

European Working Group for *Legionella* Infections
Sequence-Based Typing (SBT) protocol for epidemiological typing of
Legionella pneumophila
Version 4.1

SUMMARY

This procedure describes the European Working Group for Legionella Infections (EWGLI) method for Sequence-Based Typing of *Legionella pneumophila*. Genomic DNA is extracted then amplified using primers targeting seven specific gene loci (i.e. *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, *neuA*). Following purification, amplicons are sequenced directly with forward and reverse primers, and the resulting consensus sequences trimmed and compared to previously assigned allele numbers using the online database. Using a pre-determined order (i.e., *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, *neuA*), the combination of alleles is defined as a 7-digit allelic profile (e.g. 1,4,3,1,1,1,1) and a sequence type represented by a number (e.g., ST1). Putative new allele types can be submitted following the instructions on the website. This method can be used in the epidemiological typing of *L. pneumophila*.

DNA extraction:

The following methods have been found to be reliable and the first two are rapid:

- (i) BIO-RAD InstaGene Matrix (Catalogue no. 732-6030), using 5-10 µl supernatant as template DNA in the primary PCR amplification,
- (ii) Heating emulsified colonies in 0.5 ml sterile water at 100°C for 8 minutes, using 5-10 µl lysate as template DNA in the primary PCR amplification,
- (iii) Nucleon BACC2 DNA extraction kit (Amersham Pharmacia Biotech), using 10-100 ng DNA as template DNA in the primary PCR amplification

SBT targets:

Currently we recommend that all 7 loci (i.e. *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, *neuA*) are determined if possible.

Amplification primers:

The positions of the primers with respect to the GenBank accession number of the reference sequence are shown below:

| Gene | Primer name ¹ | Position ² | Primer sequence (5'-3') | Annealing temperature |
|--------------|--------------------------|-----------------------|-----------------------------------|-----------------------|
| <i>flaA</i> | flaA-587F | 568-587 | GCG TAT TGC TCA AAA TAC TG | 55 °C |
| | flaA-960R | 981-960 | CCA TTA ATC GTT AAG TTG TAG G | |
| <i>pilE</i> | pilE-35F | 12-35 | CAC AAT CGG ATG GAA CAC AAA CTA | 55 °C |
| | pilE-453R | 471-453 | GCT GGC GCA CTC GGT ATC T | |
| <i>asd</i> | asd-511F | 487-511 | CCC TAA TTG CTC TAC CAT TCA GAT G | 55 °C |
| | asd-1039R | 1062-1039 | CGA ATG TTA TCT GCG ACT ATC CAC | |
| <i>mip</i> | mip-74F | 58-74 | GCT GCA ACC GAT GCC AC | 55 °C |
| | mip-595R | 616-595 | CAT ATG CAA GAC CTG AGG GAA C | |
| <i>mompS</i> | mompS-450F | 430-450 | TTG ACC ATG AGT GGG ATT GG | 55 °C |
| | mompS-1126R | 1140-1126 | TGG ATA AAT TAT CCA GCC GGA CTT C | |
| <i>proA</i> | proA-1107F | 1090-1107 | GAT CGC CAA TGC AAT TAG | 55 °C |
| | proA-1553R | 1570-1553 | ACC ATA ACA TCA AAA GCC | |
| <i>neuA</i> | neuA-196F | 176-196 | CCG TTC AAT ATG GGG CTT CAG | 55 °C |
| | neuA-611R | 634-611 | CGA TGT CGA TGG ATT CAC TAA TAC | |

¹ the number in the primer name is the position in the reference sequence where the 3'-terminus of the oligonucleotide binds

² shows the binding positions of the primer with respect to the reference sequence on the website and table

Sequencing primers:

As above except that the *mompS*-1015R primer is used for the reverse sequencing reaction of *mompS*.

| Primer name ¹ | Position ² | Primer sequence (5'-3') |
|--------------------------|-----------------------|-------------------------|
| <i>mompS</i> -1015R | 1032-1015 | CAG AAG CTG CGA AAT CAG |

Please note that the *mompS* forward primer (*mompS*-450F) is a new standard primer and **MUST** be used for the primary amplification and sequencing of the *mompS* target:

for primary amplification:

mompS-450F: 5'-TTG ACC ATG AGT GGG ATTG G-3'

mompS-1126R: 5'-TGG ATA AAT TAT CCA GCC GGA CTT C-3'

for sequencing:

mompS-450F: 5'-TTG ACC ATG AGT GGG ATT GG-3'

mompS-1015R: 5'-CAG AAG CTG CGA AAT CAG-3'

Fragment sizes of amplified products and regions used for allele assignment

| Gene | Fragment size of amplified product (bp) | Size in nucleotides of region used to determine allele type | Region used for allele assignment | GenBank accession number of reference sequence |
|--------------|---|---|-----------------------------------|--|
| <i>flaA</i> | 394 | 182 | 653-749 | X83232 |
| <i>pilE</i> | 459 | 333 | 103-435 | AF048690 |
| <i>asd</i> | 575 | 473 | 538-1010 | AF034213 |
| <i>mip</i> | 558 | 402 | 117-518 | AJ496265 |
| <i>mompS</i> | 710 | 352 | 523-1010 | AF078136 |
| <i>proA</i> | 480 | 405 | 1134-1230 | M31884 |
| <i>neuA</i> | 459 | 354 | 229-583 | A6017354 |

Oligonucleotide storage and handling recommendations:

Appropriate storage of oligonucleotides will extend their shelf life. Most oligonucleotides are supplied dry (lyophilised) and manufacturers (e.g., MWG Biotech AG, Germany) usually supply detailed information on appropriate resuspension and storage conditions. Information on data-sheets and instructions from the manufacturers provided with the oligonucleotides should be read, noted and followed. To maximise the shelf-life of oligonucleotides it is recommended that concentrated "stock solutions" (e.g., 100pmol/ μ l) are made in 1x nuclease-free TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5-8.0) and stored below -20°C, and dilute "working solutions" (e.g., 10 pmol/ μ l) made in nuclease-free 10 mM Tris pH 7.5-8.0. Working solutions are frozen in 50 μ l aliquots and should **NOT** be subjected to repeat freeze-thaw cycles and must be kept for **no longer than one week at 4°C** before discarding.

Primary amplification:

PCR amplification is performed in a total volume of 50 μ l. Place PCR plate or tubes on a cold block. Start with adding 10 μ l of Taq DNA polymerase (2.5 units/reaction) to each PCR tube or 96 well plate, then add 35 μ l of the master mix, finally add the DNA template (5 μ l /tube or /well). Gently spin tubes/plates briefly in an appropriate centrifuge before placing them in a thermal cycler heating block.

PCR reaction mix

| Reagent | Stock concentration | Final concentration | Volume/reaction (μ l) |
|---------------------|---------------------|---------------------|----------------------------|
| PCR buffer | 10 x | 1 x | 5.0 |
| MgCl ₂ | 50 mM | 2.5 mM | 2.5 |
| Primer 1 | 10 pmol/ μ l | 10 pmoles | 1.0 |
| Primer 2 | 10 pmol/ μ l | 10 pmoles | 1.0 |
| dNTPs | 5 mM | 200 μ M | 2.0 |
| Nuclease-free water | | | 23.5 |
| Volume/reaction | | | 35.0 |

Prepare *Taq* DNA polymerase by adding:

| | |
|--|--------------|
| Nuclease-free water | 9.0 μ l |
| PCR buffer (x10) | 0.5 μ l |
| <i>Taq</i> DNA polymerase (5 units/ μ l) | 0.5 μ l |
| Volume/reaction | 10.0 μ l |

Controls

For each PCR run include a negative (5 μ l of nuclease-free water) and a positive control (5 μ l of known DNA template). Each sample is amplified in duplicate. Pipette solutions in the order of: negative control, test samples, positive control.

Thermal cycler parameters

| | Time | Temperature (°C) | No. cycles |
|-----------------|--------|---------------------|------------|
| Denaturation | 5 min | 95°C | 1 |
| Denaturation | 30 sec | 95°C | } 35 |
| Annealing | 30 sec | 55° C | |
| Extension | 30 sec | 72°C | |
| Final extension | 5 min | 72°C | 1 |

Primary PCR amplicon purification

Duplicate PCR products are pooled and then purified using: Montage PCR₉₆ filter plates (Millipore); Qiaquick PCR Purification Kit (Qiagen); Wizard PCR Preps Purification System (Promega) or equivalent.

Analysis of purified PCR products by E-Gel electrophoresis

Run PCR products (2 μ l) + 1x BlueJuice (Invitrogen) (18 μ l). Also run E-Gel low range quantitative DNA ladder (Invitrogen) (10 μ l) + nuclease free water (10 μ l). After 20 minutes of run, the E-Gel is analysed using the UVP gel imaging system. If positive and negative controls as well as the test sample results are satisfactory, proceed with the DNA sequencing reaction.

DNA Sequencing using the CEQ 8000 Genetic Analysis System (Beckman Coulter)

DNA Sequencing materials

Dye Terminator Cycle Sequencing is undertaken using the DTCS Quick Start Kit (Beckman Coulter). Follow the manufacturer's instructions carefully. The Sample Loading Solution contains formamide. Formamide waste (i.e., sequencing reactions and plates) should **NOT** be autoclaved and must be disposed for direct incineration.

The CEQ DTCS Quick Start Kit contains the following reagents (sufficient for 100 reactions):

- 1) Quick Start Mix (800 μ l)
dATP, dCTP, dTTP, dITP
ddUTP, ddGTP, ddCTP, ddATP (WellRED label)
Tris-HCl, MgCl₂, reaction buffer - pH 8.9
Thermo Sequenase DNA Polymerase
Pyrophosphatase
- 2) 47 Sequencing Primer (240 μ l)
- 3) pUC18 Control Template (20 μ l)
- 4) Glycogen (110 μ l)
- 5) Mineral Oil (5 ml)
- 6) Sample Loading Solution (SLS) 6 ml

DNA Sequencing procedure

DNA sequencing reactions are prepared **following the manufacturer's instructions** by combining the following:

- 1) PCR-grade water (e.g., Nuclease-Free Water, W4502, Sigma)
- 2) DNA template
- 3) User supplied sequencing primer or (-) 47 Sequencing Primer (1.6 pmol/ μ l or 1.6 μ M)
- 4) DTCS Premix (Beckman Coulter)

N.B. Can run half-volume reactions (total of 10 μ l/reaction).

Thermocycler parameters (for DNA sequencing):

| Temperature (°C) | Time | No. cycles |
|------------------|--------|------------|
| 96 | 20 sec | } 30 |
| 50 | 20 sec | |
| 60 | 4 min | |
| 4 | Hold | |

N.B. The above parameters are those recommended for the CEQ 8000 Genetic Analysis System (Beckman Coulter). For alternative sequencing platforms the relevant manufacturer's recommendations should be followed.

Ethanol precipitation is carried out as per manufacturer's instructions using a suitable centrifuge. For ethanol plate precipitation in a CEQ sample plate, please refer to the "Dye terminator cycle sequencing chemistry protocol", page 13. For ethanol precipitation in 1.5 ml eppendorf tubes, please refer to the manufacturer's instructions provided with the DTCS Quick Start Kit (Beckman Coulter).

N.B. for half volume reactions add 10 µl of H₂O, before proceeding to ethanol precipitation.

Sample preparation for loading into the CEQ

Add 55 µl of Sample Loading Solution to each well, leave on plate-shaker for 10 minutes. Overlay each re-suspended sample with one-drop of mineral oil.

From this point follow the instructions provided by the manufacturer of the Sequencing Instrument.

Sequence analysis:

EWGLI Sequence Quality Tool

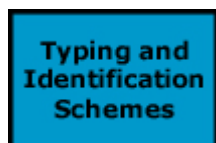
Forward and reverse sequence trace files (scf. or abi.) of all seven targets are submitted to the “**Sequence Quality Tool**” which can be accessed from the EWGLI website (www.ewgli.org). This tool assembles contig(s) from the traces, finds start and end (reference) positions in the contig, trims the contig using these positions and finally matches the trimmed sequence against those in the SBT database. The tool identifies individual alleles as well as an allelic profile and a sequence type (ST). Sequences with <100% match are identified as -1, and the position of mismatches are also indicated. Sequences of poor quality should either be repeated or further analysed using a DNA analysis software e.g., BioNumerics (Applied Maths).

EWGLI SBT Database

- Data from both forward and reverse sequencing reactions are combined and aligned to produce a consensus sequence. Reference sequences trimmed to the correct length can be downloaded from the website in order to aid contig assembly. Consensus sequences trimmed to the correct length are submitted to the SBT database as flat text file. The database returns an allele type eg., 1, when submitted sequence shows a 100% match to a pre-designated allele type. Sequences with <100% match are identified as the closest match to a pre-existing allele type with the number of mismatches specified. The SBT database also returns an alignment with mismatches highlighted.
- For each isolate, the combination of alleles at each of the loci is defined as the allelic profile using a pre-determined order, i.e., *flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, *neuA*. For example, for strain EUL no. 120, the allelic profile is 4,7,11,3,11,12,9.
- If an individual allele number has not been determined, a zero is entered into the allelic profile, thus maintaining its integrity. For example, if the *proA* allele number was not determined for the examples above, the profile would be 4,7,11,3,11,0,9 and if the *mompS* allele was not determined, it would be 4,7,11,3,0,12,9.

Instructions for accessing the EWGLI SBT Website

- Go to (www.ewgli.org)
- Click the “**Typing and identification schemes**” button:



Follow the link for the EWGLI Sequence-Based Typing (SBT) Database for *Legionella pneumophila*

Instructions for submitting sequence data to the EWGLI Sequence Quality Tool

- Under Query Functions, click on the “**Sequence Quality Tool**”
- Choose the number of sequence trace files to be uploaded.
- Click on the “**Browse**” button to download sequence files.
- Forward and reverse sequences must be downloaded for each allele in standard file format (*.scf) or ABI trace file format (*.abi).
- Click the “**Analyse Trace Files**” button.
- The tool also produces a sequence quality report for each uploaded contig and generates a seven-figure allelic profile (e.g., 3,4,1,1,14,9,1), as well as indicating whether the obtained allelic profile is of a novel combination. The tool also identifies a single designated sequence type (ST) for a pre-existing allelic profile.
- You can print a summary sheet by clicking the link: “**Print Friendly Summary**”
- For further details on the sequence quality tool please click on this link: “**explanation and frequently asked questions**”

- Putative novel alleles must be submitted for verification following the link “**New Allele Submission**”.
- To submit a novel allelic profile click on the link “**Strain Data Submission**”.

Instructions for submitting sequence data to the EWGLI SBT Database

- Under Query Functions, click on the “**Check a sequence for it’s allele number**”
- Before pasting in a query sequence the user must ensure that the consensus sequence is trimmed to the correct length.
- Reference sequences trimmed to the correct position and length can be downloaded to aid contig assembly.
- Consensus sequences of the correct length can be submitted for more than one target, alternatively you can submit consensus sequences one at a time before hitting the “**upload sequence**” button.
- Please note that sequences of incorrect length will be rejected.
- Sequences of correct length and with a 100% match to a sequence in the database will return an allele number e.g., “1”.
- Sequences of correct length and with <100% match to a sequence in the database will return an alignment with the closest matches showing where differences are.
- To enter consensus sequences for another strain, click the “**Reset form**” button and proceed as before.

N.B. The website is under constant development, so please see the website for any changes, new tools and to ensure you are using the latest version of this protocol etc.

References

Ratzow S, Gaia V, Helbig JH, Fry NK, Lück PC (2007). Addition of *neuA*, the gene encoding N-acetylneuraminase cytidyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. *J Clin Microbiol* **45**,1965-1968.

Gaia V, Fry NK, Afshar B, Lück PC, Meugnier H, Etienne J, Peduzzi R, Harrison TG (2005). A consensus sequence-based epidemiological typing scheme for clinical and environmental isolates of *Legionella pneumophila*, *J Clin Microbiol* **43**, 2047-2052.

Gaia V, Fry NK, Harrison TG, Peduzzi R (2003). Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. *J Clin Microbiol* **41**, 1491-1502.

For further information please contact Massimo Mentasti or Norman Fry by email.

Massimo.Mentasti@HPA.org.uk

Norman.Fry@HPA.org.uk