors that include adhesins, capsular proteins, and siderophores, which greatly enhance the ability of *E. coli* to cause extraintestinal disease [19-23]. Currently, interrogations of these ExPEC hallmark genes rely on conventional PCR [24, 25], DNA probe microarrays and reverse probe hybridization [26, 27], which can be time-consuming and labor intensive. To address these issues, we have designed, optimized, and validated real-time PCR assays that can evaluate the presence of the six ExPEC hallmark genes in a rapid, sensitive, and multiplexed fashion.

RESULTS

I. Real-time PCR assay design and optimization

All primers and probes were designed from highly conserved regions identified within each virulence gene. The conserved regions were detected by aligning multiple sequences for each gene, as described in the methods section. Each virulence-factor-specific set of primers (F and R) and TaqMan® dye-labeled probe generated a PCR amplicon size of ≤ 220 bp. Each probe was labeled with a different reporter dye to allow triplexing (Table 1). Probe concentration was identical (0.13 μM) across all assays (n=6), whereas the concentration of each assay-specific primer set was individually optimized using standards and a no template control. The optimal primer concentration (either 0.10 μM, 0.30 μM or 0.50 μM) was determined based on the quality of amplification curves and the numerical value of the crossing point (Cp, fractional cycle number at which the fluorescent signal crosses a threshold) yielded by the Light Cycler software. The concentration yielding the lowest Cp values for a set of positive standards indicated earliest amplification and was used for subsequent validation experiments (Figure 1). The probe and primer sequences, optimized primer concentrations, and amplicon size for each assay are listed in table 1.

II. Whole genome sequencing data analysis

To validate our assays we used a collection of 174 *E. coli* isolates; 133 were recovered from retail poultry products and 41 were obtained from clinical UTIs. All isolates were analyzed using whole genome sequencing (WGS), which provided a basis to assess the presence or absence of each virulence gene (Table S1). Briefly, the number of isolates with each hallmark gene (total number, % id), in order of descending gene prevalence, was as follows: *iutA* (100, % id ≥ 98.59), *papA* (53, % id ≥ 98.60), *kpsMII* (51, % id ≥ 94.86), *papC* (40, % id ≥ 99.45), *sfaE* (13, % id ≥ 99.57) and *afaC* (4, % id ≥ 99.38). The WGS data was used as the gold standard to assess assay sensitivity and specificity.

III. Evaluation of assay sensitivity and specificity

E. coli isolates were identified 1) as true positive (TP) when a virulence gene detected by WGS was also detected by the real-time assay (with Cp values ≤ 27) or 2) as false negative (FN) when a virulence gene detected by WGS was not detected by the assay. Our validation experiments showed that all assays were ≥ 99.01% sensitive, with the exception of *sfaE* (92.86% sensitivity) (Table 2). In total, only two isolates were identified as FN for the following genes: *sfaE* (n=1)

and *iutA* (n=1). WGS data indicated that the virulence gene was present in both cases, but there was no amplification in the real-time PCR reaction (*iutA*), or it occurred at a Cp ≥ 28 (*sfaE*) (data not shown). Sensitivities for all assays are listed in table 2.

E. coli isolates were qualified 1) as true negative (TN) when a virulence gene was determined to be missing by both WGS and the real-time PCR assay or 2) as false positive (FP) when a virulence gene was not detected by WGS, but determined to be present by the assay (with Cp values ≤ 27). Specificities for all assays were ≥ 93.15% (Table 2) and no intra-assay cross-reactivity was observed (data not shown).

In addition to purified DNA extracts, real-time PCR assays were tested on unpurified, boiled DNA extracts (heat-soak DNA). Three assays (*kpsMII, sfaE* and *iutA*) were slightly less sensitive when tested on heat-soak DNA (mean % sensitivity 91.36% versus 100.00%). Two assays (*papA* and *papC*) had the same sensitivities (100.00%) and 1 assay (*afaC*) was not evaluated for sensitivity because no isolates from the tested subset were *afaC*–positive. (Table S2). Assay specificities were identical regardless of the extraction.

IV. Comparison of real-time PCR to conventional PCR: purified DNA

Additional testing was performed by traditional PCR on a set of 173 isolates. The results were comparable to those yielded by the real-time PCR assays. Briefly, 117 isolates had identical gene profiles (6 loci assayed) by both methods; 43 isolates matched at 5 loci and 10 isolates matched at 4 loci. Only 3 isolates had a profile with 3 matched genes and there were no isolates matching at <3 loci.

Comparison of the two methods revealed that in many cases, traditional PCR assays were less sensitive than the corresponding real-time PCR assays. This fact was demonstrated by non-detection of the hallmark genes for multiple isolates. More specifically, comparison of traditional PCR results to *in silico* analysis indicated that the *papA* assay had the largest number of FN results (n=40). False negative results for the remaining assays were as follows: n=14 for *kpsM II*, n=9 for *iutA*, and n=3 for *afaC, sfaE* and *papC* assays. Comparison of real-time PCR results to *in silico* analysis revealed that a few isolates were identified as FP by the real-time assays: n=3 for *papA* and *papC,* n=2 for *iutA* and n=1 for *kpsM II afaC* and *sfaE*. Our results indicate that the real-time PCR assays are more reliable at identifying the ExPEC hallmark genes.

DISCUSSION

Aligning all publicly available sequences for each hallmark virulence gene revealed variable and conserved regions for primer and probe design. This process also revealed potentially disruptive polymorphisms in the priming sites of previously published standard PCR-based ExPEC assays [28, 29]. A polymorphism disrupting the three-prime end of the XXXX primer for the *papA* PCR assay appears to have had a profoundly negative impact on its performance [This study].WGS uggests that the commonly used traditional PCR assay for *papA* only detects approximately 30% of its intended targets. Likewise, but to a lesser degree, the traditional PCR assays for *kpsMII* and *sfaE* also appear to be disrupted by polymorphisms in their priming sites (cite). While WGS of the test isolate set confirmed the conserved nature of the primer and probe sites chosen for the real-time PCR assays presented here, there is no guarantee that disruptive mutations will not arise in the future.

All of the real-time PCR assays were greater than 93% specific to their intended target. Five of the real-time assays (*papA, papC, kpsMII, afaC* and *iutA*) made between one and seven false-positive calls, representing a small portion of the isolates tested. Early amplification curves (with Cp values < 27) were observed for these isolates, contradicting the sequencing results, which indicated that the genes were missing. This may suggest that the genes were sufficiently divergent as to be missed by our stringent gene finder algorithm (% id ≥ 95.00), but sufficiently homologous as to anneal with the primers and probes.

Assay sensitivity potentially increases and becomes more reliable when evaluated on a large sample size. Since we had only 14 *sfaE*-positive isolates to validate the assay, the single failure (false negative) had a substantial impact on sensitivity calculation (92.86%). With the exception of *afaC* (n=4) and *sfaE* (n=14), the number of isolates positive for all genes was ≥ 40, which increased the sensitivities to ≥ 99.01%

The real-time PCR assays presented here performed well on both purified and unpurified DNA preparations, with only slightly decreased sensitivity on the latter. Unpurified boiled DNA preparations contain cell debris, which can function as PCR inhibitors, impeding amplification and lowering assay sensitivity. However, extractions by boiling are cheaper, faster, and easier to perform than the commercial methods and appear to provide reasonable quality DNA for real-time PCR, when cost is a significant factor.

Using our new real-time PCR assays, we demonstrated that a substantial proportion (29.32%) of *E. coli* isolates from retail poultry products met the molecular definition of ExPEC. Further studies must be conducted to quantify the risk of colonization and infection to the consumer.

CONCLUSION

The reservoirs of ExPEC in the community are not well understood, but likely include the human gastrointestinal tract, companion animals and, food animals, as well as retail meat and poultry. The real-time PCR assays described here will enable researchers to screen *E. coli* isolates from various sources to differentiate ExPEC from other types of *E. coli* in a high throughput, rapid, and reliable fashion.

MATERIAL AND METHODS

I. Real-time PCR assay design

A probe-based real-time PCR assay was designed for the following ExPEC-defining virulence genes: pyelonephritis associated pilus (*papA* and *papC*), group II polysaccharide capsule (*kpsMII*), S fimbrial and afimbrial antigens (*sfaE* and *afaC*), and siderophore receptor for iron uptake (*iutA*). Each assay was designed by downloading all publicly available sequences for these virulence genes from the Pathosystems Resource Integration Center (PATRIC), accessed at www.patricbrc.org and from Genbank, accessed at http://www.ncbi.nlm.nih.gov/genbank/. In-house Sanger sequencing data were used to increase the number of available sequences.

All downloaded sequences were checked for specificity using the NCBI Basic Local Alignment Search Tool (BLAST) against Genbank whole-genome sequence and nucleotide collections. Next, all sequences for a single gene were combined into a multiple-sequence alignment file and each alignment file was examined in SeqMan (DNASTAR Inc., Madison, WI, USA). Among the entire set, specific sequences were chosen to identify consensus conserved regions used for the primer and probe design. The number and databank accession numbers of each sequence are listed below: *papA*, seven [GenBank: CU928164.2, GenBank: CP001671.1, GenBank: AE014075.1, GenBank: CU928161.2, GenBank: CP000468.1, GenBank: CP000243.1, GenBank: CU928163.2]; *papC,* six [GenBank: CP000243.1, GenBank: CU928164.2, GenBank: AE014075.1, GenBank: CU928163.2, GenBank: CU928161.2, GenBank: CP000468.1]; *iutA*, 10 [GenBank: AE014075.1, GenBank: AP010960.1, GenBank: CU928145.2, GenBank: CP001671.1, GenBank: CP000970.1, GenBank: CU928161.2, GenBank: AP010953.1, GenBank: CU928162.2, GenBank: CU928164.2, GenBank: CU928163.2, GenBank: CP000468.1]; *afaC*, nine [GenBank: FM955458, GenBank: AF325672, GenBank: AF329316, GenBank: FM955462, GenBank: FM955461, GenBank: FM955460, GenBank: FM955459, GenBank: FJ843076, GenBank: FABCDEX76688]; *kpsMII*, eleven [GenBank: AE014075.1, GenBank: CP000243.1, GenBank: CP001671.1, GenBank: CU928162.2, GenBank: CP000468.1, GenBank: CU928163.2, GenBank: FN554766.1, GenBank: CP001509.3, GenBank: CU928161.2, GenBank: CU928164.2, GenBank: CP0009701] and *sfaE,* four

[GenBank: CP001509.3, GenBank: CP000243.1, GenBank: AE014075.1 and GenBank: CP000819.1]. The design was performed using Primer Express 3.0 (Applied Biosystems, Carlsbad, CA, USA). Each assay consisted of two primers (F and R) and a dye-labeled TaqMan® probe; each probe was labeled with a reporter dye (FAM, VIC, or NED) at the 5'-terminus and a Minor-Groove-Binding Non-Fluorescent Quencher (MGBNFQ) at the 3'-terminus. Primer and probe sequences are listed in Table 1.

II. Assay optimization using quantified and normalized plasmid standards

II A. Generation of normalized plasmid standards

The James Johnson Laboratory provided six ExPEC isolates, which were used as positive controls, as each isolate carried one of the following virulence genes: *papA*, *papC, kpsMII, sfaE, afaC* or *iutA*. Total bacterial DNA was extracted using the DNeasy 96 Blood and Tissue Kit (Qiagen, CA, USA) and each target gene was amplified using the corresponding primers. The PCR amplicons were visualized by gel electrophoresis and used as target gene inserts for cloning reactions. Cloning was performed with the TOPO ® TA PCR4 Cloning Kit (Invitrogen, NY, USA). Chemically competent *E. coli* cells were transformed with the cloned vectors and plasmid DNA was isolated using the QIAprep Spin Miniprep Kit protocol (Qiagen, CA, USA). All commercially available kits used for DNA extraction and cloning were used according to manufacturer's instructions. Following extraction, the DNA was digested using EcoRI enzyme and the insert sizes were visualized by gel-electrophoresis and sequence-verified using the M13Fwd (5'-GTAAAACGACGGCCAGTG-3') and M13Rev (5'-GGAAACAGCTATGACCATG-3') cloning sequencing primers. Quantification of the cloned plasmids was performed in triplicates with the PlasQuant real-time PCR assay using Plasquant2 F and Plasquant2 R Primers and Plasquant2 FAM Probe in neat and 1:5 dilutions. Plasmid normalization was performed using the dilution factor $2^{\Delta Cp}$, where ΔCp = (10 – Cp value of the non-normalized cloned plasmids).

II B. Assay Optimization

The optimal primer concentration for each assay was determined using the corresponding normalized plasmid standards. The 6 assays were combined into 2 triplexes based on the type of reporter dye associated with each probe, as shown in Table 1. Seven positive standard curves were generated with the normalized plasmids in 10-fold dilutions, starting with a concentration of 10^9 copies/µL. A no template control (NTC) was included in each experiment. Three

optimizations were done for each triplex, with primer (n=6) concentrations of either 0.10 μM, 0.30 μM, or 0.50 μM. Each 10-μL reaction contained 2 μL of the cloned standard plasmid DNA in 7 dilutions, 5 μL of 5X Quanta Perfecta Multiplex Mastermix (Quanta Biosciences) and 0.06 μL (0.13 μM) of each probe. The volume of each primer set varied according to the concentration tested: 0.05 μL (0.10 μM), 0.15uL (0.30 μM), or 0.25 μL (0.50 μM). Molecular Biology Grade (MBG) water was used to reach the 10 μL reaction volume. All reactions were performed in triplicates on the LightCycler 480 System (Roche). Thermocyling conditions were identical for all runs and were as follows: initial amplification was performed for 3 min at 95°C for *Taq* activation, followed by 15sec at 95°C for denaturation, and 1 min at 55°C for annealing/extension x 45 cycles, then a cooling step at 40°C for 10 sec. A Crossing point value (Cp, fractional cycle number at which the fluorescent signal crosses a threshold) was obtained for each reaction using the Absolute Quantification 2nd Derivative Maximum algorithm from the Light Cycler 480 Software. Cp-mean values and quality of amplification curves for all standards at different primer concentrations were compared and evaluated to determine the optimal primer concentration for each hallmark assay.

III. Assay validation using whole-genome sequenced *E. coli* isolates

III A. Collection of *E. coli* isolates.

Assay validation was performed using a set of 174 *E. coli* isolates, which was composed of 133 poultry-associated isolates and 41 clinical isolates from urinary tract infection samples. Both sample types were collected over a 12-month period. The clinical isolates were collected from Flagstaff Medical Center (FMC) in Flagstaff, AZ and stored at -80°C on cryobeads (CryoCare Cryobead, Hardy Diagnostics, CA, USA) before shipment to TGen. Species confirmation was performed at the FMC clinical laboratory using the BD Phoenix™ Automated Microbiology System (Becton and Dickinson, NJ, USA). The poultry-associated isolates were recovered from commercial poultry products purchased in 9 Flagstaff, AZ grocery stores. Bacteria from poultry products were isolated according to an in-house adapted standard operating procedure (SOP) provided in the supplemental material (Document S1). Putative *E. coli* colonies were selected based on colony morphology and confirmed by an in-house *uidA* qPCR assay. Total genomic DNA from all poultry-associated and clinical isolates was extracted using the DNeasy 96 Blood and Tissue Kit (Qiagen, CA, USA) following the manufacturer's instructions. In addition, genomic DNA from the poultry-associated isolates was extracted by boiling, which entailed heating a sealed, homogeneous mixture of bacterial colonies in PBS (100 μL/sample) for 15 min at 90°C in

a PCR thermal cycler (Bio-Rad, CA, USA). The isolated DNA was used in the real-time PCR reactions without further purification steps.

III B. Laboratory Whole Genome Sequencing DNA Preparation

All isolates were prepared for in-house whole genome sequencing on the Illumina GAII Platform (Illumina, CA, USA). Briefly, genomic DNA was sheared to 500bp fragments using Sonicman™ (Matrical, Bioscience, WA, USA) and purified with the QIAquick PCR Purification Kit (Qiagen, CA, USA) according to manufacturer's instructions. Libraries were prepared on the Biomek® FX^P Laboratory Automation Workstation (Beckman Coulter, CA, USA) using the SPRI Works HT (SWHT) Reagent Kit, (Agencourt, MA, USA) and customized SWHT1 Method from the Biomek Software. Libraries were amplified and indexed with 6-nucleotide barcoded primers (n=48) (AIR DNA Primer Barcode) (Bioo Scientific, TX, USA). Each amplified library was manually purified using the Agencourt® AMPure®XP PCR Purification protocol (Agencourt, MA, USA) according to manufacturer's instructions. Real-time qPCR was performed in triplicates in 10-µL reactions on the 7900HT Fast Real-Time PCR System (Applied Biosystems) using the KAPA Library Quantification Kit (KAPA Biosystems, MA, USA). Libraries were quantified based on a regression curve generated by the 6 KAPA standards included in the quantification kit and manually pooled at equal molar concentration. Pools were sequenced at TGen using the Illumina GAII Platform and Illumina Sequencing Chemistry.

III C. Bioinformatics Whole Genome Sequencing Data Analysis

All sequenced genomes were assembled using VelvetOptimiser (Version 2.2.2) [30] and Velvet [31]. All isolates were assessed *in silico* for presence of the selected virulence genes by aligning each set of sequences used for assay design against each of the assembled genomes using Nucleotide-Nucleotide BLAST (Version 2.2.25+) [32]. Sequence similarity matches for each gene were determined using thresholds of 100% nucleotide identity and 100% coverage of the query sequence length. Isolates that were identified *in silico* as positive for a given virulence gene were considered as true positives (TP) and were selected for assay validation and sensitivity evaluations. Isolates that were determined *in silico* as negative for a virulence gene were considered as true negatives (TN) and served to calculate assay specificity.

III D. First assay validation: using pure DNA extracts as template

Each triplex was validated against all 174 sequenced *E. coli* isolates. ExPEC strains from the J. Johnson Laboratory were included in each experiment as positive controls. Criteria for real-time PCR result analysis were as follows: 1) Positive result: presence of amplification curves at Cp ≤ 27 for at least 2 out of the 3 replicates. 2) Negative result: a) presence of amplification curves at Cp > 27, b) presence of amplification curves at Cp ≤ 27 for 1 triplicate only c) absence of amplification curves. Criteria for classifying isolates as true positives (TP), true negatives (TN), false positives (FP) or false negatives (FN) were defined as follows: 1) TP: positive result yielded by the assay and presence of virulence gene confirmed by WGS; 2) TN: negative result yielded by the assay and absence of gene confirmed by WGS; 3) FP: positive result yielded by the assay and absence of gene confirmed by WGS; 4) FN: negative result yielded by the assay and presence of gene confirmed by WGS. Sensitivity and specificity were calculated using the following equations:

1) Analytical sensitivity: Sn = TP/ (TP+FN)

2) Analytical specificity: Sp=TN/ (TN+FP).

III E. Second assay validation: using non-purified DNA extracts as template

DNA from poultry-associated *E. coli* isolates (n=133) was extracted by boiling as described in section IIIA. This non-purified DNA was directly added to the real-time PCR reactions for assay validation and the results compared with data from the first validation. Evaluation of both sets of results provided additional information about the assay sensitivity and specificity performance when two different extraction methods were used.

IV. Real-time PCR data analysis

A total of 3 Cp values were obtained for each isolate (1 Cp-value/triplicate). The Cp values were obtained using the Absolute Quantification 2nd Derivative Maximum algorithm from the Light Cycler 480 Software. The raw file from each hallmark assay was analyzed by calculating the Cp-mean and Cp-standard deviations from the 3 replicates for all isolates. If necessary, Cp-standard deviation values were adjusted to ≥ 0.15 by excluding 1 triplicate from the calculation. Cp-mean and standard deviation values were also calculated for the corresponding cloned plasmid DNA included within each assay as positive control.

V. Traditional PCR data analysis

173 isolates from the complete collection were also screened for ExPEC Hallmark genes using traditional PCR. The testing and raw result analysis was carried out at Veterans Affairs Medical Center and University of Minnesota, Minneapolis, MN, USA. The reactions were performed in 25.0µL final volume containing 2.0µL DNA sample, 2.5µL 10X Buffer, 4.0 µL $MgCl_2$ (25mM), 2.0µL dNTPs (10mM), 0.25µL Amplitaq Gold (5U/µL) and 14.25µL of 200mM Primers and water mix. The thermocycling conditions were used as described below: activation step at 95ºC for 10 min followed by 28 cycles of denaturation at 94ºC for 30 sec, annealing at 63ºC for 30 sec and extension at 68ºC for 3 min. Final extension was done at 72ºC for 10 min; afterwards, the reaction was held at 4ºC.

COMPETING INTERESTS

Authors declare that they have no competing interests. JRJ has received research grants and/or contracts from Merck, Rochester Medical, and Syntiron.

AUTHORS' CONTRIBUTIONS

KS contributed to the experimental design, data acquisition, analysis and interpretation and drafted the manuscript; TC helped with laboratory sample processing MA and OM participated in the bioinformatics analysis of the sequencing data; JB carried out the sequence alignments and assay design; CML contributed to the experimental design and drafting the manuscript; LBP contributed to the funding, experimental design, and drafting the manuscript, JRJ and BJ provided the positive ExPEC isolates and consulted on various aspects of the project, BJ performed the traditional PCR reactions, LG provided the collection of FMC UTI-associates *E. coli* isolates and consulted on the clinical aspects of the project.

ACKNOWLEDGEMENTS

We thank Kara Waits and Heidi Grande at the Translational Genomic Research Institute for assistance with the sample processing and preparation of the boiled DNA extracts.

REFERENCES

[1] Russo TA, Johnson JR: Proposal for a new inclusive designation for extraintestinal pathogenic isolates of Escherichia coli: ExPEC. *J Infect Dis.* 2000, 181:1753-1754.

[2] Ewers C, Li G, Wilking H, et al. Avian pathogenic, uropathogenic, and newborn meningitis-causing Escherichia coli: how closely related are they? *Int J Med Microbiol*. 2007, 297:163-176

[3] Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes Infect*. 2000, 5:449-456.

[4] Taniuchi M, Walters CC, Gratz J, et al. Development of a multiplex polymerase chain reaction assay for diarrheagenic Escherichia coli and Shigella spp. and its evaluation on colonies, culture broths, and stool. *Diagn Microbiol Infect Dis*. 2012, 73:121-128.

[5] Jafari A, Aslani MM, Bouzari S. Escherichia coli: a brief review of diarrheagenic pathotypes and their role in diarrheal diseases in Iran. *Iran J Microbiol*. 2012, 4:102-117.

[6] Mao BH, Chang YF, Scaria J, et al. Identification of Escherichia coli genes associated with urinary tract infections. *J Clin Microbiol*. 2012, 50:449-456.

[7] Köhler CD, Dobrindt U. What defines extraintestinal pathogenic Escherichia coli? *Int J Med Microbiol*. 2011, 301:642-647.

[8] Johnson JR, Russo TA. Extraintestinal pathogenic Escherichia coli: "the other bad E coli". *J Lab Clin Med*. 2000, 139:155-162.

[9] Pitout JDD. Extraintestinal pathogenic *Escherichia coli*: a combination of virulence with antibiotic resistance. *Front. Microbiol.* 2012, 3:1-7.

[10] Wiles TJ, Kulesus RR, Mulvey MA. Origins and virulence mechanisms of uropathogenic Escherichia coli. *Exp Mol Pathol*. 2008, 85:11-19.

[11] Zaffanello M, Malerba G, Cataldi L, et al. Genetic risk for recurrent urinary tract infections in humans: a systematic review. *J Biomed Biotechnol*. 2010, 321082:1-9.

[12] Li D, Liu B, Chen M, et al. A multiplex PCR method to detect 14 Escherichia coli serogroups associated with urinary tract infections. *J Microbiol Methods*. 2010, 82:71-77.

[13] Bergeron CR, Prussing C, Boerlin P, et al. Chicken as reservoir for extraintestinal pathogenic Escherichia coli in humans, Canada. *Emerging Infect Dis*. 2012, 18:415-421.

[14] Johnson JR, Kuskowski MA, Smith K, O'bryan TT, Tatini S. Antimicrobial-resistant and extraintestinal pathogenic Escherichia coli in retail foods. *J Infect Dis*. 2005, 191:1040-1049.

[15] Jakobsen L, Spangholm DJ, Pedersen K, et al. Broiler chickens, broiler chicken meat, pigs and pork as sources of ExPEC related virulence genes and resistance in Escherichia coli isolates from community-dwelling humans and UTI patients. *Int J Food Microbiol.* 2010, 142:264-272.

[16] Johnson JR, Sannes MR, Croy C, Johnston B, Clabots C, Kuskowski MA, Bender J, Smith KE, Winokur PL, Belongia EA. Antimicrobial drug-resistant Escherichia coli from humans and poultry products, Minnesota and Wisconsin, 2002-2004. *Emerg Infect Dis*. 2007, 6:838-46.

[17] Vieira AR, Collignon P, Aarestrup FM, McEwen SA, Hendriksen RS, Hald T, Wegener HC. Association between antimicrobial resistance in Escherichia coli isolates from food animals and blood stream isolates from humans in Europe: an ecological study. *Foodborne Pathog Dis.* 2011, 12:1295-301.

[18] Abdallah KS, Cao Y, Wei DJ. Epidemiologic Investigation of Extra-intestinal pathogenic E. coli (ExPEC) based on PCR phylogenetic group and fimH single nucleotide polymorphisms (SNPs) in China. *Int J Mol Epidemiol Genet*. 2011, 2:339-353.

[19] Mora A, López C, Dabhi G, et al. Extraintestinal pathogenic Escherichia coli O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol*. 2009, 9:1-11.

[20] Vejborg RM, Hancock V, Schembri MA, Klemm P. Comparative genomics of Escherichia coli strains causing urinary tract infections. *Appl Environ Microbiol*. 2011, 77:3268-3278.

[21] Ziegler T, Jacobsohn N, Fünfstück R. Correlation between blood group phenotype and virulence properties of Escherichia coli in patients with chronic urinary tract infection. *Int J Antimicrob Agents.* 2004, 24:70-75.

[22] Krieger JN, Dobrindt U, Riley DE, Oswald E. Acute Escherichia coli prostatitis in previously health young men: bacterial virulence factors, antimicrobial resistance, and clinical outcomes. *Urology*. 2011, 77:1420-1425.

[23] Obeng AS, Rickard H, Ndi O, Sexton M, Barton M. Antibiotic resistance, phylogenetic grouping and virulence potential of Escherichia coli isolated from the faeces of intensively farmed and free range poultry. *Vet Microbiol*. 2012, 154:305-315.

[24] Morin MD, Hopkins WJ. Identification of virulence genes in uropathogenic Escherichia coli by multiplex polymerase chain reaction and their association with infectivity in mice. *Urology*. 2002, 60:537-541.

[25] Yamamoto S, Terai A, Yuri K, Kurazono H, Takeda Y, Yoshida O. Detection of urovirulence factors in Escherichia coli by multiplex polymerase chain reaction. FEMS *Immunol Med Microbiol*. 1995, 12:85-90.

[26] Johnson JR, Russo TA. Molecular epidemiology of extraintestinal pathogenic (uropathogenic) Escherichia coli. *Int J Med Microbiol*. 2005, 295:383-404.

[27] Korczak B, Frey J, Schrenzel J, et al. Use of diagnostic microarrays for determination of virulence gene patterns of Escherichia coli K1, a major cause of neonatal meningitis. *J Clin Microbiol.* 2005, 43:1024-1031.

[28] Boyd EF, Hartl DL. Diversifying selection governs sequence polymorphism in the major adhesin proteins fimA, papA, and sfaA of Escherichia coli. *J Mol Evol.* 1998, 47:258-267.

[29] Johnson JR, Delavari P, Kuskowski M, Stell AL. Phylogenetic Distribution of Extraintestinal Virulence-Associated Traits in *Escherichia coli. J Infect Dis.* 2001, 183:78-88.

[30] http://bioinformatics.net.au/software.velvetoptimiser.shtml

[31] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008, 18:821-829.

[32] Altschul, SF. et al. Basic Local Alignment Search Tool. *J.Molec Biol*. 1990, 215:403–410.

FIGURE LEGENDS

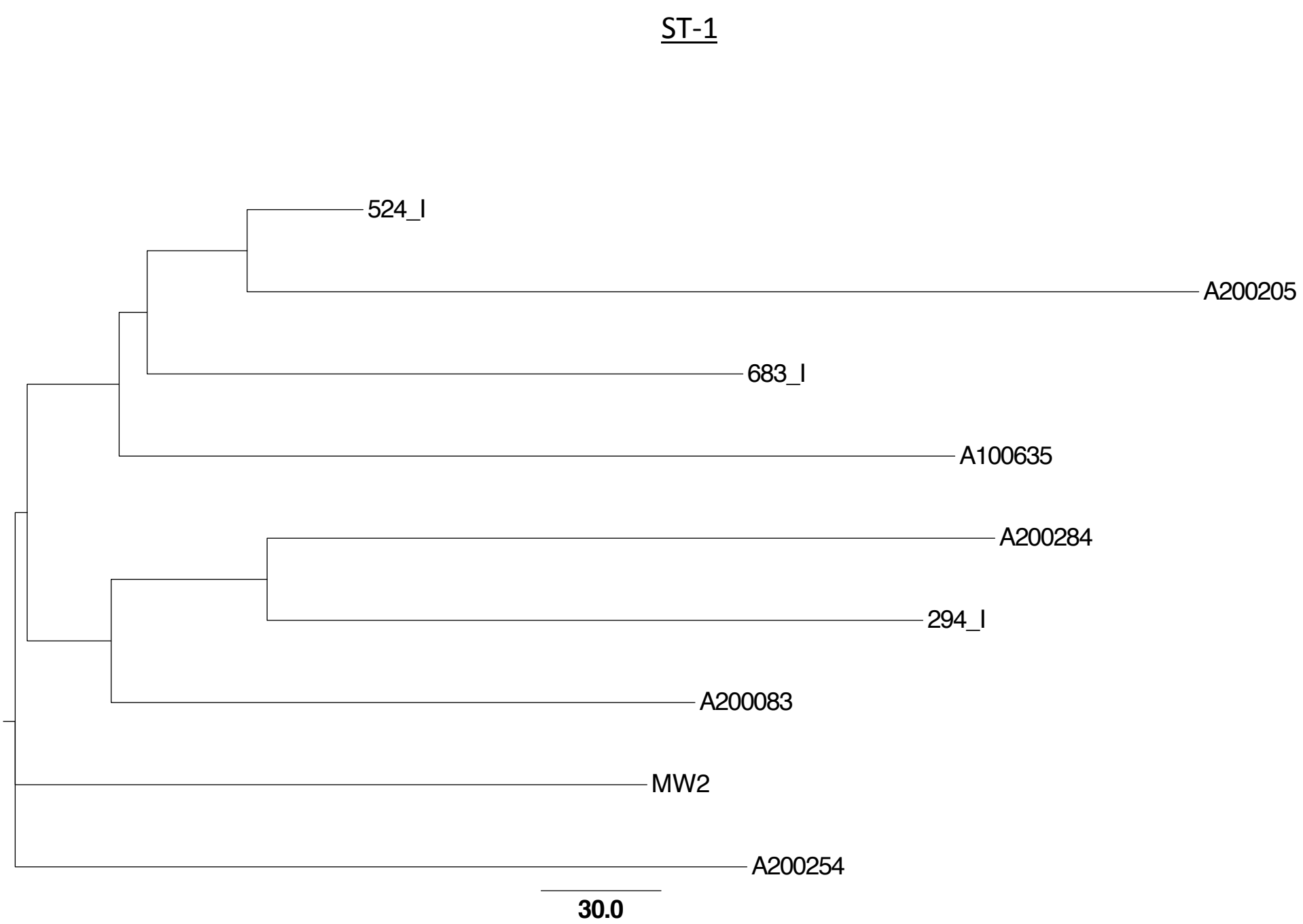
Figure 1A-C

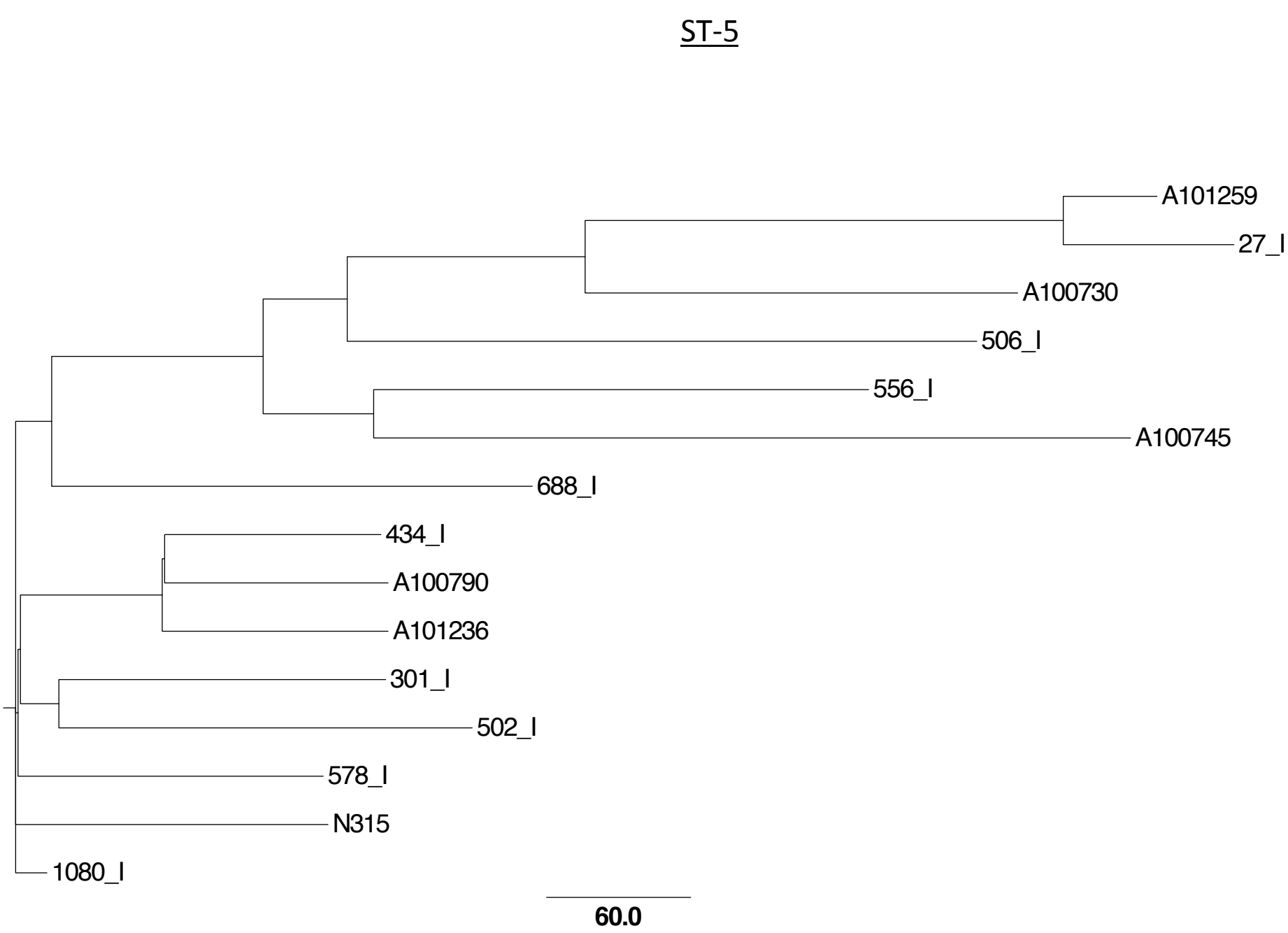
Light Cycler real-time amplification plots showing amplification curves for iutA assay at 0.1μM (Fig 1A), 0.3 μM (Fig. 1B) and 0.5 μM (Fig. 1C) respectively. The Cp values for the set of positive standards varied slightly based on the primer concentration. The primer concentration that enabled the earliest amplification (standard curves with the lowest Cp-mean values) was qualified as optimal and chosen for further hallmark VF gene assay validation experiments. For the iutA assay the amplification curve plots look similar, but the according to the Absolute Quantification 2nd Derivative Maximum algorithm lowest Cp values were obtained when primers at 0.5 μM were used.

Figure 2A-B.

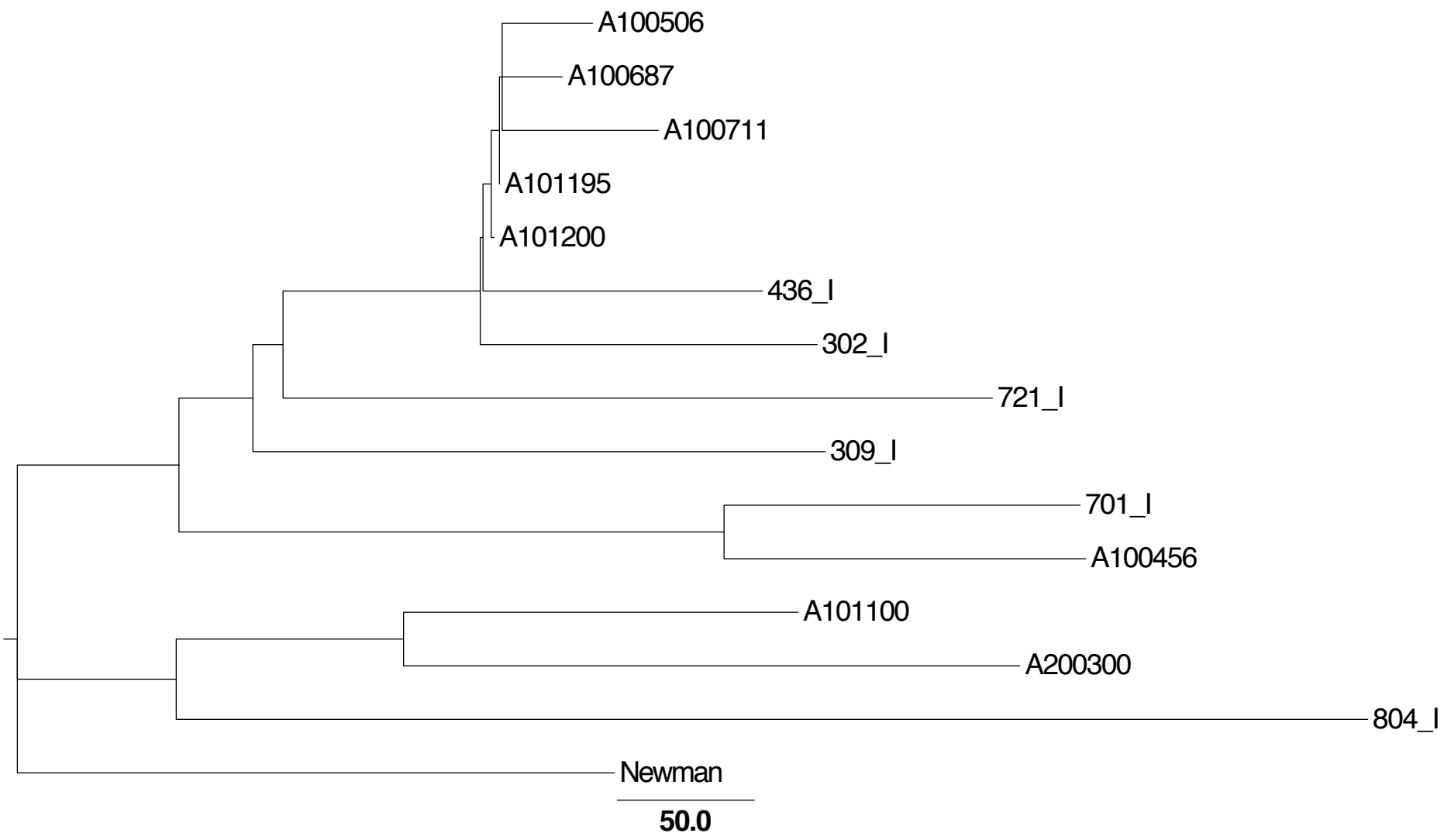
Light Cycler data analysis file from kpsMII assay showing dense amplification curves (red) from the validation experiments using all 174 sequenced *E. coli* isolates extracted using the DNeasy 96 Blood and Tissue Kit. Each amplification curve (Cp ≤ 27) represents a known *kpsM II*-positive *E. coli* isolate which was retrospectively matched to the corresponding whole genome sequencing results and compared for the presence of the *kpsMII* gene. Isolates that amplified at Cp≥ 27 were considered as negative. The known *kpsMII*-negative *E. coli* isolates (green) did not amplify at all by the assay. They were also matched to the corresponding whole genome sequencing data and assessed for the absence of the *kpsMII* gene (Fig 2A).

Figure 2B displays amplification curves for the same assay (kpsMII) tested on a small portion (collection II) of the poultry-associated *E. coli* isolates extracted using the heat-soak method. All isolates from this collection amplified at Cp≤ 27 and were considered as positive for the *kpsMII* gene. The amplification curves are similar and consistent between both assay runs irrespective of the DNA extraction method used.

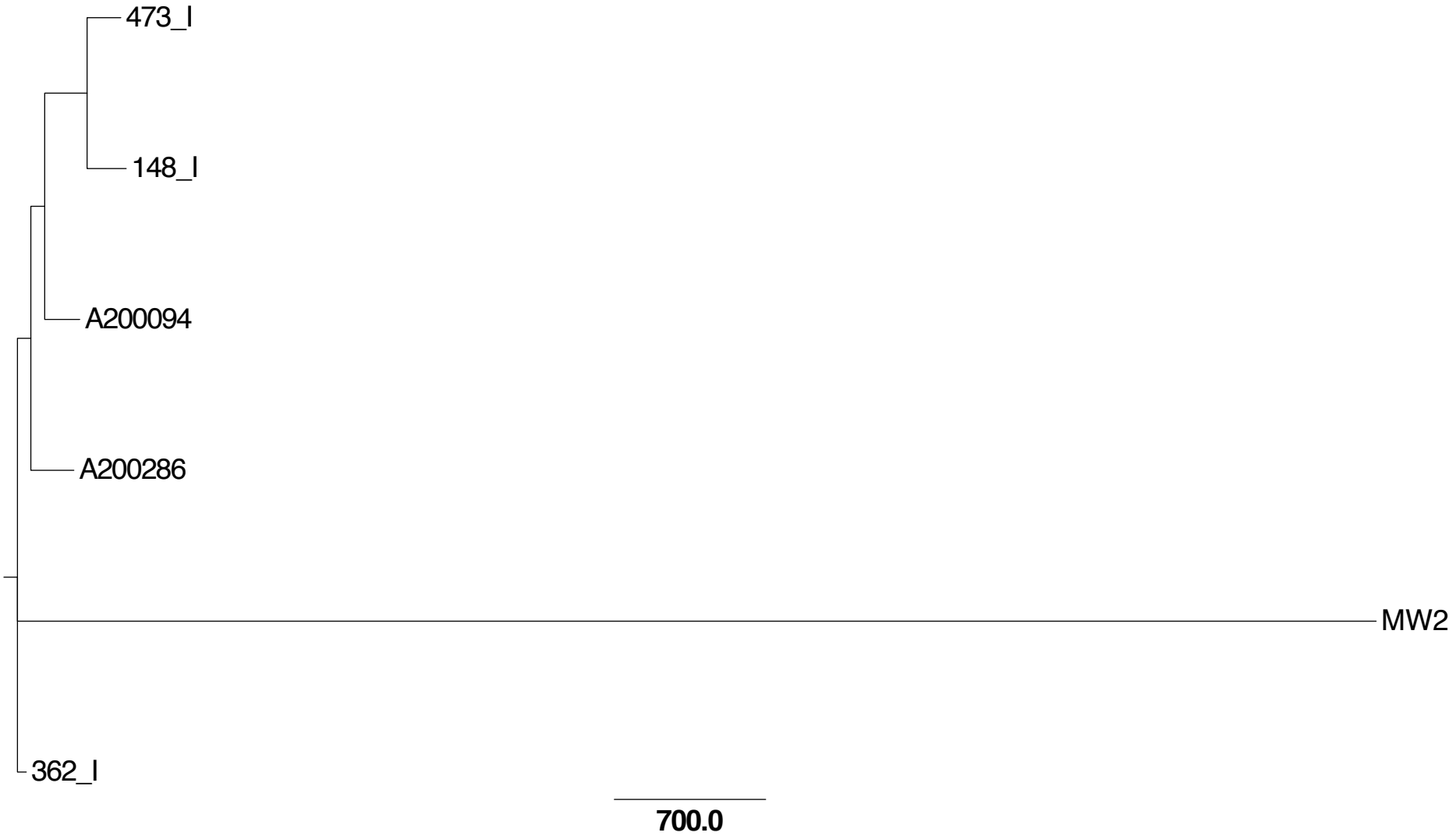

ST-1
524_I
A200205
683_I
A100635
A200284
294_I
A200083
MW2
A200254
30.0


ST-5
A101259
27_I
A100730
506_I
556_I
A100745
688_I
434_I
A100790
A101236
301_I
502_I
578_I
N315
1080_I
60.0

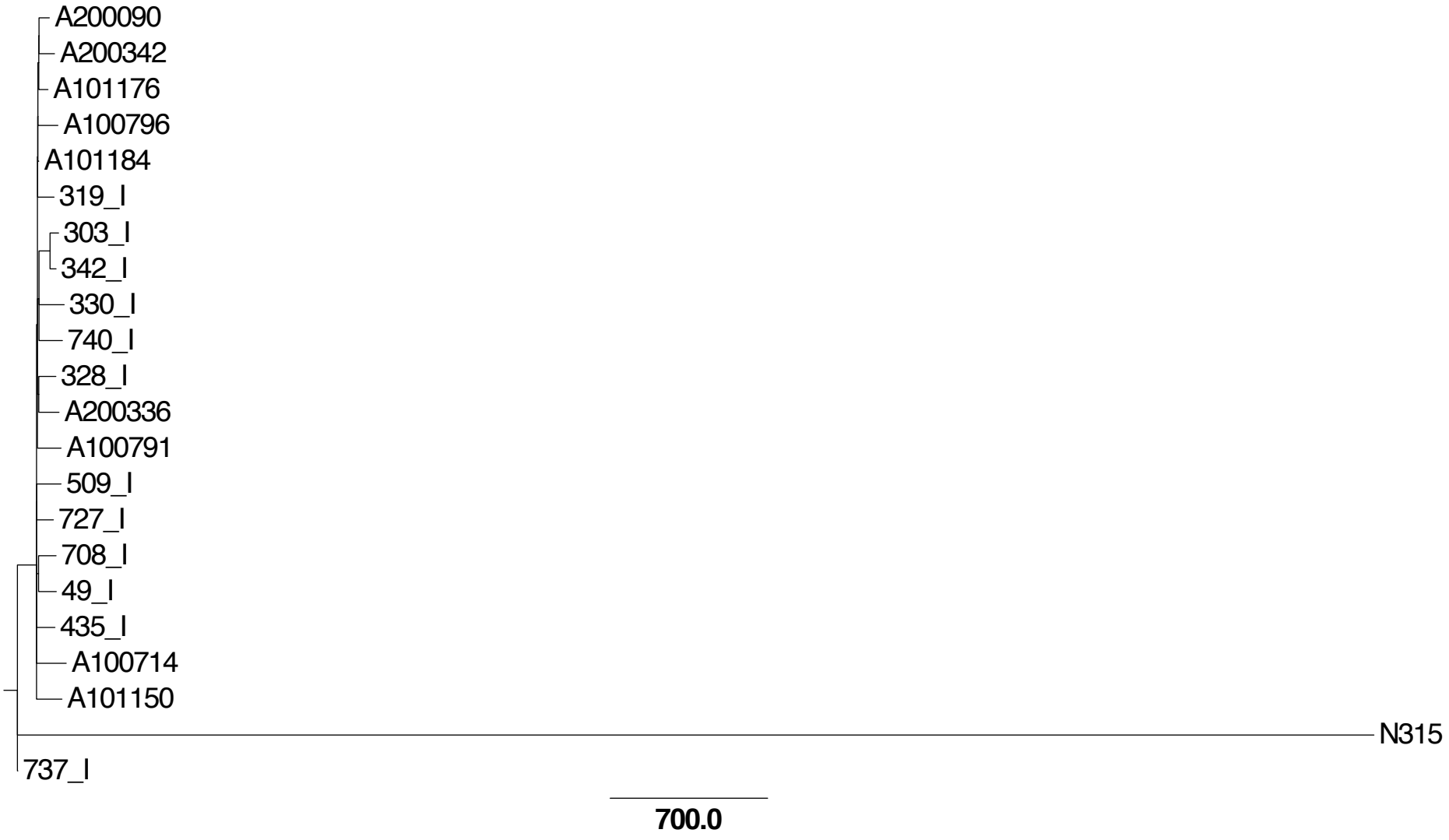
ST-8

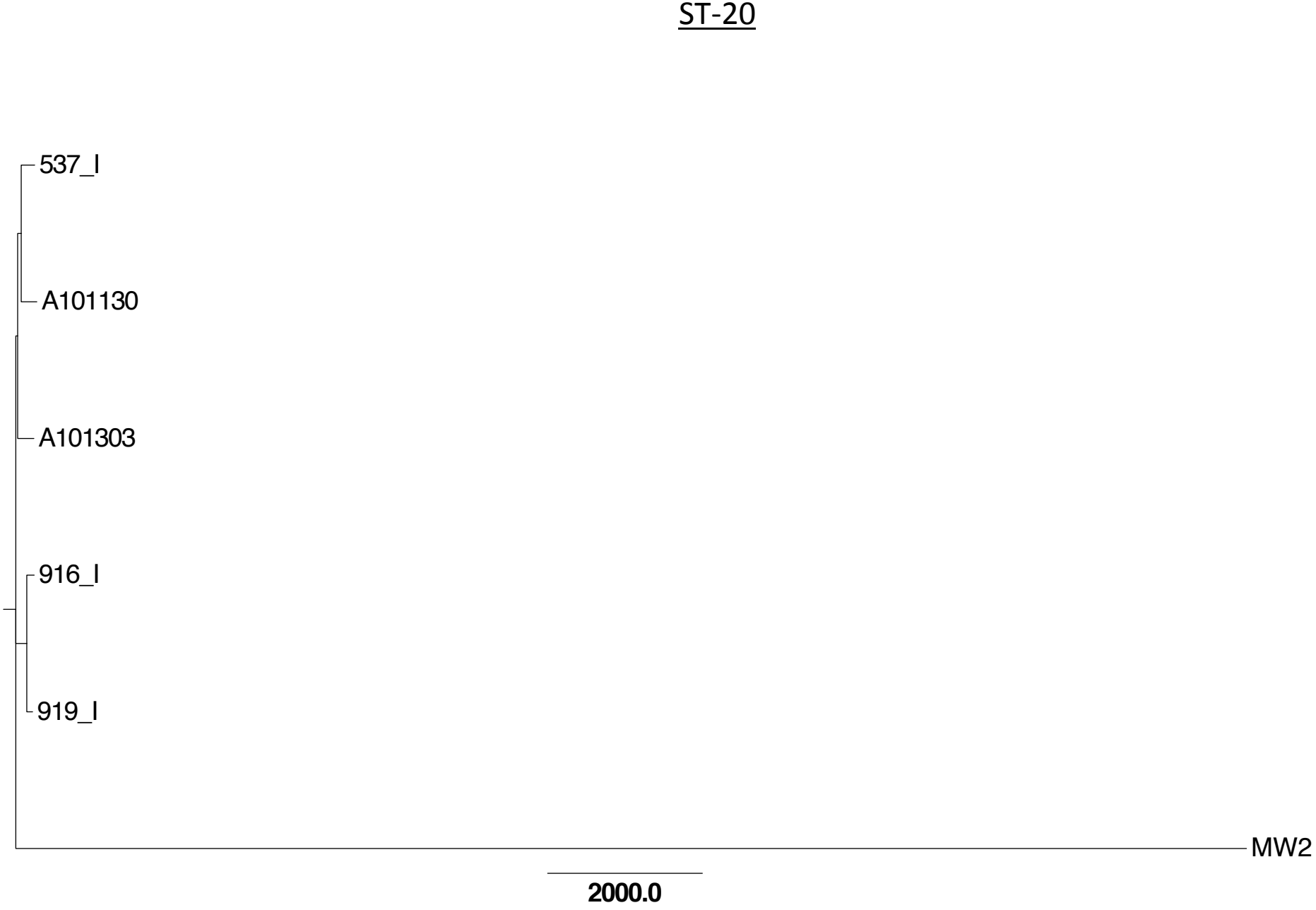

A100506
A100687
A100711
A101195
A101200
436_I
302_I
721_I
309_I
701_I
A100456
A101100
A200300
804_I
Newman
50.0

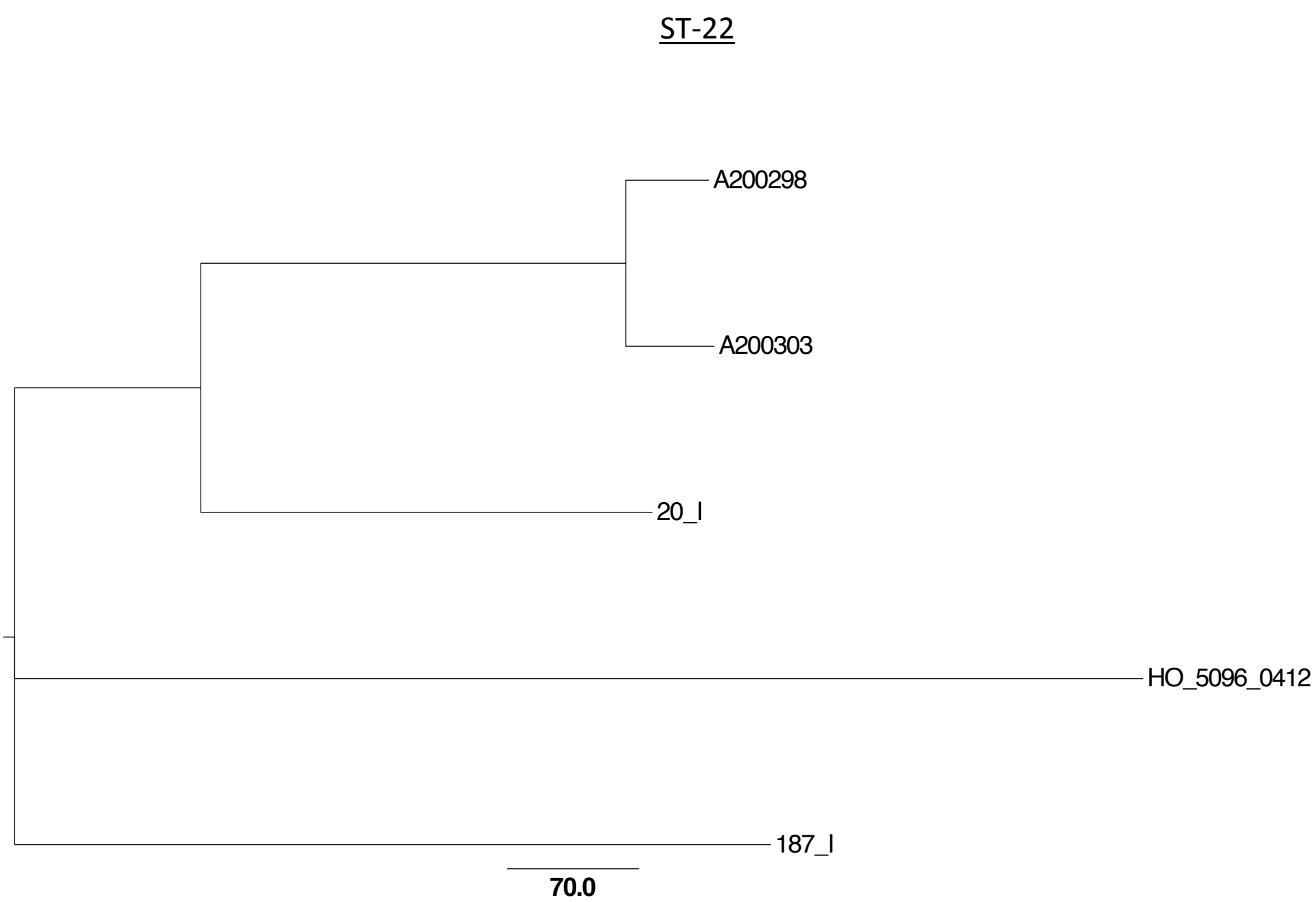
ST-12



ST-15

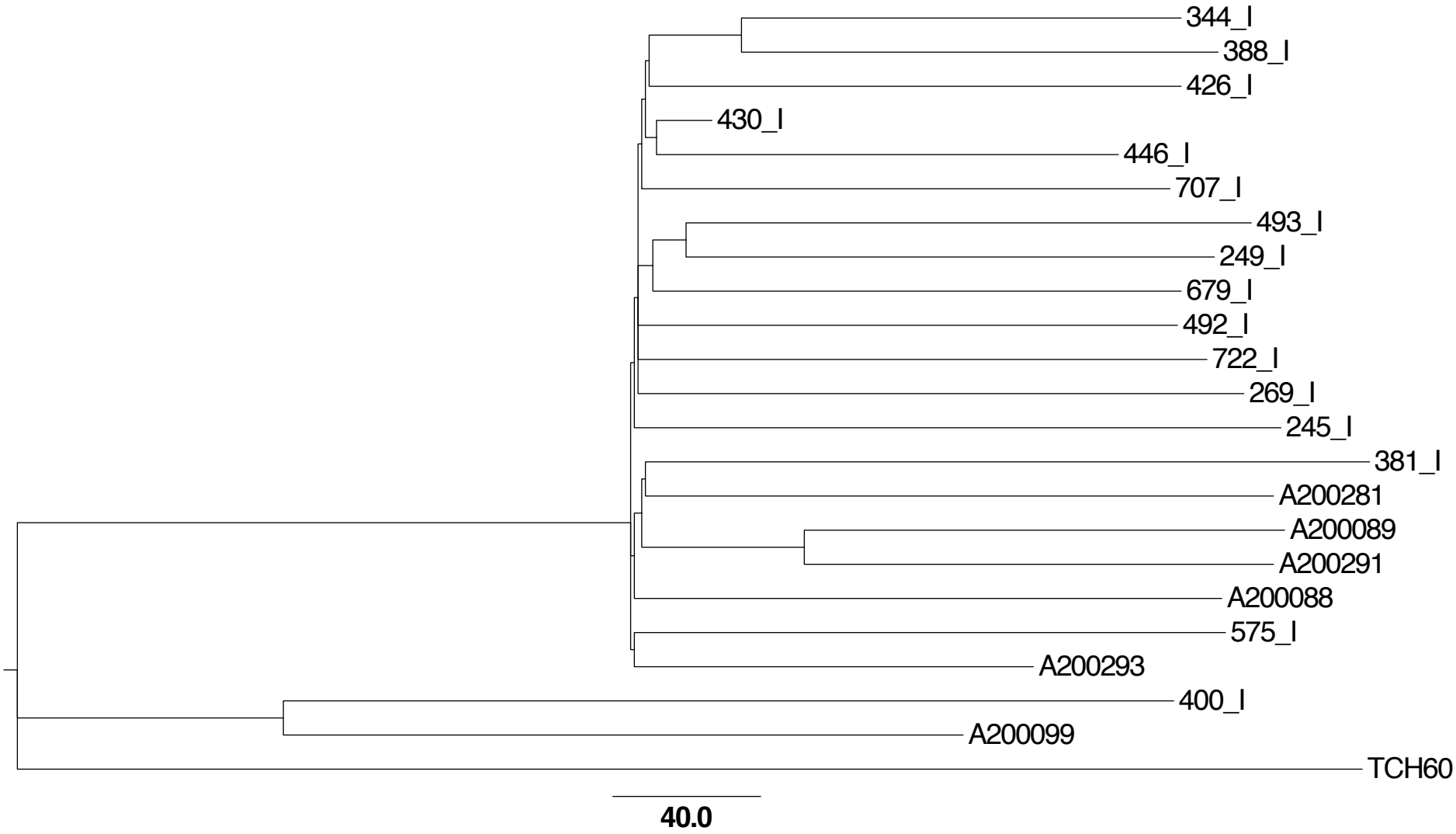

A200090
A200342
A101176
A100796
A101184
319_I
303_I
342_I
330_I
740_I
328_I
A200336
A100791
509_I
727_I
708_I
49_I
435_I
A100714
A101150
N315
737_I
700.0


ST-20
537_I
A101130
A101303
916_I
919_I
MW2
2000.0

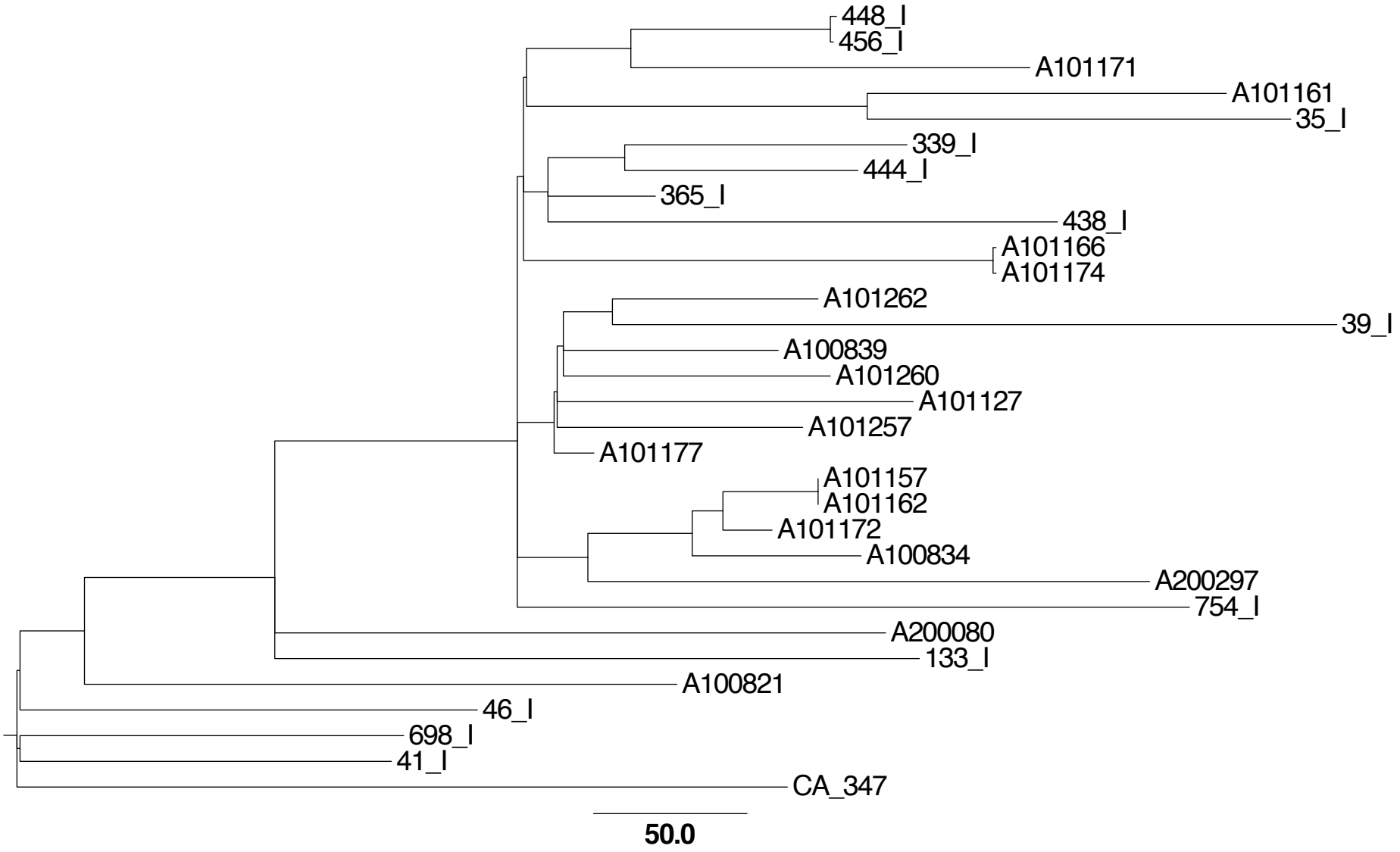

ST-22
A200298
A200303
20_I
HO_5096_0412
187_I
70.0

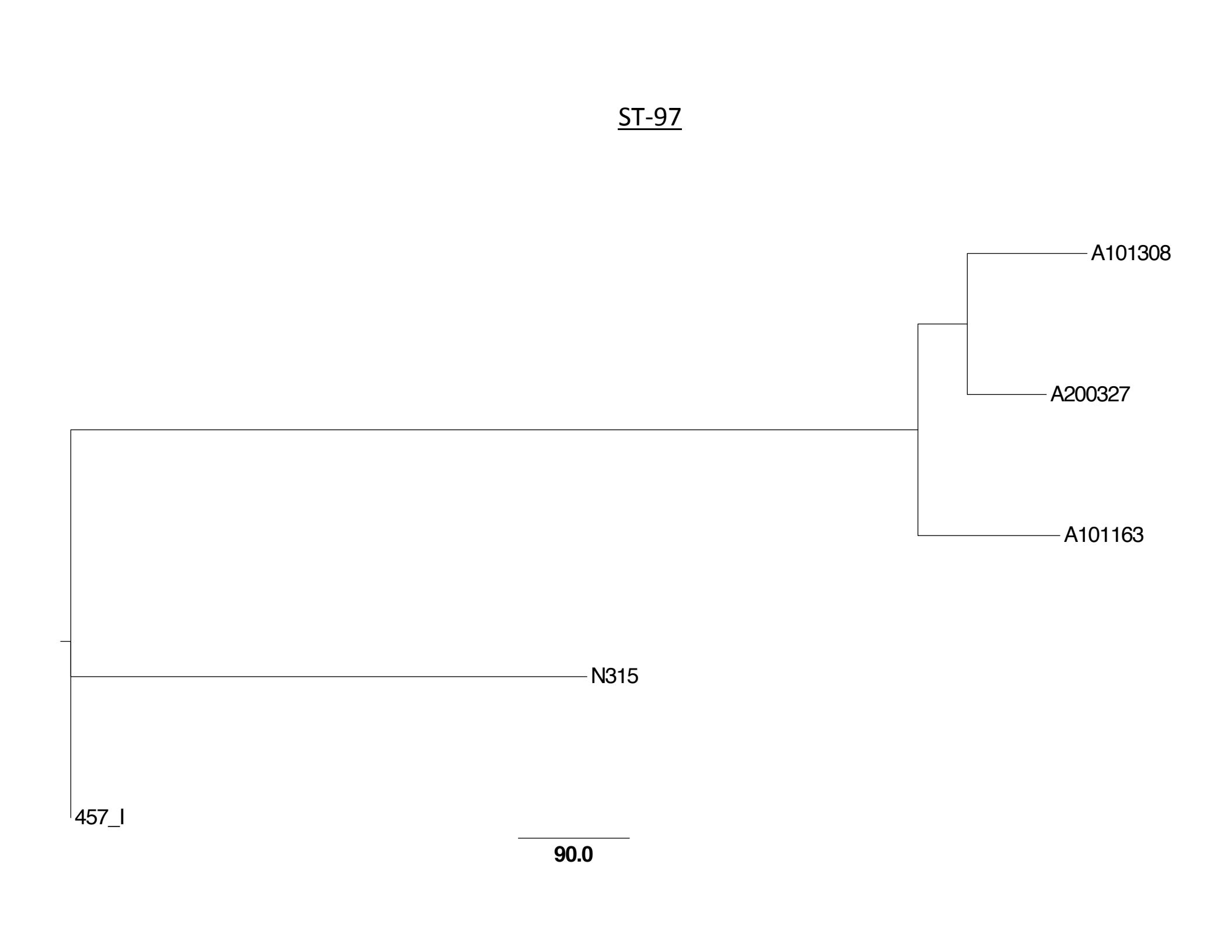

ST-25
A100721
A100831
428_I
Newman
A200292
900.0

ST-30

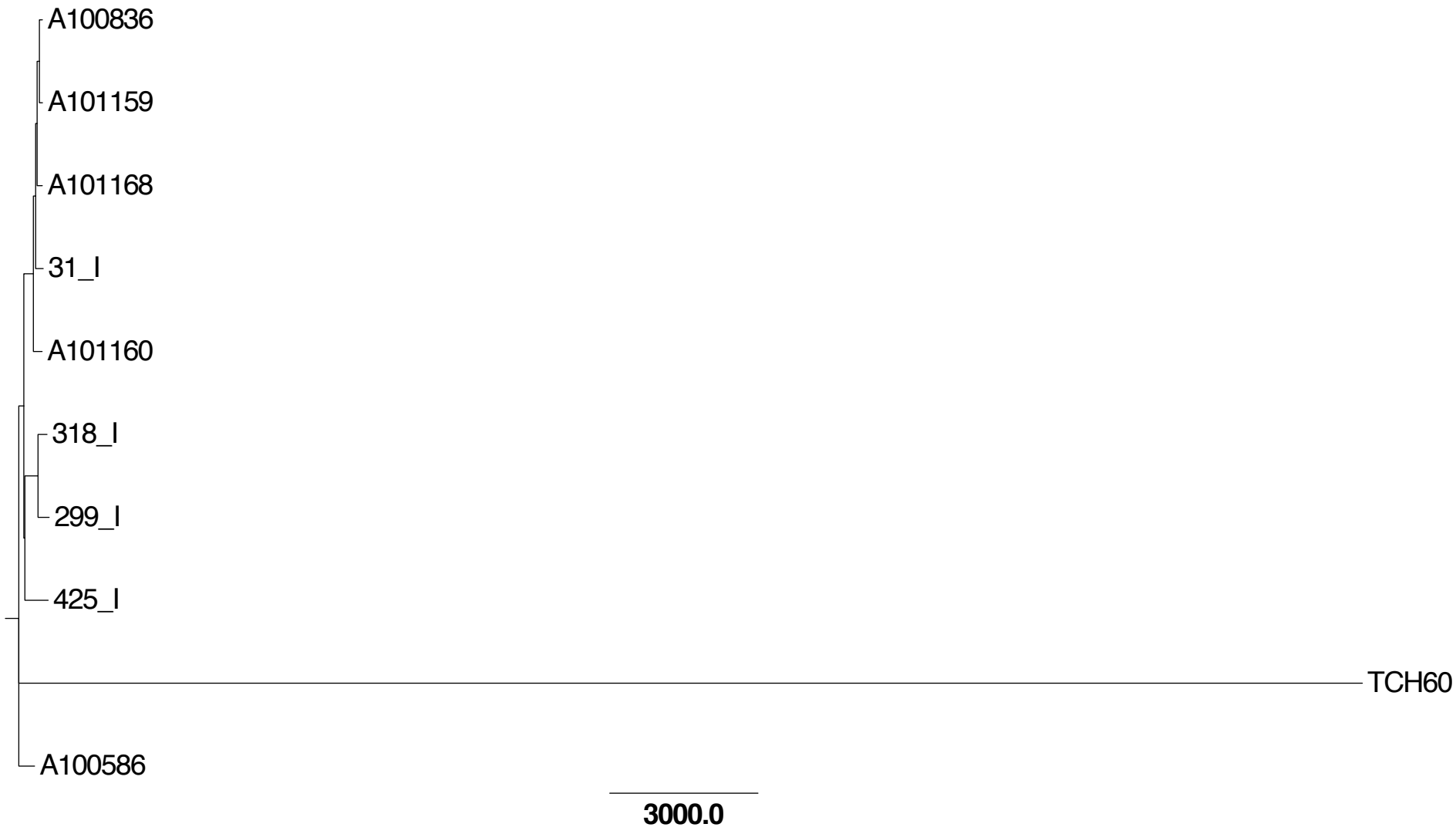

344_I
388_I
426_I
430_I
446_I
707_I
493_I
249_I
679_I
492_I
722_I
269_I
245_I
381_I
A200281
A200089
A200291
A200088
575_I
A200293
400_I
A200099
TCH60
40.0

ST-45


448_I
456_I
A101171
A101161
35_I
339_I
444_I
365_I
438_I
A101166
A101174
A101262
39_I
A100839
A101260
A101127
A101257
A101177
A101157
A101162
A101172
A100834
A200297
754_I
A200080
133_I
A100821
46_I
698_I
41_I
CA_347
50.0


ST-97
A101308
A200327
A101163
N315
457_I
90.0

ST-121


A100836
A101159
A101168
31_I
A101160
318_I
299_I
425_I
TCH60
A100586
3000.0