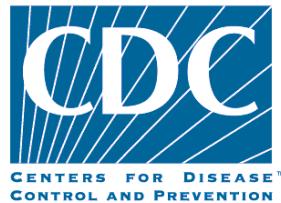


# **Logically Inferred Tuberculosis Transmission (LITT) Algorithm**

User's Manual

Version 1.1, Released November 13, 2020



# Contents

<b>Introduction.....</b>	<b>3</b>
<b>Intended use.....</b>	<b>3</b>
<b>Glossary .....</b>	<b>4</b>
<b>Acronyms.....</b>	<b>7</b>
<b>Algorithm overview .....</b>	<b>8</b>
<b>Preparing data.....</b>	<b>9</b>
<i>A. Data overview.....</i>	<i>9</i>
<i>B. Case data table .....</i>	<i>12</i>
<i>C. Epi link table.....</i>	<i>22</i>
<i>D. SNP distance matrix.....</i>	<i>26</i>
<i>E. Table of risk factor weights.....</i>	<i>29</i>
<b>Performing an analysis.....</b>	<b>30</b>
<i>A. Accessing the algorithm.....</i>	<i>30</i>
<i>B. Uploading input files .....</i>	<i>32</i>
<i>C. Setting analysis parameters.....</i>	<i>35</i>
<i>D. Initiating an analysis.....</i>	<i>36</i>
<b>Using analytic outputs .....</b>	<b>40</b>
<i>A. Interpreting analytic outputs.....</i>	<i>40</i>
<i>B. Using analytic outputs .....</i>	<i>58</i>
<i>C. Limitations.....</i>	<i>59</i>
<b>Training datasets.....</b>	<b>60</b>
<b>Contacts and additional resources.....</b>	<b>61</b>
<b>Citations.....</b>	<b>61</b>
<b>Acknowledgments .....</b>	<b>62</b>
<b>Appendix 1 — Accessing the OAMD web portal.....</b>	<b>62</b>
<b>Appendix 2 — Calculation of time rating scores .....</b>	<b>68</b>
<b>Appendix 3 — Using MicrobeTrace to visualize transmission networks .....</b>	<b>69</b>

# Introduction

The Logically Inferred Tuberculosis Transmission algorithm, or LITT, is a free analytic tool created in the programming language R [1] by staff of the Centers for Disease Control and Prevention (CDC) to facilitate the investigation of clusters of tuberculosis (TB) disease. No program files beyond a web browser need to be downloaded or installed on the user's computer to perform a LITT analysis. LITT automates the integration of information from multiple data streams (e.g., disease surveillance, clinical summaries, epidemiologic investigations, and whole genome sequencing (WGS) with phylogenetic analysis) to identify and rank potential sources of infection for individual TB cases within clusters. LITT applies a hierarchy of evaluation criteria, including filtering, scoring, and weighting of input data as well as decision rules modelled on those applied by U.S. public health practitioners investigating TB clusters and outbreaks of TB disease in the field [2]. By enabling the rapid and systematic evaluation of large volumes of disparate data within a formalized analytic framework, LITT can be used to generate timely, iterative, repeatable assessments of epidemiologic data collected as part of TB outbreak investigations. Additionally, use of LITT promotes the systematic formatting and organization of these data in ways that can simplify and expedite cluster and outbreak investigations.

## Intended use

*LITT is available free from the CDC without warranty or guarantee of any kind as to the appropriateness of its application or the accuracy of resulting outputs. CDC technical assistance with LITT may not always be available for all user analyses; the availability of these support services may be discontinued at any time.*

LITT was developed in the United States, a low TB-burden country. LITT is intended to be used by public health practitioners conducting investigations of clusters of TB disease in the field where chains of TB transmission may be more readily discernible. While it can be applied retrospectively to evaluate conclusions and actions associated with historical investigations, it is primarily envisioned as a tool to inform ongoing investigations. LITT can also be used by persons who are indirectly supporting field investigations (e.g., state health department staff collaborating with county-based colleagues investigating a cluster in their local jurisdiction) provided that epidemiologic data generated by the investigation and required for a LITT analysis can be shared. Similarly, academic researchers can use LITT to study characteristics of both clusters of TB disease and of associated epidemiologic investigations.

LITT should supplement, not replace, the careful and thorough collection, management, and analysis of data and samples achieved by investigations. While LITT can characterize possible and probable transmission pathways within TB clusters (i.e., likely sources of infection for individual cases), investigators should always review and compare results of a LITT analysis with insights generated using frontline epidemiologic and molecular investigation methods. Furthermore, LITT outputs are only as good as the quality (e.g., accuracy, completeness) of the data used to generate them. For example, missing epidemiologic links or imprecisely calculated dates may compromise the accuracy and value of LITT results. Beyond being able to interpret LITT outputs, users should become familiar with the algorithm's decision rules to understand the various ways that LITT outputs are constrained by assumptions and limitations of the underlying analytic framework [2].

Within this user's manual, explicit names of variables appear in **green font**, explicit names of files, components of files (e.g., worksheets within Microsoft Excel files), webpages, and features of webpages (e.g., check boxes) appear in **blue font**, and hyperlinks appear as **underlined italicized** text.

## Glossary

**Acid-fast bacteria (AFB):** a class of bacteria, including *Mycobacterium tuberculosis*, the etiological agent of tuberculosis disease, that are resistant to decolorization by acids during laboratory staining procedures.

**Case:** a person infected with *M. tuberculosis*, the etiological agent of TB disease, who is experiencing TB disease. Note that more accurately, a case refers to an instance of disease in a person (i.e., “patient X had a case of tuberculosis disease in 2018”). To be consistent with more colloquial usage of the term, case is used to refer to a person rather than an episode of disease within LITT and associated documentation (including this user's manual).

**Cluster:** a group of cases that are known or presumed to be epidemiologically linked by recent *M. tuberculosis* transmission (e.g., based on results of genotyping of pathogen isolates from sputum samples).

**Contact:** a person with known or presumed exposure to an infectious case. Contacts can be identified through name-based (i.e., one person names another) and location-based (i.e., records or interview

findings indicate that persons were in the same place (e.g., homeless shelter, workplace) at the same time) methods. The goal of a contact investigation is to generate a list of contacts for follow up.

**Epidemiologic (epi) link:** two persons known or presumed to have had direct or indirect contact at a location are said to share an epidemiologic, or epi, link. Epi link is a commonly used shorthand for the formal term epidemiologic link. Epi links are identified during contact investigations and generally speaking can vary in strength based on the proximity, duration, or frequency of contact between the persons. As part of LITT analyses the strength of epi links is assigned based on the duration that two persons overlapped and the confidence the user has in the dates of overlap. For additional information about defining epi link strength see Box 2 *below*.

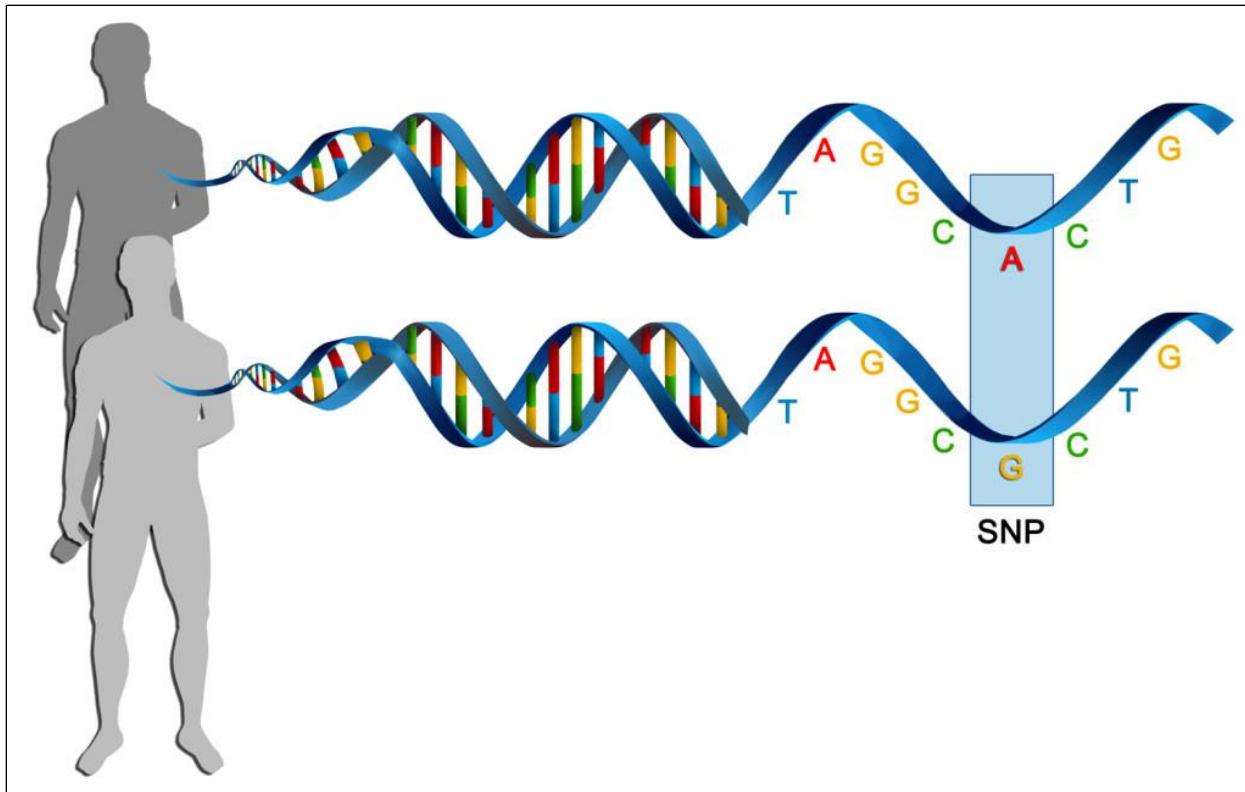
**Given case:** of a pair of cases in a cluster, the case for which LITT is attempting to identify a possible source of infection.

**Most recent common ancestor (MRCA):** a hypothesized genome type, inferred as part of a whole-genome single nucleotide polymorphism (wgSNP) analysis, from which all other *M. tuberculosis* isolates from cases in a cluster are believed to be descendant. The MRCA serves as a reference point for examining the direction of genetic change within phylogenetic trees.

**Potential source case:** of a pair of cases in a cluster, the case being considered by LITT as a possible source of infection for the given case.

**Risk factor:** any attribute, characteristic, or exposure within a cluster that may increase the likelihood of transmission of *M. tuberculosis* or the development of TB disease. As examples, spending time in a congregate setting (e.g., homeless shelter, correctional facility) is often a risk factor for infection/transmission, while positive HIV status is a risk factor for development of tuberculosis disease following infection. Ideally, shared risk factors should be considered epidemiologically relevant (i.e., important for or contributing to pathogen transmission, infection, symptom expression, etc.) and not simply shared characteristics. Risk factors can be identified through routine surveillance or during contact, cluster, or outbreak investigations.

**Single nucleotide polymorphism (SNP):** base pair variation at a single position (locus) when two *M. tuberculosis* deoxyribonucleic acid (DNA) sequences are compared (*Figure 1*).



**Figure 1. Diagram illustrating a single nucleotide polymorphism (SNP) [3].**

**SNP distance:** the number of SNPs identified when corresponding gene sequences from two *M. tuberculosis* isolates are compared. SNP distance is considered a measure of the genetic relatedness of the isolates.

**Transmission link:** two persons share a transmission link when they share an epidemiologically linked, the timing of their overlap was during the infectious period of one of the persons (i.e., case), and there is evidence (e.g., results of genotyping of pathogen isolates from sputum samples from both persons) that this person transmitted *M. tuberculosis* to the other person.

**Whole-genome sequencing (WGS):** method of genetic analysis in which the complete or nearly complete DNA sequence of an organism is determined.

**Whole-genome single nucleotide polymorphism (wgSNP) analysis:** method of genetic analysis based upon the number of single nucleotide polymorphisms (SNPs) identified when corresponding gene sequences from two or more *M. tuberculosis* isolates are compared.

## **Acronyms**

The following is a list of acronyms that appear in this user's manual

**AFB:** acid-fast bacilli

**CDC:** U.S. Centers for Disease Control and Prevention

**CITGO:** CDC information technology on the go

**DNA:** deoxyribonucleic acid

**DTBE:** Division of Tuberculosis Elimination

**LITT:** Logically Inferred Tuberculosis Transmission

**MRCA:** most recent common ancestor

**NTCA:** National Tuberculosis Controllers Association

**OAMD:** Office of Advanced Molecular Detection

**OCIO:** Office of the Chief Information Officer

**SAMS:** Secure Access Management System

**SNP:** single nucleotide polymorphism

**TB:** tuberculosis

**TB GIMS:** Tuberculosis Genotyping Information Management System

**VPN:** virtual private network

**WGS:** whole-genome sequencing

**wgSNP:** whole-genome single nucleotide polymorphism (analysis)

## Algorithm overview

The principal objective of a LITT analysis is to identify possible and probable transmission pathways within TB clusters. For each case in a cluster, LITT attempts to identify the most likely source of infection from among other cases in the cluster. In an iterative process, LITT considers each case as the given case and sequentially creates pairwise combinations of this given case and every other potential source case in the cluster. For each pair of cases, LITT assesses the likelihood that the potential source case was the source of infection for the given case. This is accomplished in a two-step process.

### Step 1. Filtering out cases that could not have been the source of infection

LITT filters out cases that could not have been the source of infection for a given case under consideration. These include:

- *Cases that did not have pulmonary or laryngeal disease:* Because they are not considered infectious except in rare circumstances, LITT filters out cases that did not have pulmonary or laryngeal disease.
- *Pediatric cases:* Because they are not considered infectious except in rare circumstances, LITT filters out cases in children <10 years of age.
- *Cases that were not closely related by whole genome sequencing:* LITT filters out cases that were genetically distant to the given case under consideration based on wgSNP analysis and the SNP cutoff specified by the user as a parameter for the LITT analysis.
- *Cases that were infectious much earlier or later in time than when the given case likely became infected.* Please see [Appendix 2](#) for additional details.

### Step 2. Scoring and ranking

From among the remaining cases (i.e., those not filtered out as a potential source of infection for a given case under consideration), LITT assesses the likelihood that a particular potential source case was the source of infection for the given case under consideration based on four ratings:

- *Infectious rating*: How infectious was the potential source case in terms of sputum smear result and cavitary status?
- *SNP rating*: How closely related were the *M. tuberculosis* isolates from the potential source case and the given case under consideration based on wgSNP analysis (i.e., SNP distance)?
- *Time rating*: To what extent did the known or estimated infectious period of the potential source case precede, overlap, or follow the known or estimated infectious period start date of the given case under consideration?
- *Epi and risk factor rating*: To what extent were the potential source case and given case under consideration epi linked or similar in terms of shared and relevant risk factors for transmission?

For each pairwise combination of given and potential source cases, LITT assigns a score that reflects the likelihood that the potential source case was the source of infection for the given case. The score is calculated as the sum of values of the four ratings, which can have values ranging from zero to 7. The magnitude of the score is inversely proportional to the likelihood of infection, with a lower score indicating a higher likelihood that the potential source case infected the given case. Based on these scores, cases are then ranked according to the relative likelihood that they are a potential source. Details of the LITT evaluation criteria and scoring system will be detailed in a forthcoming peer-reviewed publication [2]. By elucidating pathways of disease spread, LITT analyses can help to focus public health actions and resources on persons, settings, or activities driving the spread of infection within transmission networks.

## Preparing data

### A. Data overview

Four types of data, each uploaded as a separate input file, can be included in a LITT analysis:

1. **Case data table**: this input file contains information on demographic and epidemiologic characteristics of each case included in the cluster being analyzed.
2. **Epi link table**: this input file contains qualitative information about the known epidemiologic links between some or all pairs of cases included in the cluster being analyzed.
3. **SNP distance matrix**: this input file contains quantitative information about the genetic distance between *M. tuberculosis* isolates collected from some or all cases included in the cluster being analyzed as determined by WGS analyses.

4. **Table of risk factor weights:** this input file contains quantitative information about risk factors associated with some or all cases included in the cluster being analyzed.

The [Case data table](#) input file and either the [Epi link table](#) input file or the [SNP distance matrix](#) input file must be uploaded to perform a LITT analysis. Including both the [Epi link table](#) and the [SNP distance matrix](#) input files, or these two and the [Table of risk factor weights](#) input file, should improve the accuracy and completeness of LITT outputs, assuming that these files contain high quality data (e.g., accuracy, completeness).

A separate input file template (Microsoft Excel format) is available for the [Case data table](#), [Epi link table](#), and [Table of risk factor weights](#) input files. These templates can be downloaded from the [TB molecular epidemiology GitHub](#) webpage, which can be accessed using this [\*link\*](#) or through a link on the LITT online user interface ([\*see below\*](#)). Each input file template contains a data worksheet (named [data](#)) into which the data to be analyzed are entered, and a metadata worksheet (named [metadata](#)) that provides information about the names, definition, and formatting of variables included in the [data](#) worksheet. The [data](#) worksheet must be in the first position (i.e., to the left of the [metadata](#) worksheet) as this is the worksheet position from which LITT pulls data for the analysis.

Each input file template contains up to three different types of variables. Required variables (columns shaded green) include variables for which LITT requires a value in order to run. If a cell in a column associated with a required variable is left blank, either a default value will be assigned (so that the cell is no longer blank) or the associated row of data will be deleted. Conditionally required variables (columns shaded blue) include variables for which LITT requires a value in order to run but only if one or more other variables are associated with specific conditional values. If a cell in a column associated with a conditionally required variable is left blank, and one or more other variables are associated with the conditional values, the associated row of data will be deleted. Optional variables (columns shaded yellow) include variables for which LITT does not require a value in order to run. If a cell in a column associated with an optional variable is left blank, the cell will remain blank and the analysis will proceed. The one exception is the risk factor variable(s), which are not required for a LITT analysis and can be optionally added by the user. If one or more columns for risk factor variables are added by the user, and any cells in these columns are left blank, a default value (“N”) will be assigned.

The [SNP distance matrix](#) input file will be generated by CDC upon request from LITT users. Instructions for how to request a [SNP distance matrix](#) input file for a LITT analysis are provided [\*below\*](#). Note that

CDC technical assistance with LITT may not always be available for all user analyses; the availability of these support services may be discontinued at any time.

**Note:** All screenshots and discussion of analytic outputs presented in this user's manual are based on a LITT analysis of the input files associated with [LITT training dataset 01](#). This analysis can be replicated by downloading the [LITT training dataset 01](#) input files (see section on training datasets *below*) and using them to perform a LITT analysis as described in this user's manual. Additionally, screenshots of the LITT online user interface presented in this user's manual were taken while running LITT in the Google Chrome web browser. Note that LITT seems to consistently perform well if any issues when used in the Chrome web browser; the LITT online user interface may appear slightly different in other browsers.

**IMPORTANT: DO NOT INCLUDE PERSONALLY IDENTIFIABLE INFORMATION (PII) ABOUT CASES**

When populating the [Case data table](#), [Epi link table](#), and [Table of risk factor weights](#) input files, and/or when editing content of the [SNP distance matrix](#) input file (e.g., changing values for [Case ID](#) after receiving the file from CDC and before uploading files to CDC servers for a LITT analysis), deidentified data should be used. Personally identifiable information about cases should never be uploaded to CDC servers as part of a LITT analysis. This is because any information uploaded to a CDC server might be considered a federal record and therefore be subject to disclosure under the federal Freedom of Information Act or other laws. Do not use the real and complete names of cases as values for [Case ID](#) or include the real and complete names of locations (e.g., schools, shelters, workplaces), especially in geographically small areas. Further, be mindful of situations in which combinations of data can allow for personal identification. Users should endeavor to deidentify potentially identifying information whenever possible (e.g., if the location is a restaurant, consider assigning an identifier like "Restaurant A" or even "Location A" instead of using the name of the restaurant). A separate key to allow translation of generic identifiers back to actual names can be maintained by the user as a separate file that is stored locally (i.e., not on a CDC server).

## **B. Case data table**

The [Case data table](#) input file contains information on demographic and epidemiologic characteristics of each case included in the cluster being analyzed. A screenshot of the [Case data table](#) input file from [LITT training dataset 01](#) provides an example of how a populated [data](#) worksheet might appear ([\*Figure 2\*](#)).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Case ID	Pediatric	XRAYCAV	SPSMEAR	Extrapulmonary Only	Infectious Period Start	Infectious Period End	Infection Acquisition Start	Infection Acquisition End	Symptom Onset	Investigation Presumed Source	Investigation Presumed Source Strength	Homelessness	HIV_pos	Inj_drug_use	Sex	Country_origin	Race_ethnicity
2	Case 1	N	N	NEG	Y				02/01/18				N	N	N	Female	US born	Asian
3	Case 2	Y	N	NEG	N			04/01/16	05/01/18		Case 3	Definite	N	N	N	Female	US born	Asian
4	Case 3	N	Y	POS	N	09/01/17	05/31/18			12/01/17			Y	Y	N	Male	non-US born	Asian
5	Case 4	N	Y	POS	N	01/01/18	07/15/18			04/01/18			Y	Y	Y	Male	non-US born	Asian
6	Case 5	N	Y	POS	N	06/20/18	11/15/18			09/20/18			Y	N	Y	Male	US born	White
7	Case 6	N	N	POS	N	05/15/18	08/01/18			08/15/18			Y	N	Y	Male	US born	White
8	Case 7	N	Y	POS	N	02/20/18	07/15/18			05/20/18			Y	Y	Y	Female	US born	White
9	Case 8	N	Y	POS	N	06/15/18	10/01/18			09/15/18	Case 7	Probable	Y	Y	Y	Male	non-US born	Black

	A	B	C	D	E	F	G	H	I
1	Case ID	Pediatric	XRAYCAV	SPSMEAR	Extrapulmonary Only	Infectious Period Start	Infectious Period End	Infection Acquisition Start	Infection Acquisition End
2	Case 1	N	N	NEG	Y				02/01/18
3	Case 2	Y	N	NEG	N				04/01/16
4	Case 3	N	Y	POS	N	09/01/17	05/31/18		
5	Case 4	N	Y	POS	N	01/01/18	07/15/18		
6	Case 5	N	Y	POS	N	06/20/18	11/15/18		
7	Case 6	N	N	POS	N	05/15/18	08/01/18		
8	Case 7	N	Y	POS	N	02/20/18	07/15/18		
9	Case 8	N	Y	POS	N	06/15/18	10/01/18		

	J	K	L	M	N	O	P	Q	R
Symptom Onset	Investigation Presumed Source	Investigation Presumed Source Strength	Homelessness	HIV_pos	Inj_drug_use	Sex	Country_origin	Race_ethnicity	
			N	N	N	Female	US born	Asian	
	Case 3	Definite	N	N	N	Female	US born	Asian	
12/01/17			Y	Y	N	Male	non-US born	Asian	
04/01/18			Y	Y	Y	Male	non-US born	Asian	
09/20/18			Y	N	Y	Male	US born	White	
08/15/18			Y	N	Y	Male	US born	White	
05/20/18			Y	Y	Y	Female	US born	White	
09/15/18	Case 7	Probable	Y	Y	Y	Male	non-US born	Black	

**Figure 2.** Screenshot of the **data** worksheet in the **LITT training dataset 01 Case data table** input file (top panel) and below it, magnified images of left-hand portion (middle panel) and right-hand portion (bottom panel) of the **data** worksheet.

The **data** worksheet in the **Case data table** input file includes five required variables (green columns), three conditionally required variables (blue columns), and four or more optional variables (yellow columns) described below.

### **Required variables (green columns):**

1. **Case ID:** unique case identifier. If a cell in the **Case ID** column is left blank, LITT will delete the associated row of data during the analysis. Note that values of **Case ID** must match exactly in all input files. In the **SNP distance matrix** input file provided to users by CDC, accession numbers from the CDC's Tuberculosis Genotyping Information Management System (TB GIMS) will generally be used as values of **Case ID**. In order to maintain exactly matching values of **Case ID** across input files, users will need to either use accession numbers as values of **Case ID** in all relevant input files or edit the case identifiers in the **SNP distance matrix** input file to reflect the identifier of choice. If editing values of **Case ID** in the **SNP distance matrix** input file, **do not include personally identifiable information about cases.**
2. **Pediatric:** status of the case as to being pediatric (less than 10 years of age). Options are “Y” (case is less than 10 years of age) or “N” (case is 10 years of age or older). If a cell in the **Pediatric** column is left blank, LITT will assign a default value of “N” during the analysis.
3. **XRAYCAV:** status of the case as to having one or more chest radiographs showing evidence of one or more lung cavities. Options are “Y” (case had one or more chest radiographs showing evidence of one or more lung cavities) or “N” (case had no chest radiographs showing evidence of one or more lung cavities). If a cell in the **XRAYCAV** column is left blank, LITT will assign a default value of “N” during the analysis.
4. **SPSMEAR:** status of the case as to having results of any sputum smear examinations that were positive for any acid-fast bacilli (AFB), of which *M. tuberculosis* is one. Options are “POS” (case had one or more sputum smear examinations that were positive for any AFB) or “NEG” (case had no sputum smear examination that was positive for any AFB). If a cell in the **SPSMEAR** column is left blank, LITT will assign a default value of “NEG” during the analysis.
5. **Extrapulmonary Only:** status of the case as to having only extrapulmonary TB (i.e., does not have either pulmonary or laryngeal disease). Options are “Y” (case only has extrapulmonary disease and does not have either pulmonary or laryngeal disease) or “N” (case has pulmonary disease, laryngeal disease, or a combination of either/both of these and extrapulmonary disease). If a cell in the **Extrapulmonary Only** column is left blank, LITT will assign a default value of “N” during the analysis.

### **Conditionally required variables (blue columns):**

1. **Infectious Period Start:** start date of the case's infectious period. This variable is only required and can only be calculated for infectious cases (i.e., adult cases with pulmonary or laryngeal disease). To calculate **Infectious Period Start**, use recommendations from the National Tuberculosis Controllers Association (NTCA) and CDC [4] summarized in *Figure 3*.

<b>TABLE 2. Guidelines for estimating the beginning of the period of infectiousness of persons with tuberculosis (TB), by index case characteristic</b>				
Characteristic				
TB symptoms	AFB* sputum smear positive	and	Cavitory chest radiograph	Recommended minimum beginning of likely period of infectiousness
Yes	No	and	No	3 months before symptom onset or first positive finding (e.g., abnormal chest radiograph) consistent with TB disease, whichever is longer
Yes	Yes	or	Yes	3 months before symptom onset or first positive finding consistent with TB disease, whichever is longer
No	No	and	No	4 weeks before date of suspected diagnosis
No	Yes	or	Yes	3 months before first positive finding consistent with TB

**SOURCE:** California Department of Health Services Tuberculosis Control Branch; California Tuberculosis Controllers Association. Contact investigation guidelines. Berkeley, CA: California Department of Health Services; 1998.

\* Acid-fast bacilli.

**Figure 3. Screenshot of Table 2 from Guidelines for the investigation of contacts of persons with infectious tuberculosis [4] showing decision rules for how to determine the infectious period start date based on case characteristics.**

**Infectious Period Start** is generally calculated using the value for **Symptom Onset** and/or the date of the first positive finding consistent with TB disease (e.g., abnormal chest x-ray). If a cell in the **Infectious Period Start** column is left blank, LITT will delete the associated row of data during the analysis if **Pediatric**=”N” and **Extrapulmonary Only**=”N.” The format for values is MM/DD/YY. See *Box 1* for additional details about the concept of infectious period.

2. **Infection Acquisition End:** latest date that the case could have been infected. It is calculated using the same method as **Infectious Period Start** by using recommendations from the NTCA and CDC [4] summarized in *Figure 3*; non-infectious cases are nearly always sputum smear negative for AFB and non-cavitory on chest radiograph such that **Infection Acquisition End** is 4 weeks before date of suspected diagnosis. If a cell in the **Infection Acquisition End** column is left blank, LITT will delete the associated row of data if **Pediatric**=”Y” or **Extrapulmonary Only**=”Y.” For infectious cases (i.e., adult cases with pulmonary or laryngeal disease), **Infection Acquisition End** should be left blank because it is identical to **Infectious Period Start**. The format for values is MM/DD/YY. See *Box 1* for additional details about the concept of infection acquisition.

### **Optional variables (yellow columns):**

- **Infection Acquisition Start:** earliest date that the case could have been infected. Calculation of **Infection Acquisition Start** is dependent on case or episode characteristics. For example, for adult cases, the date the case was first present in the geographic area where the cluster occurred could be used. If the cluster occurred in a specific setting (e.g., workplace) then the date the case was first present in that setting (e.g., employment start date in the workplace) could be used. For pediatric cases, the date of birth could be used. If a cell in the **Infection Acquisition Start** column is left blank, the cell will remain blank. While **Infection Acquisition Start** is an optional variable, users are strongly encouraged to populate this date field. The format for values is MM/DD/YY. See [Box 1](#) for additional details about the concept of infection acquisition.
- **Infectious Period End:** end date of the case's infectious period. This variable is only required and can only be calculated for infectious cases (i.e., adult cases with pulmonary or laryngeal disease). To calculate **Infectious Period End**, use recommendations from the NTCA and CDC [4], specifically:

*"The infectious period is closed when the following criteria are satisfied: 1) effective treatment (as demonstrated by susceptibility results) for ≥2 weeks; 2) diminished symptoms; and 3) mycobacteriologic response (e.g., decrease in grade of sputum smear positivity detected on sputum-smear microscopy)... A patient returning to a congregate living setting or to any setting in which susceptible persons might be exposed should have at least three consecutive negative sputum AFB smear results from sputum collected >8 hours apart (with one specimen collected during the early morning) before being considered noninfectious."*

For contact investigation purposes, the infectious period can also end when the case is isolated under airborne infection isolation precautions, even if not all the criteria above have been met. This is because these precautions limit the case's ability to transmit *M. tuberculosis* to additional contacts.

If a cell in the **Infectious Period End** column is left blank, the cell will remain blank. While **Infectious Period End** is an optional variable, users are strongly encouraged to populate this date field. The format for values is MM/DD/YY. See [Box 1](#) for additional details about the concept of infectious period.

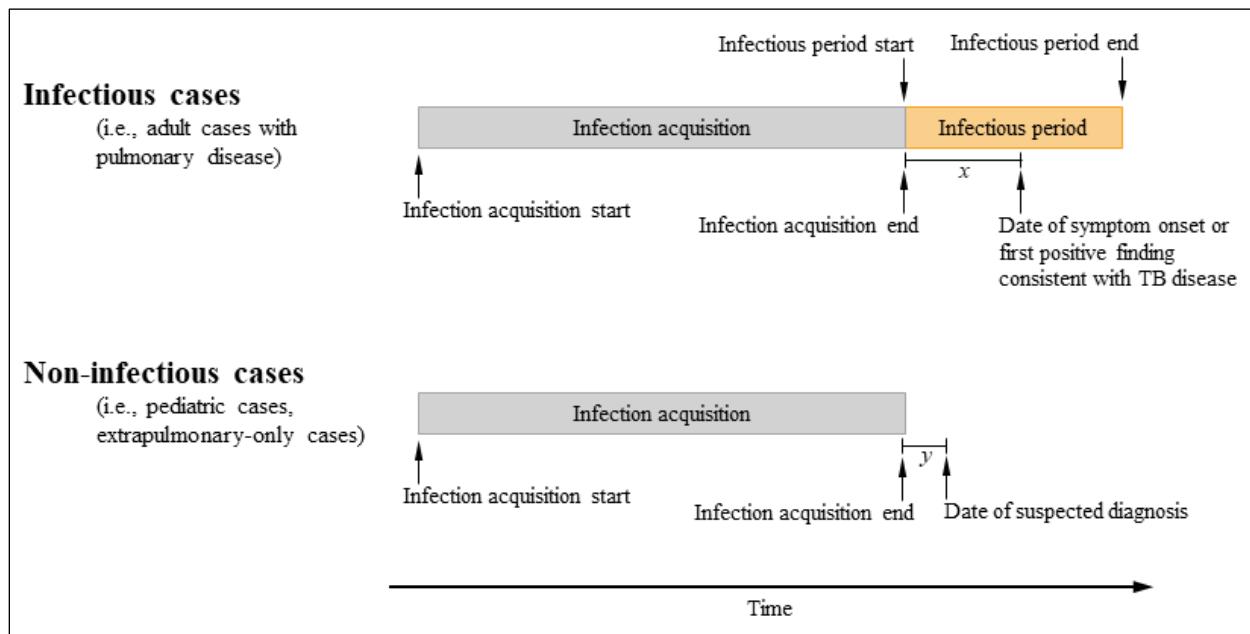
- **Symptom Onset:** earliest date that a case first experienced symptoms consistent with TB disease. Because LITT analyses focus on transmission, **Symptom Onset** is generally synonymous with onset of cough even though TB disease is associated with several other symptoms. If a cell in the **Symptom**

**Onset** column is left blank, the cell will remain blank. While **Symptom Onset** is an optional variable, users are strongly encouraged to populate this date field as values for **Symptom Onset** may be needed to estimate an infectious period. The format for values is MM/DD/YY.

---

#### Box 1: Infection Acquisition Dates and Infectious Period

Values for the variables **Infection Acquisition Start** and **Infection Acquisition End** are the start and end dates of infection acquisition and represent the earliest and latest dates that a case could have been infected by *M. tuberculosis*, respectively (*Figure 4*). Similarly, values for the variables **Infectious Period Start** and **Infectious Period End** are the start and end dates of the calculated infectious period and represent the earliest and latest dates that a case could have been a source of *M. tuberculosis* infection for another case, respectively. For infectious cases, **Infectious Period Start** is identical to **Infection Acquisition End**.



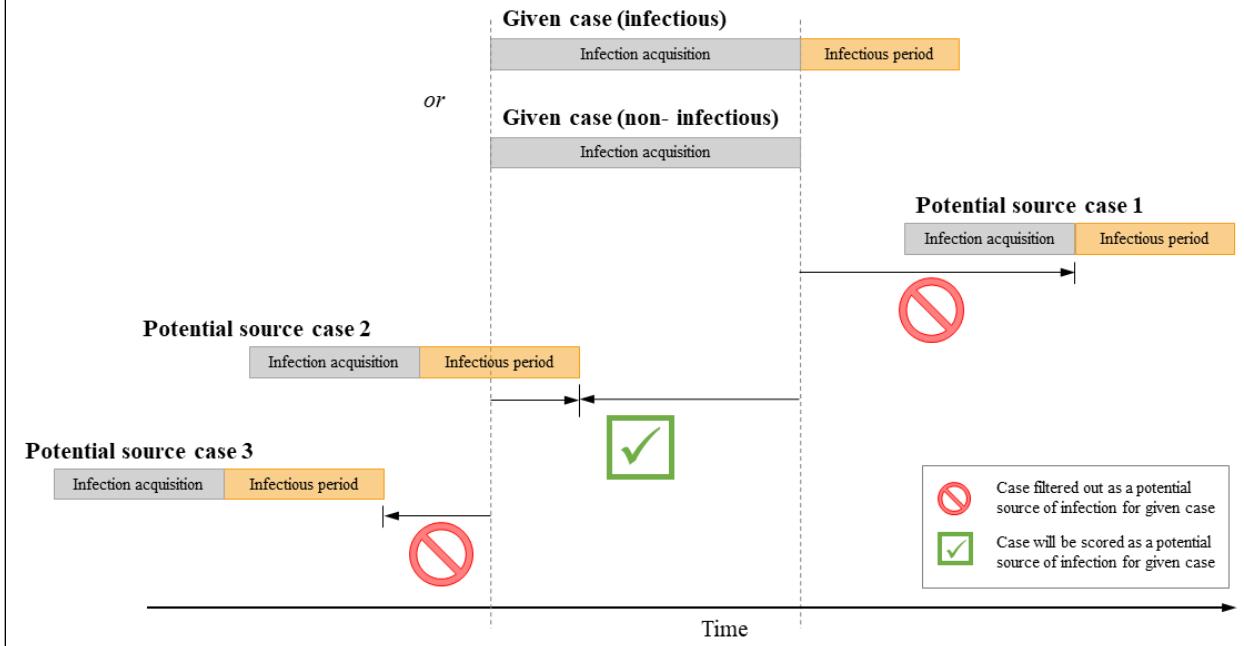
**Figure 4.** Schematic showing the start and end dates for infection acquisition and infectious periods of an infectious case, and for infection acquisition of a non-infectious case.

Only infectious cases (i.e., adult cases with pulmonary or laryngeal disease) have an infectious period. **Infectious Period Start** is calculated using the value for **Symptom Onset** and/or the date of first positive finding consistent with TB disease (e.g., abnormal chest x-ray). Based on this information, recommendations from the NTCA and CDC [4] (*Figure 3*), can be used to determine the value for  $x$  in *Figure 4* and derive the corresponding value for **Infectious Period Start**. For example, if an adult case with pulmonary disease reported cough onset on 09/10/2018 (i.e., case had TB symptoms) and was cavitary on chest radiograph and/or sputum smear positive for AFB, then  $x$  would be 3 months and accordingly, the values for **Infectious Period Start** and **Infection Acquisition End** would 06/10/2018. Note that for infectious cases, **Infection Acquisition End** should be left blank in the **Case data table** input file because it is identical to **Infectious Period Start**.

Although non-infectious cases (i.e., pediatric cases, extrapulmonary-only cases) will not have an infectious period, values for **Infection Acquisition Start** and **Infection Acquisition End** can be calculated. Most non-infectious cases will be sputum smear negative for AFB and non-cavitory on chest radiograph. **Infection Acquisition End** is calculated by extrapolating from the same method used to calculate **Infectious Period Start** (*Figure 3*). For example, if a pediatric case diagnosed with TB disease on 09/10/2018 was asymptomatic, sputum smear negative for AFB, and non-cavitory on chest radiograph, then  $y$  would be 4 weeks and the value for **Infection Acquisition End** would be 08/10/2018. The case's date of birth could be used as the value for **Infection Acquisition Start**.

During a LITT analysis, values for **Infectious Period Start**, **Infectious Period End**, **Infection Acquisition Start**, and/or **Infection Acquisition End** are used to filter out potential source cases that could not have been the source of infection for a given case. *Figure 5* shows an example in which values for **Infection Acquisition Start** and **Infection Acquisition End** are known for an infectious or a non-infectious given case. Potential source case 1 could not have transmitted *M. tuberculosis* to either of the given cases if its **Infectious Period Start** was significantly later in time than the **Infection Acquisition End** of the given cases. Similarly, Potential source case 3 could not have transmitted *M. tuberculosis* to either of the given cases if its **Infectious Period End** was earlier in time than the **Infection Acquisition Start** of the given cases.

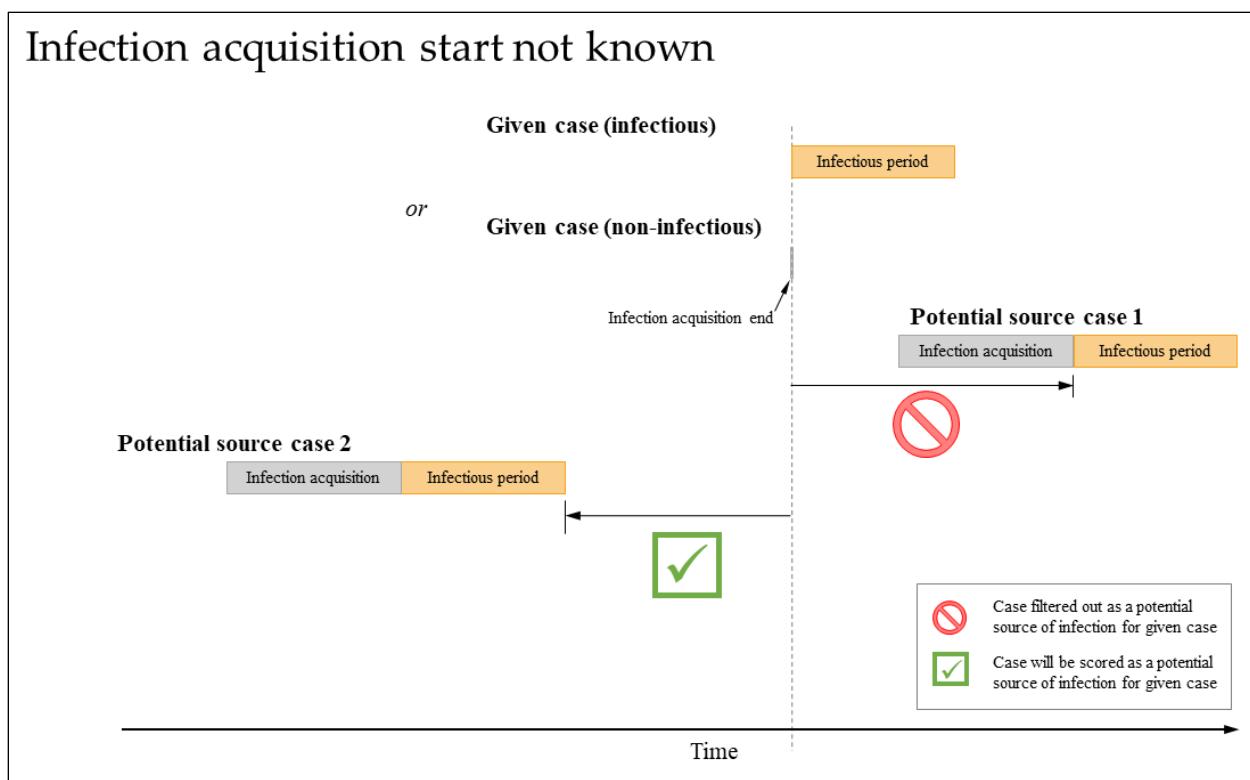
## Infection acquisition start and end are known



**Figure 5.** Schematic showing the start and end dates for infection acquisition and infectious periods of hypothetical infectious and non-infectious given cases and potential source cases and illustrating how LITT uses the temporal sequence of dates to filter out potential source cases when values of infection acquisition start and end are known for the given case.

*Figure 6* shows an example in which a value for **Infection Acquisition Start** is not known for an infectious or a non-infectious given case. For the infectious given case, even though a value for **Infection Acquisition Start** is not known, **Infection Acquisition End** is known if **Infectious Period Start** is known as the two are identical. Potential source case 1 could not have transmitted *M. tuberculosis* to this given case if its **Infectious Period Start** was significantly later in time than the **Infection Acquisition End** of the given case. In contrast, if the **Infectious Period End** for potential source case 2 was earlier in time than the **Infection Acquisition End** of the given case, it could have transmitted *M. tuberculosis* to the given case. The same applies for the non-infectious given case as long as a value for **Infection Acquisition End** is known.

## Infection acquisition start not known



**Figure 6. Schematic showing the start and end dates for infection acquisition and infectious periods of hypothetical infectious and non-infectious given cases and potential source cases and illustrating how LITT uses the temporal sequence of dates to filter out potential source cases when a value of infection acquisition start is not known for the given case.**

- **Investigation Presumed Source:** value (or comma separated list of the values) of **Case ID**(s) indicating the user's presumed source determination(s) (i.e., one or more most likely potential source case(s) for the given case, as identified by the user prior to performing a LITT analysis). Note that the presumed source(s) must be one or more other cases included in the **Case data table** input file, and the value(s) of **Case ID** must match exactly in all input files. The number of values of **Case ID** in each cell must match the number of values entered in the associated cell in the **Investigation Presumed Source Strength** column. **Investigation Presumed Source** is not used in a LITT analysis. However, if included in the **data** worksheet, values entered for **Investigation Presumed Source** will be appended in LITT output files to facilitate comparisons between LITT and user-generated characterization of possible

and probable transmission pathways within the cluster. If a cell in the **Investigation Presumed Source** column is left blank, the cell will remain blank.

- **Investigation Presumed Source Strength:** value (or comma separated list of values) of numbers or text data indicating the confidence the user has in their presumed source determination(s) (i.e., values in the corresponding cell in the **Investigation Presumed Source** column). Note that the order and number of values in each cell must match the order and number of **Case ID**(s) entered in the corresponding cell in the **Investigation Presumed Source** column. **Investigation Presumed Source Strength** is not used in a LITT analysis. However, if included in the **data** worksheet, values entered for **Investigation Presumed Source Strength** will be appended in LITT output files to facilitate comparisons between LITT and user generated characterization of possible and probable transmission pathways within the cluster. If a cell in the **Investigation Presumed Source Strength** column is left blank, the cell will remain blank.
- Risk factor(s): one or more columns containing information on risk factors for transmission or development of TB disease (e.g., homelessness, HIV infection, injection drug use). If one or more risk factors are to be included in a LITT analysis, each risk factor must be added as a separate column, beginning with the first blank column to the right of the last optional variable in the **data** worksheet (**Investigation Presumed Source Strength**). Options are “Y” (case was associated with the risk factor) or “N” (case was not associated with the risk factor). While determining what constitutes association of a case with a risk factor is a subjective process (i.e., there is no single “correct” way to define association), assignments should be made consistently across all cases within a cluster in accordance with some system of formalized, user-defined rules, criteria, or definitions. Column header(s) associated with risk factor(s) must match risk factor name(s) in the **Variable** column in the **data** worksheet of the **Table of risk factor weights** input file exactly; the order in which the risk factors are listed in the two files can differ. If a cell in any risk factor column is left blank, LITT will assign a default value of “N” during the analysis.

**Note:** “Risk factor(s)” does not appear in green text because it is not an explicit variable name. In contrast, users specify the specific names of any risk factors they want to include in a LITT analysis (e.g., homelessness, HIV infection, injection drug use) and these names become the associated variable names.

- **Visualization characteristic(s):** one or more columns containing text for use in labelling visualizations generated from results of a LITT analysis (e.g., case characteristics such as sex, country of origin, race/ethnicity). If one or more visualization characteristics are to be included in a LITT analysis, each characteristic must be added as a separate column, beginning with the first blank column to the right

of the last optional variable in the [data](#) worksheet (either [Investigation Presumed Source Strength](#) or the last Risk factor). Column header(s) associated with visualization characteristics do not need to match content in any other input file; visualization characteristics can be listed in any order. Options will be dependent on pertinent characteristics of cases in the cluster. **Do not include personally identifiable information about cases.** Visualization characteristic(s) are not used in a LITT analysis but rather are appended in one of more LITT output files to facilitate visualization of results. If a cell in a visualization characteristic column is left blank, the cell will remain blank.

**Note:** “Visualization characteristic(s)” does not appear in green text because it is not an explicit variable name. In contrast, users specify the specific names of any visualization characteristics they want to include in a LITT analysis (e.g., case characteristics such as sex, country of origin, race/ethnicity) and these names become the associated variable names.

### C. Epi link table

The [Epi link table](#) input file contains qualitative information about the known epi links between some or all pairs of cases included in the cluster being analyzed. A screenshot of the [Epi link table](#) input file from [LITT training dataset 01](#) provides an example of how a populated [data](#) worksheet might appear (*Figure 7*).

	A	B	C	D
1	Case1	Case2	Epi Link Strength	Label
2	Case 1	Case 3	Definite	Homeless shelter A
3	Case 1	Case 5	Definite	Homeless shelter A
4	Case 2	Case 3	Definite	Household
5	Case 3	Case 5	Definite	Homeless shelter A

**Figure 7.** Screenshot of the [data](#) worksheet in the [LITT training dataset 01](#) [Epi link table](#) input file.

While the [Epi link table](#) input file is not required to perform a LITT analysis, data on epi links can be highly informative and exert strong effects on LITT outputs. See [Box 2](#) for additional details about epi links.

The **data** worksheet in the **Epi link table** input file includes three required variables (green columns) and one optional variable (yellow column) described below.

#### **Required variables (green columns):**

1. **Case1**: unique case identifier (**Case ID**) for one case in an epi linked pair of cases. If a cell in the **Case1** column is left blank, LITT will delete the associated row of data during the analysis. Note that values of **Case ID** must match exactly in all input files. The directionality/order in which the cases of a case pair are listed (i.e., which case of the pair is designated as **Case1**) is not important, and there is no need to list each pair twice so that each case in the pair is designated as **Case1** (e.g., if “Case A” is epi linked to “Case B,” it is assumed that “Case B” is epi linked to “Case A”). In the **SNP distance matrix** input file provided to users by CDC, accession numbers from TB GIMS will generally be used as values of **Case ID**. In order to maintain exactly matching values of **Case ID** across input files, users will need to either use accession numbers as values of **Case ID** in all relevant input files or edit the case identifiers used in the **SNP distance matrix** input file to reflect the identifier of choice. If editing values of **Case ID** in the **SNP distance matrix** input file, **do not include personally identifiable information about cases**.
2. **Case2**: unique case identifier (**Case ID**) for the other case in the epi linked pair of cases. All other features of this variable are the same as those for **Case1**.
3. **Epi Link Strength**: strength of the epi link between the pair of epi linked pair of cases. Options are “Definite” (cases are definitely epi linked), “Probable” (cases are probably epi linked), or “Possible” (cases are possibly epi linked). While assigning values of **Epi Link Strength** to epi links is a subjective process (i.e., there is no single correct way to assign values), assignments should be made consistently across all epi links within a cluster in accordance with some system of formalized, user-defined rules, criteria, or definitions. If a cell in the **Epi Link Strength** column is left blank, LITT will assign a default value of “Probable” during the analysis.

#### **Optional variable (yellow column):**

- **Label**: text describing the relationship between the pair of epi linked cases. For example, “Hotel A,” “Homeless Shelter B,” “Siblings,” “Coworkers at facility C.” **Do not include personally identifiable information about cases**. **Label** data are not used in a LITT analysis, but rather are appended in one of more LITT output files to facilitate visualization of results. If a cell in the **Label** column is left blank, the cell will remain blank.

## **Box 2: Some guidance on epi links and risk factors**

### *How to know which epi links to include in a LITT analysis?*

For the purpose of thorough and complete record keeping, it is recommended that all known epi links be included in the [Epi link table](#) input file. Multiple links can be entered for a given pair of cases in a cluster, but LITT will only consider the link with the strongest (i.e., most definitive) value of [Epi Link Strength](#) in the analysis.

### *How to know what strength to assign an epi link?*

When assigning a strength value to an epi link, be mindful of the distinction between the strength of the connection between the two persons (i.e., your sense of whether the epi link is real) and the likely contribution of the link to transmission. For example, within a hypothetical cluster there may be clear evidence of two epi links: Case 1 and Case 2 lived together in a small bunkhouse at a summer camp, and Case 1 and Case 3 were friends on social media. Arguably, since there is clear evidence for the existence of both of these epi links, they could both be assigned an [Epi Link Strength](#) value of "Definite." In this case, LITT will assign the same quantitative weight to both links when evaluating possible transmission pathways. It could be argued, however, that cohabitation in a confined space has a relatively high likelihood of resulting in transmission compared to a connection on social media, which might not be associated with any actual physical contact. In light of this, a user might decide to assign an [Epi Link Strength](#) value of "Probable" or "Possible" to the second link to reduce its importance in the LITT analysis, even though there is clear evidence that it is a real link.

While assigning values of [Epi Link Strength](#) to epi links is a subjective process (i.e., there is no single correct way to assign values), assignments should be made consistently across all epi links within a cluster in accordance with some system of formalized, user-defined rules, criteria, or definitions. As an example, the following system of [Epi Link Strength](#) definitions was formalized by staff of multiple state and local TB control programs and CDC at a meeting of the Outbreak Detection Working Group on April 27, 2012:

1. "Definite" epi link:
  - a. Two cases where at least one named the other as a contact.

*or*

- b. Two cases shared airspace at the same location (e.g., workplace, bar, shelter) at the same time.
- 2. “Probable” epi link:

- a. Two cases shared a common named contact but did not name each other as contacts.

*or*

- b. Two cases shared airspace at the same location (e.g., workplace, bar, shelter) during the same general time period, but the health department was unable to document that they were there at the same time.

*or*

- c. A person-search database (e.g., Accurint) or Internet search indicates the patients are relatives, associates, co-workers, or have shared a home address.

- 3. “Possible” epi link:

- a. Two cases lived or worked in the same neighborhood during the same general time period.

*or*

- b. Two cases shared activities or social/behavioral traits that increase the chances that they were in contact with each other (e.g., were experiencing homelessness at the same time and may have spent time in common locations for shelter or food).

*or*

- c. The two patients are connected on a social network site (e.g., are Facebook friends, are contacts on LinkedIn, one or both patients follow the other on Twitter).

*How to know which risk factors to include in a LITT analysis? (see below for more information about risk factors)*

For the purpose of thorough and complete record keeping it is recommended that all known risk factors be documented, but only those suspected of meaningfully contributing to transmission within the cluster

being analyzed should be included in the [Table of risk factor weights](#) input file. While a seemingly more comprehensive approach, including risk factors known to affect TB transmission in general but not suspected of meaningfully contributing to transmission within the cluster being analyzed will act to dilute the effect of the contributing factors in the LITT analysis. LITT will only consider risk factors shared by a given pair of cases in a cluster if that pair of cases does not share any epi links.

*How to know what weights to assign to risk factors? (see [below](#) for more information about risk factors)*

As with values of [Epi Link Strength](#), assigning weights to risk factors is a subjective process (i.e., there is no single correct way to assign weights). Weights are relative and should reflect the user's understanding of the importance of various risk factors to TB transmission and disease development in general as well as the importance of those risk factors within the cluster being analyzed given available information about cases and routes of transmission.

*How to know whether to code a connection between cases as an epi link or a risk factor? (see [below](#) for more information about risk factors)*

While it is technically possible to code risk factors as epi links, it is recommended that the former be reserved for attributes, characteristics, or exposures shared by two or more cases within a cluster while the latter be reserved for instances of known or suspected contact between cases.

---

## **D. SNP distance matrix**

The [SNP distance matrix](#) input file contains quantitative information about the genetic distance between *M. tuberculosis* isolates (as determined by WGS analyses) collected from some or all cases included in the cluster being analyzed. A screenshot of the [SNP distance matrix](#) input file from [LITT training dataset 01](#) provides an example of how a populated [data](#) worksheet might appear ([Figure 8](#)).

	A	B	C	D	E	F	G	H	I	J
1		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	MRCA
2	Case 1	0	0	1	20	1	21	25	26	10
3	Case 2	0	0	1	20	1	21	25	26	10
4	Case 3	1	1	0	21	2	22	26	27	11
5	Case 4	20	20	21	0	21	1	5	6	10
6	Case 5	1	1	2	21	0	22	26	27	11
7	Case 6	21	21	22	1	22	0	6	7	11
8	Case 7	25	25	26	5	26	6	0	1	15
9	Case 8	26	26	27	6	27	7	1	0	16
10	MRCA	10	10	11	10	11	11	15	16	0

**Figure 8. Screenshot of the [data](#) worksheet in the LITT training dataset 01 SNP distance matrix input file.**

While the [SNP distance matrix](#) input file is not required to perform a LITT analysis, data on SNP distances can be highly informative and exert strong effects on LITT outputs. For example, SNP distance data can often be used to rule out recent transmission between cases that are epi linked based on other sources of information (see [above](#)).

Two aspects of the format of the [SNP distance matrix](#) input file are noteworthy. First, the cells across the diagonal of the matrix all contain the value of 0 since the SNP distance between any *M. tuberculosis* isolate and itself is 0). Cells in the triangular blocks above and below this diagonal are symmetrical since, for example, the SNP distance between the *M. tuberculosis* isolates from Case 3 and Case 7 is the same as the SNP distance between the *M. tuberculosis* isolates from Case 7 and Case 3) ([Figure 8](#)). LITT can process [SNP distance matrix](#) input files that are fully populated (i.e., all cells in the diagonal of the matrix and in the upper and lower triangular blocks contain SNP distance values) or partially populated (i.e., only cells in the upper or the lower triangular blocks contain SNP distances values, with cells in the unpopulated triangular block blanks; cells across the diagonal may also be blank) ([Figure 9](#)).

	A	B	C	D	E	F	G	H	I
1		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
2	Case 1								
3	Case 2	0							
4	Case 3	1	1						
5	Case 4	20	20	21					
6	Case 5	1	1	2	21				
7	Case 6	21	21	22	1	22			
8	Case 7	25	25	26	5	26	6		
9	Case 8	26	26	27	6	27	7	1	
10	MRCA	10	10	11	10	11	11	15	16

**Figure 9.** Screenshot of the [data](#) worksheet in a [SNP distance matrix](#) input file showing a partially populated matrix in which cells in the upper triangular block and along the diagonal of the matrix are blank.

Second, notice that in addition to the pairs of cases in the cluster for which SNP distance data are available, the distance matrix in the [SNP distance matrix](#) input file will also contain a distance between each included case and the most recent common ancestor, or MRCA. The MRCA is not used in a LITT analysis.

CDC will provide [SNP distance matrix](#) input files upon request to LITT users in state and local TB control programs in the United States. To request a SNP distance matrix for a LITT analysis, users should send an email to their point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov). In the body of the email, specify the accession number for every sample that should be included in the SNP distance matrix. Note that CDC technical assistance may not always be available for all user analyses and that the availability of these support services may be discontinued at any time.

In the [SNP distance matrix](#) input file, CDC will generally use accession numbers from TB GIMS as values of [Case ID](#). [Case ID](#) is used for both the row labels and column headers. In order to maintain exactly matching values of [Case ID](#) across input files, users will need to either use accession numbers as

values of Case ID in all relevant input files or edit the case identifiers used in the SNP distance matrix input file to reflect the identifier of choice. If editing values of Case ID in the SNP distance matrix input file, **do not include personally identifiable information about cases**.

### E. Table of risk factor weights

The Table of risk factor weights input file contains quantitative information about risk factors associated with some or all cases included in the cluster being analyzed. A screenshot of the Table of risk factor weights input file from LITT training dataset 01 provides an example of how a populated data worksheet might appear (*Figure 10*).

	A	B
1	Variable	Weight
2	Homelessness	2.00
3	HIV_pos	6.00
4	Inj_drug_use	4.00

**Figure 10.** Screenshot of the data worksheet in the LITT training dataset 01 Table of risk factor weights input file.

While the Table of risk factor weights input file is not required to perform a LITT analysis, data on risk factors can inform LITT outputs. See *Box 2* for additional details about risk factors. The data worksheet in the Table of risk factor weights input file includes two required variables (green columns) described below.

#### Required variables (green columns):

1. **Variable:** name of the risk factor. Risk factor names in the **Variable** column must exactly match column headers associated with risk factors in the **data** worksheet in the **Case data table** input file. Note that multiple factors can be combined for the purpose of simplifying the analysis (e.g., alcohol use, non-intravenous drug use, and/or intravenous drug use could be combined into a single risk factor called any substance use).
2. **Weight:** numerical value assigned to the risk factor by the user that measures the perceived importance of that factor in terms of transmission relative to other factors. The larger the numerical value, the higher the perceived relative importance of the risk factor. In the example from **LITT training dataset 01** illustrated in *Figure 10*, HIV positivity (**Weight**=6.00) is being considered three times as important as homelessness (**Weight**=2.00), and injection drug use (**Weight**=4.00) is being considered twice as important as homelessness (**Weight**=2.00). Results of this LITT analysis would be identical if the weights for these three risk factors were instead set at [20, 60, 40], [200.0, 600.0, 400.0], [1.00, 3.00, 2.00], [5, 15, 10], [3.0, 9.0, 6.0], [2.5, 7.5, 5], [1.25, 3.75, 2.50], etc. **Weight** values should be positive numbers with up to two decimal places (e.g., **Weight**="1.25"). If a cell in the **Weight** column is left blank, LITT will delete the associated row of data during the analysis *except* if every cell in the **Weight** column is left blank, in which case LITT will assign an equal value of **Weight** to all risk factors. Risk factors with a **Weight**>"0" will be included in a LITT analysis; risk factors with a **Weight**="0" will not be included in a LITT analysis but will be included in LITT output tables for use in visualizations generated from results of a LITT analysis.

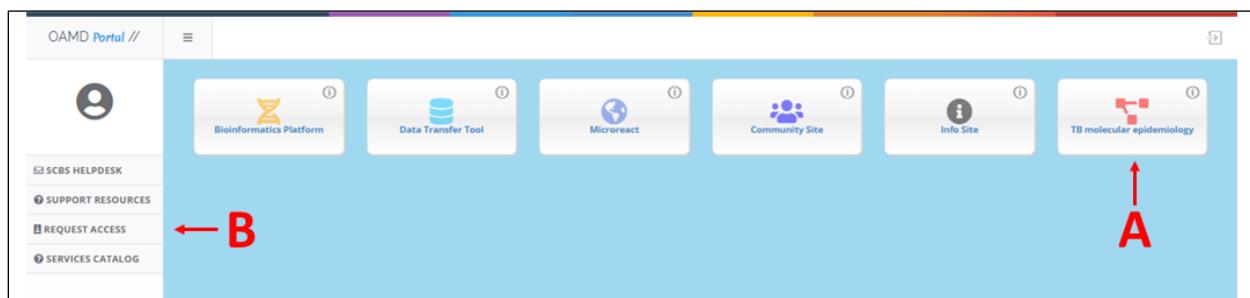
## Performing an analysis

### A. Accessing the algorithm

LITT is available as an interactive webpage hosted by the CDC's Office of Advanced Molecular Detection (OAMD). Users access LITT via a web browser by logging into the OAMD web portal (*Figure 11*), an online clearinghouse for select digital products and services curated by OAMD. For additional information about how to access the OAMD web portal, see Appendix 1 *below*.

Once logged into the OAMD web portal (*Figure 11*), users will see their approved list of accessible OAMD applications; each application will appear as a visual tile. LITT is accessed via the **TB molecular epidemiology** tile (A in *Figure 11*); click on this tile to navigate to the **Algorithms for TB molecular epidemiology analysis** webpage (*Figure 12*). If this tile is not present, click on the **Request access** button

on the left of the screen (B in *Figure 11*), then click on the “TB molecular epidemiology” option in the left side of the pop-up window to move it to the right side (“Requested”) of the pop-up window, and then click on the **Go** button at the top right of the pop-up window. If the tile is still not present or working properly, users should send an email to their Division of Tuberculosis Elimination (DTBE) point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGGenotyping@cdc.gov](mailto:TBGGenotyping@cdc.gov) to request assistance. Note that CDC technical assistance may not always be available for all user analyses and that the availability of these support services may be discontinued at any time.



**Figure 11.** Annotated screenshot of the OAMD web portal. Annotation appears in red.

On the [Algorithms for TB molecular epidemiology analysis](#) webpage, users will see links to TB-related data analysis algorithms and associated files and documentation. LITT is accessed via the [LITT](#) hyperlink (A in *Figure 12*), which will direct users to the LITT online user interface (*Figure 13*).

 An annotated screenshot of the 'Algorithms for TB molecular epidemiology analysis' webpage. The page title is 'Algorithms for TB molecular epidemiology analysis'. Below the title is a section titled 'Logically Inferred Tuberculosis Transmission (LITT)'. A red arrow labeled 'A' points to the 'LITT' link. Below this is another section titled 'Location And Time To Epi (LATTE)'. A red arrow labeled 'A' also points to the 'LATTE' link.
 

**Algorithms for TB molecular epidemiology analysis**

**Logically Inferred Tuberculosis Transmission (LITT)**

LITT is an algorithm that can be used to integrate whole genome sequencing, clinical, and epidemiological data to identify and rank potential sources for tuberculosis cases.

[LITT](#) ← A

**Location And Time To Epi (LATTE)**

LATTE identifies all overlaps in space and time from a list of dates that cases (or contacts) were in a particular location, then combines that with infectious periods (IP) to identify and categorize epi or IP epi links.

[LATTE](#)

**Figure 12.** Annotated screenshot of the [Algorithms for TB molecular epidemiology analysis](#) webpage. Annotation appears in red.

Logically Inferred Tuberculosis Transmission (LITT) ?

**Input files**

Warning: do not upload personally identifiable information (PII)

**Case data table**

Output extra columns in case data table

**Epi link table**

**SNP distance matrix**

Include distance matrix in outputs

**Set up outputs**

Name prefix for output files

**SNP cutoff**  
 Maximum number of SNPs for two cases to be linked by transmission.  


**Advanced options**

**Table of risk factor weights**

This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table.

**Figure 13.** Screenshot of the LITT online user interface.

### B. Uploading input files

To upload a **Case data table** input file, click the **Browse** button under “Case data table” on the left of the screen (A in *Figure 14*) and select the **Case data table** input file from the finder window (i.e., all LITT input files can be stored on the user’s local computer in any drive or folder that can be accessed via the browser’s finder). This file is required to perform a LITT analysis.

Logically Inferred Tuberculosis Transmission (LITT) I → ?

**Input files**

Warning: do not upload personally identifiable information (PII)

**Case data table** A →  No file selected

**Epi link table** B →  Output extra columns in case data table

**SNP distance matrix** D →  No file selected

**Include distance matrix in outputs** E →

**Set up outputs**

Name prefix for output files F →

**SNP cutoff**  
Maximum number of SNPs for two cases to be linked by transmission.

0 G → 5 H 7

**Advanced options**

Table of risk factor weights I → H

No file selected

This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table.

Run

**Figure 14.** Annotated screenshot of the LITT online user interface. Annotation appears in red.

To upload an [Epi link table](#) input file, click the [Browse](#) button under “Epi link table” on the left of the screen (C in [Figure 14](#)) and select the [Epi link table](#) input file from the finder window. To upload a [SNP distance matrix](#) input file, click the [Browse](#) button under “SNP distance matrix” on the left of the screen (D in [Figure 14](#)) and select the [SNP distance matrix](#) input file from the finder window. Although the [Epi link table](#) and the [SNP distance matrix](#) input files are both optional for performing a LITT analysis, at least one of the two is required; inclusion of both is recommended to improve the accuracy and completeness of LITT outputs. To upload a [Table of risk factor weights](#) input file, click the [Browse](#) button under “Table of risk factor weights” on the right of the screen (H in [Figure 14](#)) and select the [Table of risk factor weights](#) input file from the finder window. This file is optional for performing a LITT analysis, but inclusion should improve the accuracy and completeness of LITT outputs when epi link data are missing.

Clicking on the [Help](#) (“?”) button ([Figure 14](#)) will reveal a pop up window (A in [Figure 15](#)) that provides links to supporting materials on the [TB molecular epidemiology GitHub](#) webpage including the LITT user’s manual and training presentation, input file templates (see [above](#)), training datasets (see [below](#)), the LITT R source code, and a reference (peer-reviewed publication) describing LITT [2]. The [TB molecular epidemiology GitHub](#) can also be accessed using this [link](#).

# Logically Inferred Tuberculosis Transmission (LITT)

A → ?

### Input files

**Warning:** do not upload personally identifiable information (PII)

**Case data table**

Output extra columns in case data table

**Epi link table**

**SNP distance matrix**

Include distance matrix in outputs

### Set up outputs

**Name prefix for output files**

**SNP cutoff**  
Maximum number of SNPs for two cases to be linked by transmission.

This table contains a column of risk factors, with their weights. variable names must exactly match the name of the column in the case data table.

**Table of risk factors**

Help  
User's manual & training presentation  
Input file templates  
Training datasets  
Reference

**Figure 15.** Annotated screenshot of the LITT online user interface showing the pop up window revealed by clicking the **Help** button. Annotation appears in red.

Once an input file has been successfully uploaded, the file name will appear in the box to the right of the associated **Browse** button and a hatched blue bar reading “Upload complete” will appear underneath the box (*Figure 16*).

Logically Inferred Tuberculosis Transmission (LITT)

?

**Input files**

Warning: do not upload personally identifiable information (PII)

**Case data table**

Browse... LITT\_training\_dataset\_01\_Cas  
Upload complete

Output extra columns in case data table

**Epi link table**

Browse... LITT\_training\_dataset\_01\_Epi  
Upload complete

**SNP distance matrix**

Browse... LITT\_training\_dataset\_01\_SNI  
Upload complete

Include distance matrix in outputs

**Set up outputs**

Name prefix for output files

**SNP cutoff**  
Maximum number of SNPs for two cases to be linked by transmission.  


**Advanced options**

**Table of risk factor weights**

Browse... LITT\_training\_dataset\_01\_Tab  
Upload complete

This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table.

**Figure 16.** Screenshot of the LITT online user interface showing successfully uploaded [Case data table](#), [Epi link table](#), [SNP distance matrix](#), and [Table of risk factor weights](#) input files.

### C. Setting analysis parameters

After uploading input files, review and select parameter settings for the LITT analysis.

1. **Output extra columns in case data table:** If this box is checked (B in [Figure 14](#)), the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) output file (see [below](#)) will contain all columns in the [Case data table](#) input file (i.e., required and optional variables). If the box is not checked, the [Calculated\\_Case\\_Data](#) output file will contain only the columns of data in the [Case data table](#) input file that are used in the LITT analysis (i.e., required variables). The default setting is for the box to be checked and the extra columns output to the output file.
2. **Include distance matrix in outputs:** If this box is checked (E in [Figure 14](#)), the [\(Prefix\)LITT\\_Distance\\_Matrix](#) output file (see [below](#)) will be generated. This file returns the matrix of SNP distances uploaded for the LITT analysis in the [SNP distance matrix](#) input file. The default setting is for the box not to be checked and the extra output file not to be generated.

3. **Name prefix for output files:** The text entered here (F in *Figure 14*) will form the prefix for the name of the zipped folder that is generated by the LITT analysis and all LITT output files contained within it. Additionally, LITT will append the text “LITT” to the name prefix when naming the folder and files. For instance, if the text “Example” is entered, the zipped folder generated by the associated LITT analysis will be named “ExampleLITT” and the names of all of the output files contained within it will begin with “ExampleLITT.”
4. **SNP cutoff:** the value selected using this slider (G in *Figure 14*) sets the SNP distance threshold at which the LITT analysis will consider it possible for two cases in the cluster to be linked by transmission. The default value for the **SNP cutoff** is 5 SNPs but can be adjusted to vary between zero and 7 SNPs. In practice, a **SNP cutoff** value of 5 means that if *M. tuberculosis* isolates from two cases within a cluster differ by 6 or more SNPs, LITT will not consider either case as a possible source of infection for the other. In contrast, if the two cases differ by 5 or fewer SNPs, LITT may consider each case as a possible source of infection for the other.

**Note:** There is no single, correct, empirically-based SNP distance threshold that can be used consistently to characterize transmission within TB clusters. However, thresholds between zero and 10 SNPs have been shown to be biologically appropriate in multiple investigations of TB clusters [5], and a value of 5 SNPs has been proposed as a reasonable reference threshold [6].

#### **D. Initiating an analysis**

Once input files have been successfully uploaded and analysis parameters have been set, hit the **Run** button (A in *Figure 17*) to initiate the LITT analysis. If the analysis is proceeding successfully, a progress bar will appear in the bottom right corner of the screen above text that reads “Running LITT” (B in *Figure 17*).

Logically Inferred Tuberculosis Transmission (LITT) ?

**Input files**

Warning: do not upload personally identifiable information (PII)

**Case data table**

Browse... LITT\_training\_dataset\_01\_Cas Upload complete

Output extra columns in case data table

**Epi link table**

Browse... LITT\_training\_dataset\_01\_Epi Upload complete

**SNP distance matrix**

Browse... LITT\_training\_dataset\_01\_SNI Upload complete

Include distance matrix in outputs

**Set up outputs**

Name prefix for output files

**SNP cutoff**  
Maximum number of SNPs for two cases to be linked by transmission.

0 1 2 3 4 5 6 7

**Advanced options**

Table of risk factor weights

Browse... LITT\_training\_dataset\_01\_Tab Upload complete

This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table.

A → Run

B → Running LITT ×

**Figure 17. Annotated screenshot of the LITT online user interface showing a LITT analysis proceeding successfully. Annotation appears in red.**

After completion of the analysis, text that reads “Analysis complete” and a [Download Results](#) button will appear beneath the [Run](#) button (A in *Figure 18*).

**Logically Inferred Tuberculosis Transmission (LITT)**

**Figure 18.** Annotated screenshot of the LITT online user interface after successful completion of a LITT analysis. Annotation appears in red.

After clicking on the [Download Results](#) button, an icon of a zipped download file will appear in the bottom left hand corner of the screen (A in *Figure 19*). Click on this icon to choose the location where the zipped folder containing the results of the LITT analysis should be saved. Note that after clicking on the [Download Results](#) button, the specific steps to follow to save results of the analysis, and the appearance of the screen, may vary slightly depending on the browser and browser settings being used.

Logically Inferred Tuberculosis Transmission (LITT)

Input files

Warning: do not upload personally identifiable information (PII)

**Case data table**

Browse... LITT\_training\_dataset\_01\_Cas  
Upload complete

Output extra columns in case data table

**Epi link table**

Browse... LITT\_training\_dataset\_01\_Epi  
Upload complete

**SNP distance matrix**

Browse... LITT\_training\_dataset\_01\_SNP  
Upload complete

Include distance matrix in outputs

Set up outputs

Name prefix for output files  
Example

SNP cutoff  
Maximum number of SNPs for two cases to be linked by transmission.  
0 1 2 3 4 5 6 7

Advanced options

Table of risk factor weights

Browse... LITT\_training\_dataset\_01\_Tab  
Upload complete

This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table.

Clear inputs

Run

Analysis complete

Download Results

ExampleLITT.zip | -A | Show all | X

**Figure 19.** Annotated screenshot of the LITT online user interface after successful completion of a LITT analysis and clicking on the [Download Results](#) button. Annotation appears in red.

If the LITT analysis cannot be successfully completed for some reason, an error message may appear. For example, the message “No case data. Please input a case data table” will appear underneath the **Run** button (A in *Figure 20*) if the **Case data table** input file was not successfully uploaded or contained one or more formatting errors in the **data** worksheet.

The screenshot shows the LITT online user interface with three main sections: Input files, Set up outputs, and Advanced options.

- Input files:**
  - Case data table:** A browse button and a "No file selected" button.
  - Epi link table:** A browse button, a file name input field containing "LITT\_training\_dataset\_01\_Epi", and a "Upload complete" button.
  - SNP distance matrix:** A browse button, a file name input field containing "LITT\_training\_dataset\_01\_SNP", and a "Upload complete" button.
  - Output extra columns in case data table:** A checked checkbox.
  - Include distance matrix in outputs:** An unchecked checkbox.
- Set up outputs:**
  - Name prefix for output files:** An input field containing "Example".
  - SNP cutoff:** A slider scale from 0 to 7 with a value of 5 highlighted.
- Advanced options:**
  - Table of risk factor weights:** A table header with "Browse..." and "LITT\_training\_dataset\_01\_Tab" buttons, and a "Upload complete" button.
  - A descriptive text: "This table contains a list of the columns in the case data table to use as risk factors, with their weights. Variable names must exactly match the name of the column in the case data table."

**Annotations:**

- A →** Red arrow pointing to the error message: "No case data. Please input a case data table."
- B →** Red arrow pointing to the "Clear inputs" button.

**Figure 20. Annotated screenshot of the LITT online user interface after unsuccessful completion of a LITT analysis resulting in generation of an error message. Annotation appears in red.**

In the event of unsuccessful completion of the LITT analysis, inspect all input files for proper content and formatting and attempt to repeat the analysis. When ready to repeat the analysis, click on the [Clear inputs](#) button (B in *Figure 20*) to clear any previously uploaded input files from the webpage and then re-upload modified input files.

## Using analytic outputs

### A. Interpreting analytic outputs

Results of a successful LITT analysis are downloaded as a single zipped folder. Opening the folder will unzip it and extract the following seven or eight individual LITT output files. All but one of these files is in Microsoft Excel format. The file names listed below include “(Prefix)” which represents the [Name prefix for output files](#) (F in *Figure 14*) entered by the user.

1. **(Prefix)LITT\_log**: this text file provides a high-level summary of the LITT analysis (*Figure 21*). It reports the **SNP cutoff** value selected by the user, the number of cases in the **Case data table** input file, the number of epi links in the **Epi link table** input file, and the number of risk factors in the **Table of risk factor weights** input file. If any issues with data quality or completeness were encountered during the LITT analysis, associated changes made by the algorithm are detailed (e.g., missing values that were set to default values). If the analysis stopped unexpectedly due to an error, information on the associated issue(s) may be detailed.

```
LITT analysis
SNP cutoff = 5
Number of cases: 8
Number of epi links: 4
Number of risk factors: 3
```

**Figure 21.** Screenshot of the **(Prefix)LITT\_log** output file.

2. **(Prefix)LITT\_All\_Potential\_Sources**: this file contains two worksheets, the first of which is named **potential sources** (*Figure 22*). The second worksheet in the output file, **filtered cases**, is described *below*. During the analysis, LITT iteratively considers each case in the cluster as the given case and sequentially creates pairwise combinations of this given case and every other case in the cluster that has not been filtered (potential source cases). For each case pair, LITT assesses the likelihood that the potential source case was the source of infection for the given case. The **potential sources** worksheet contains information about all case pairs for which LITT determined that the potential source case could have been the source of infection for the given case.

Given Case	Potential Source	Rank	Score	Without SNP Score	Score Category	SNP Rating	Infectious Rating	Time Rating	Epi and Risk Factor Rating	Investigation Presumed Source	Label
Case 1	Case 3	1	1	0	high likelihood	1	0 (cavitory)	0 (source earlier than given case and within two years)	0 (definite epi link)	Definite presumed source	Homeless shelter A
Case 2	Case 3	1	1	0	high likelihood	1	0 (cavitory)	0 (source earlier than given case and within two years)	0 (definite epi link)	Household	
Case 3	Case 5	2	6	5	low likelihood	1	0 (cavitory)	2 (source IP start later than given case IP start)	3 (no epi link or shared risk factors)		
Case 3	no potential sources										
Case 4	no potential sources										
Case 5	Case 3	1	2	0	medium likelihood	2	0 (cavitory)	0 (source earlier than given case and within two years)	0 (definite epi link)		Homeless shelter A
Case 6	Case 4	1	3.5	2.5	low likelihood	1	0 (cavitory)	0 (source earlier than given case and within two years)	2.5 (shared risk factors)		RF: Homelessness   RF: Inj_drug_use
Case 7	Case 4	1	7	2	low likelihood	5	0 (cavitory)	0 (source earlier than given case and within two years)	2 (possible epi link or all risk factors shared)		RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use
Case 8	Case 7	1	3	2	medium likelihood	1	0 (cavitory)	0 (source earlier than given case and within two years)	2 (possible epi link or all risk factors shared)	Probable presumed source	RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use

Given Case	Potential Source	Rank	Score	Without SNP Score	Score Category	SNP Rating	Infectious Rating
Case 1	Case 3	1	1	0	high likelihood	1	0 (cavitory)
Case 2	Case 3	1	1	0	high likelihood	1	0 (cavitory)
Case 2	Case 5	2	6	5	low likelihood	1	0 (cavitory)
Case 3	no potential sources						
Case 4	no potential sources						
Case 5	Case 3	1	2	0	medium likelihood	2	0 (cavitory)
Case 6	Case 4	1	3.5	2.5	low likelihood	1	0 (cavitory)
Case 7	Case 4	1	7	2	low likelihood	5	0 (cavitory)
Case 8	Case 7	1	3	2	medium likelihood	1	0 (cavitory)

Time Rating	Epi and Risk Factor Rating	Investigation Presumed Source	Label
0 (source earlier than given case and within two years)	0 (definite epi link)		Homeless shelter A
0 (source earlier than given case and within two years)	0 (definite epi link)	Definite presumed source	Household
2 (source IP start later than given case IP start)	3 (no epi link or shared risk factors)		
0 (source earlier than given case and within two years)	0 (definite epi link)		Homeless shelter A
0 (source earlier than given case and within two years)	2.5 (shared risk factors)		RF: Homelessness   RF: Inj_drug_use
0 (source earlier than given case and within two years)	2 (possible epi link or all risk factors shared)		RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use
0 (source earlier than given case and within two years)	2 (possible epi link or all risk factors shared)	Probable presumed source	RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use

**Figure 22.** Screenshot of the **potential sources** worksheet within the **(Prefix)LITT\_All\_Potential\_Sources** output file (top panel) and below it, magnified images of left-hand portion (middle panel) and right-hand portion (bottom panel) of the **potential sources** worksheet.

Data for the following variables are included in the **potential sources** worksheet:

- **Given Case**: the value of **Case ID** for the given case in the case pair. LITT evaluated the likelihood that the specified given case was infected by the associated potential source case.
- **Potential Source**: the value of **Case ID** for the potential source case in the case pair. LITT evaluated the likelihood that the specified given case was infected by the associated potential source case. If LITT filtered out all cases as potential sources of infection for the given case, the text “no potential sources” will appear.
- **Rank**: value corresponding to the rank order of the specified potential source case on the list of all potential source cases that could have been the source of infection for the specified given case. The lower the value of **Rank**, the higher the likelihood that the potential source case was the source of infection for the associated given case. LITT assigns a value of **Rank=“1”** to the potential source case with the lowest value of **Score** (see *Figure 23*) for the given case (i.e., highest likelihood that the

potential source case was the source of infection for the associated given case), a value of **Rank**=”2” to the potential source case with the second lowest value of **Score**, and so on.

**Note:** If two potential source cases have an identical value of **Score**, their value of **Rank** will be tied; LITT will assign them the same value of **Rank** and will skip one unit of **Rank** when assigning a value of **Rank** to the potential source case with the next highest likelihood. For example, if two potential source cases were tied with a value of **Rank**=”2,” the assigned values of **Rank** would be “1” (potential source case with the highest likelihood of being the source of infection for the associated given case), then “2” (for the two potential source cases with identical value of **Score** and the second highest likelihoods of being the source of infection), then “4” (for the potential source case with the third highest likelihood of being the source of infection because **Rank**=”3” is skipped) (*Figure 23*).

Given Case	Potential Source	Rank	Score
Case A	Case B	1	3
Case A	Case C	2	4
Case A	Case D	2	4
Case A	Case E	4	5

**Figure 23.** Screenshot of the first four columns of the **potential sources** worksheet within a hypothetical **(Prefix)LITT\_All\_Potential\_Sources** output file showing two potential source cases with the same values of **Score** (“4”) and **Rank** (“2”), resulting in the next value of **Rank** (“3”) being skipped.

Values of **Rank** associated with an asterisk (“\*”) indicate that a SNP distance was not available for the case pair and that the values of **Score** and **Rank** for this case pair were calculated differently relative to case pairs for which a SNP distance was available.

- **Score:** value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case. The lower the value of **Score**, the higher the likelihood that the potential source case was the source of infection for the given case. The value of **Score** is calculated as the sum of the values of **SNP rating**, **Time rating**, **Infectious Rating**, and **Epi and Risk Factor Rating**. These values are calculated by LITT and also reported in the **potential sources** worksheet of the **(Prefix)LITT\_All\_Potential\_Sources** output file. Values of **Score** range from 0–8 and are not calculated if there is no value for SNP distance reported for a pair of cases; in this instance, a value of **Without SNP Score** is calculated instead.

- **Without SNP Score:** value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case. The lower the value of **Without SNP Score**, the higher the likelihood that the potential source case was the source of infection for the given case. The value of **Without SNP Score** is calculated as the sum of the **Time rating**, **Infectious Rating**, and **Epi and Risk Factor Rating** (i.e., SNP data are missing). These values are calculated by LITT and also reported in the **potential sources** worksheet of the **(Prefix)LITT\_All\_Potential\_Sources** output file. **Without SNP Score** values allow for all potential source cases to be compared when some are missing SNP values, for example because no *M. tuberculosis* isolate was successfully collected and/or sequenced. Values of **Without SNP Score** values range from 0–5.
- **Score Category:** categorical classification of the calculated value of **Score** (and/or value of **Without SNP Score**) describing the likelihood that the specified given case was infected by the associated potential source case. Categories are: “High likelihood” (**Score** or **Without SNP Score** of “0” or “1”), “Medium likelihood” (“2”  $\leq$  **Score** < 3; **Without SNP Score** = “2”), “Low likelihood” (“4”  $\leq$  **Score** < 8; “3”  $\leq$  **Without SNP Score** < 4).
- **SNP Rating:** value equivalent to the SNP distance between the *M. tuberculosis* isolates from the specified given and potential source cases. The **SNP Rating** is used to impose a penalty on the value of **Score** proportional to the SNP distance between isolates from the two cases. If no SNP distance is reported, a value of “No SNP data” will appear.
- **Infectious Rating:** value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case (i.e., the lower the value, the higher the likelihood of infection) based on the infectiousness of the potential source case. Infectiousness is determined by sputum smear examination(s) that were positive for any AFB and/or one or more chest radiograph(s) showing evidence of one or more lung cavities. The **Infectious Rating** is used to impose a penalty on the value of **Score** (or the value of **Without SNP Score**) if the potential source case did not have sputum smear examination(s) that were positive for AFB and/or one or more chest radiograph(s) showing evidence of one or more lung cavities.
- **Time Rating:** value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case based on the timing of their respective infections (i.e., the lower the value, the higher the likelihood of infection). The **Time Rating** is used to impose a penalty on the **Score** (or the value of **Without SNP Score**) if a) the value (date) of **Infectious Period Start** for the potential source case is after but within three months of the value (date) of **Infectious Period Start** (or **Infection Acquisition End**) for the given case or b) the value (date) of **Infectious Period End** for the potential source case is after the value (date) of **Infection Acquisition Start** for the given case but before the **Infectious Period Start** (or **Infection Acquisition End**) for the given case.

- **Epi and Risk Factor Rating:** value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case based on epi links or risk factors shared between the cases (i.e., the lower the value, the higher the likelihood of infection). The **Epi and Risk Factor Rating** is used to impose a penalty on the **Score** (or the value of **Without SNP Score**) if there are weak or no epi link(s) between the given and potential source cases and/or if the cases do not share all, some, or any risk factors.

**Note:** Risk factors are only analyzed when the cases do not share an epi link.

- **Investigation Presumed Source:** text specifying the source case identified by the user as the presumed source of infection for the specified given case, and the level of confidence that the user had in making that identification (e.g., “Definite presumed source” or “Probable presumed source”). These data are based on values of **Presumed Source** and **Presumed Source Strength** (optional variables) in the **data** worksheet of the **Case data table** input file. If the value of **Case ID** of the potential source case was listed as a value of **Presumed Source** for the given case in the **Case data table** input file, then the value of **Investigation Presumed Source** will combine the value of **Presumed Source Strength** in the **Case data table** input file with the phrase “presumed source” (e.g., “Definite presumed source”). This information can be used to facilitate comparisons between LITT outputs and the user’s presumed source determination(s). If the value of **Case ID** of the potential source case was not listed as a value of **Presumed Source** for the given case in the **Case data table** input file, then this cell will be blank.
- **Label:** description of the relationship between the specified given and potential source cases. This text is pulled from a) the **Label** variable in the **data** worksheet of the **Epi link table** input file (if data on epi links were included), and/or b) risk factor information in the **data** worksheet of the **Case data table** input file (if data on risk factors were included).

The second worksheet in the **(Prefix)LITT\_All\_Potential\_Sources** output file, named **filtered cases** (*Figure 24*), contains information about all case pairs for which LITT determined that the potential source case could not have been the source of infection for the given case. In the **filtered cases** worksheet, LITT provides the reason why each potential source case was filtered out as a potential source of infection for the associated given case.

A	B	C	D	E
Given Case	Filtered Case	Reason Filtered	Investigation Presumed Source	
1	Case 1	pediatric		
2	Case 1	Case 2	20 SNPs	
3	Case 1	Case 4	21 SNPs	
4	Case 1	Case 6	21 SNPs	
5	Case 1	Case 7	25 SNPs	
6	Case 1	Case 8	26 SNPs	
7	Case 1	Case 5	given case IP start - source IP start = -4.57 months with 3 month cutoff	
8	Case 2	Case 1	extrapulmonary	
9	Case 2	Case 4	20 SNPs	
10	Case 2	Case 6	21 SNPs	
11	Case 2	Case 7	25 SNPs	
12	Case 2	Case 8	26 SNPs	
13	Case 3	Case 1	extrapulmonary	
14	Case 3	Case 2	pediatric	
15	Case 3	Case 4	21 SNPs	
16	Case 3	Case 6	22 SNPs	
17	Case 3	Case 7	26 SNPs	
18	Case 3	Case 8	27 SNPs	
19	Case 3	Case 5	given case IP start - source IP start = -9.6 months with 3 month cutoff	
20	Case 4	Case 1	extrapulmonary	
21	Case 4	Case 2	pediatric	
22	Case 4	Case 3	21 SNPs	
23	Case 4	Case 5	21 SNPs	
24	Case 4	Case 8	6 SNPs	
25	Case 4	Case 6	given case IP start - source IP start = -4.41 months with 3 month cutoff	
26	Case 4	Case 7	score >= 8 (SNP=5;time=2;infectiousness=0;epi=2;score=9)	
27	Case 5	Case 1	extrapulmonary	
28	Case 5	Case 2	pediatric	
29	Case 5	Case 4	21 SNPs	
30	Case 5	Case 6	22 SNPs	
31	Case 5	Case 7	26 SNPs	
32	Case 5	Case 8	27 SNPs	
33	Case 6	Case 1	extrapulmonary	
34	Case 6	Case 2	pediatric	
35	Case 6	Case 3	22 SNPs	
36	Case 6	Case 5	22 SNPs	
37	Case 6	Case 7	6 SNPs	
38	Case 6	Case 8	7 SNPs	
39	Case 7	Case 1	extrapulmonary	
40	Case 7	Case 2	pediatric	
41	Case 7	Case 3	26 SNPs	
42	Case 7	Case 5	26 SNPs	
43	Case 7	Case 6	6 SNPs	
44	Case 7	Case 8	given case IP start - source IP start = -3.78 months with 3 month cutoff	
45	Case 8	Case 1	extrapulmonary	
46	Case 8	Case 2	pediatric	
47	Case 8	Case 3	27 SNPs	
48	Case 8	Case 4	6 SNPs	
49	Case 8	Case 5	27 SNPs	
50	Case 8	Case 6	7 SNPs	
51				
52				

◀ ▶ | potential sources **filtered cases** +

**Figure 24.** Screenshot of the **filtered cases** worksheet within the **(Prefix)LITT\_All\_Potential\_Sources** output file.

Data for the following variables are included in the **filtered cases** worksheet:

- **Given Case:** the value of **Case ID** for the given case in the case pair. LITT evaluated the likelihood that the specified given case was infected by the associated potential source case.

- **Filtered Case:** the value of **Case ID** for the potential source case in the case pair that was filtered out. LITT evaluated the likelihood that the specified given case was infected by the associated potential source case.
- **Reason Filtered:** The reason why LITT filtered out the potential source case as a possible source of infection for the specified given case.

**Note:** While a potential source case may be filtered out for multiple reasons, LITT will only report one.

- **Investigation Presumed Source:** as described above, text specifying the source case identified by the user as the presumed source of infection for the specified given case and the level of confidence that the user had in making that identification (e.g., “Definite presumed source” or “Probable presumed source”).
- 3. **(Prefix)LITT\_Calculated\_Case\_Data:** this file contains a single worksheet named **case data (*Figure 25*)**. This worksheet contains a summary of the demographic and epidemiologic characteristics of all cases included in the cluster, including content merged from the **Case data table**, **Epi link table**, and **SNP distance matrix** input files as well as a summary of select findings from the LITT analysis.

Case ID	Pediatric	XRAYCAV	SPSMEAR	Extrapulmonary Only	Infectious Period Start	Infectious Period End	Infection Acquisition Start	Infection Acquisition End	Investigation Presumed Source	Investigation Presumed Source Strength	Sequence Available In	Number of Times is Ranked 1st in Potential Source List	Number of Times is Ranked 1st in Potential Source List	Homelessness	HIV_pos	Inj_drug_use	Symptom Onset	Sex	Country_origin	Race_ethnicity
Case 1	N	N	NEG	Y			04/01/2016	05/01/2018			Y	2	0	N	N	N				
Case 2	Y	N	NEG	N							Y	1	0	N	N	N		Female	US born	Asian
Case 3	N	Y	POS	N	09/01/2017	05/31/2018			Case 3	Definite	Y	3	3	Y	Y	Y	12/01/2017	Male	non-US born	Asian
Case 4	N	Y	POS	N	01/01/2018	07/15/2018					Y	0	2	Y	Y	Y	04/01/2018	Male	non-US born	Asian
Case 5	N	Y	POS	N	06/20/2018	11/15/2018					Y	2	0	Y	N	Y	09/20/2018	Male	US born	White
Case 6	N	N	POS	N	05/15/2018	08/01/2018					Y	0	0	Y	N	Y	08/15/2018	Male	US born	White
Case 7	N	Y	POS	N	02/20/2018	07/15/2018					Y	0	1	Y	Y	Y	05/20/2018	Female	US born	White
Case 8	N	Y	POS	N	06/15/2018	10/01/2018			Case 7	Probable	Y	0	0	Y	Y	Y	09/15/2018	Male	non-US born	Black

Case ID	Pediatric	XRAYCAV	SPSMEAR	Extrapulmonary Only	Infectious Period Start	Infectious Period End	Infection Acquisition Start	Infection Acquisition End	Investigation Presumed Source	Investigation Presumed Source Strength
Case 1	N	N	NEG	Y					02/01/2018	
Case 2	Y	N	NEG	N			04/01/2016	05/01/2018	Case 3	Definite
Case 3	N	Y	POS	N	09/01/2017	05/31/2018				
Case 4	N	Y	POS	N	01/01/2018	07/15/2018				
Case 5	N	Y	POS	N	06/20/2018	11/15/2018				
Case 6	N	N	POS	N	05/15/2018	08/01/2018				
Case 7	N	Y	POS	N	02/20/2018	07/15/2018				
Case 8	N	Y	POS	N	06/15/2018	10/01/2018			Case 7	Probable

Sequence	Number of Times is Ranked 1st	Available In	Number of in Potential	Analysis	Epi Links	Source List	Homelessness	HIV_pos	Inj_drug_use	Symptom Onset	Sex	Country_origin	Race_ethnicity
Y	2	0	N								Female	US born	Asian
Y	1	0	N								Female	US born	Asian
Y	3	3	Y								Male	non-US born	Asian
Y	0	2	Y								Male	non-US born	Asian
Y	2	0	Y								Male	US born	White
Y	0	0	Y								Male	US born	White
Y	0	1	Y								Female	US born	White
Y	0	0	Y								Male	non-US born	Black

**Figure 25.** Screenshot of the **case data** worksheet within the **(Prefix)LITT\_Calculated\_Case\_Data** output file (top panel) and below it, magnified images of left-hand portion (middle panel) and right-hand portion (bottom panel) of the **case data** worksheet.

Data for the following variables are included in the **(Prefix)LITT\_Calculated\_Case\_Data** output file.

Values for variables designated with an “\*” are imported directly from the **Case data table** input file by LITT:

- **Case ID:** unique case identifier.\*
- **Pediatric:** status of the case as to being pediatric (less than 10 years of age).\*
- **XRAYCAV:** status of the case as to having one or more chest radiographs showing evidence of one or more lung cavities.\*
- **SPSMEAR:** status of the case as to having results of any sputum smear examinations that were positive for any acid-fast bacilli (AFB), of which *M. tuberculosis* is one.\*

- **Extrapulmonary Only:** status of the case as to having only extrapulmonary TB (i.e., does not have either pulmonary or laryngeal disease).\*
  - **Infectious Period Start:** start date of the case's infectious period (for adult cases with pulmonary or laryngeal disease).\*
  - **Infectious Period End:** end date of the case's infectious period (for adult cases with pulmonary or laryngeal disease).\*
  - **Infection Acquisition Start:** for pediatric and extrapulmonary-only cases, the earliest date that the case could have been infected.\*
  - **Infection Acquisition End:** for pediatric and extrapulmonary-only cases, the latest date that the case could have been infected.\*
  - **Investigation Presumed Source:** value (or comma separated list of the values) of **Case ID(s)** indicating the user's presumed source determination(s) (i.e., one or more most likely potential source case(s) for the given case, as identified by the user prior to performing the LITT analysis).\*
  - **Investigation Presumed Source Strength:** value (or comma separated list of values) of numbers or text data indicating the confidence the user has in their presumed source determination(s).\*
  - **Sequence Available in Analysis:** whether or not the specified case had available WGS data from which SNP distances could be calculated. These data are based on inclusion of the specified case in the **SNP distance matrix** input file.
  - **Number of Epi Links:** the number of epi links associated with the specified case based on the number of times it was part of an epi linked pair of cases in the **Epi link table** input file.
  - **Number of Times is Ranked 1st in Potential Source List:** the number of times the specified case was ranked as the most likely potential source case for another case in the cluster by the LITT analysis.
  - **Risk factor(s):** one or more columns containing information on risk factors for transmission or disease development (e.g., homelessness, HIV infection, injection drug use).\*
  - **Symptom Onset:** earliest date that a case first experienced symptoms consistent with TB disease.\*
  - **Visualization characteristic(s):** one or more columns containing text for use in labelling visualizations generated from results of a LITT analysis (e.g., case characteristics such as sex, country of origin, race/ethnicity).\* These columns will only be included when the **Output extra columns in case data table** analysis parameter box is checked (B in *Figure 14*).
4. **(Prefix)LITT\_Calculated\_Epi\_Data:** this file contains a single worksheet named **epi links** (*Figure 26*). This worksheet reports a de-duplicated list of all epi links present in the **Epi link table** input file and appends SNP distances (if available) to each pair of epi linked cases.

	A	B	C	D	E
1	Case1	Case2	Epi Link Strength	Label	SNP Distance
2	Case 1	Case 3	definite	Homeless shelter A	1
3	Case 1	Case 5	definite	Homeless shelter A	1
4	Case 2	Case 3	definite	Household	1
5	Case 3	Case 5	definite	Homeless shelter A	2
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					

[epi links](#) [+](#)

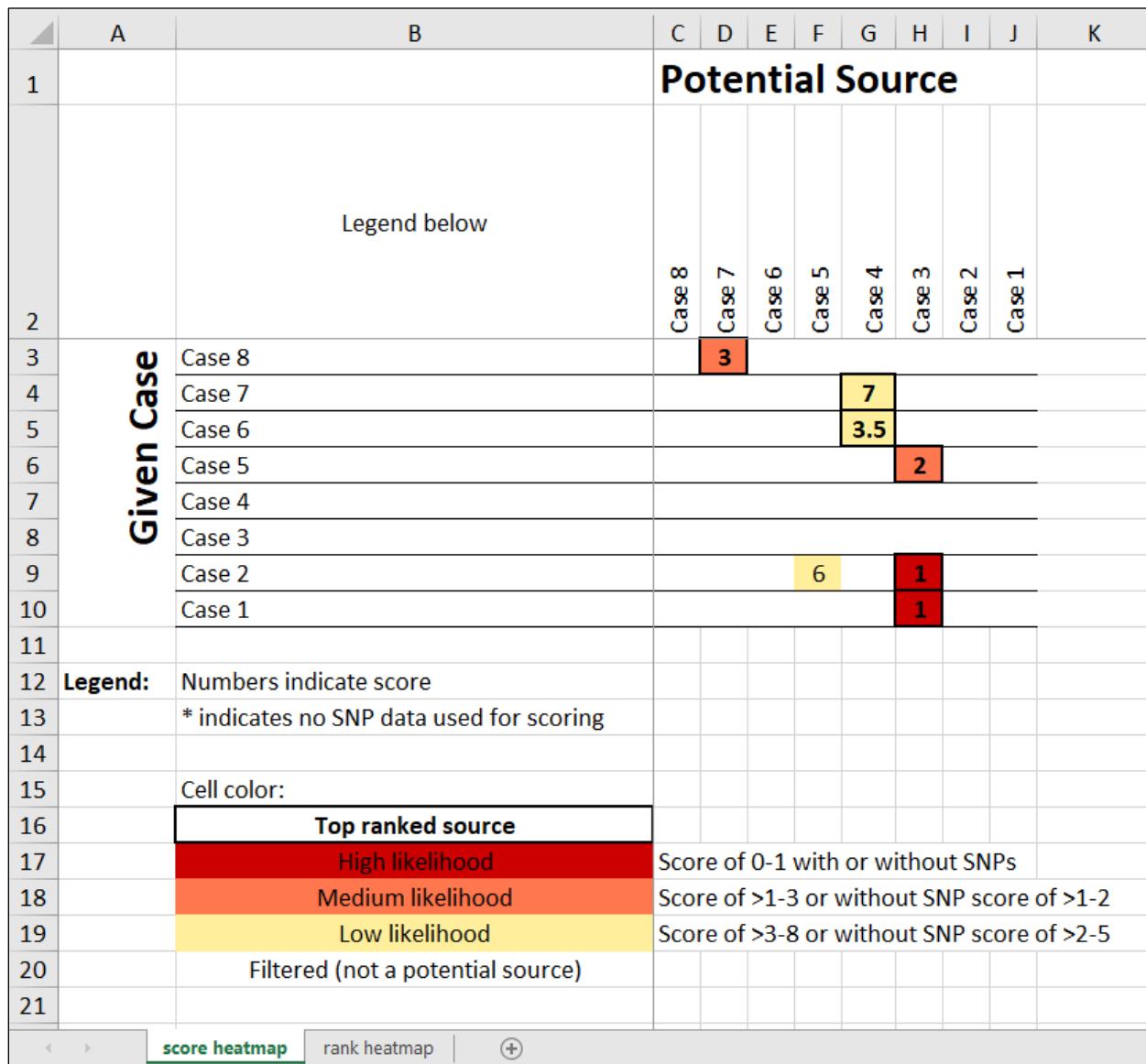
**Figure 26.** Screenshot of the [Epi links](#) worksheet within the [\(Prefix\)LITT\\_Calculated\\_Epi\\_Data](#) output file.

Data for the following variables are included in the [\(Prefix\)LITT\\_Calculated\\_Epi\\_Data](#) output file.

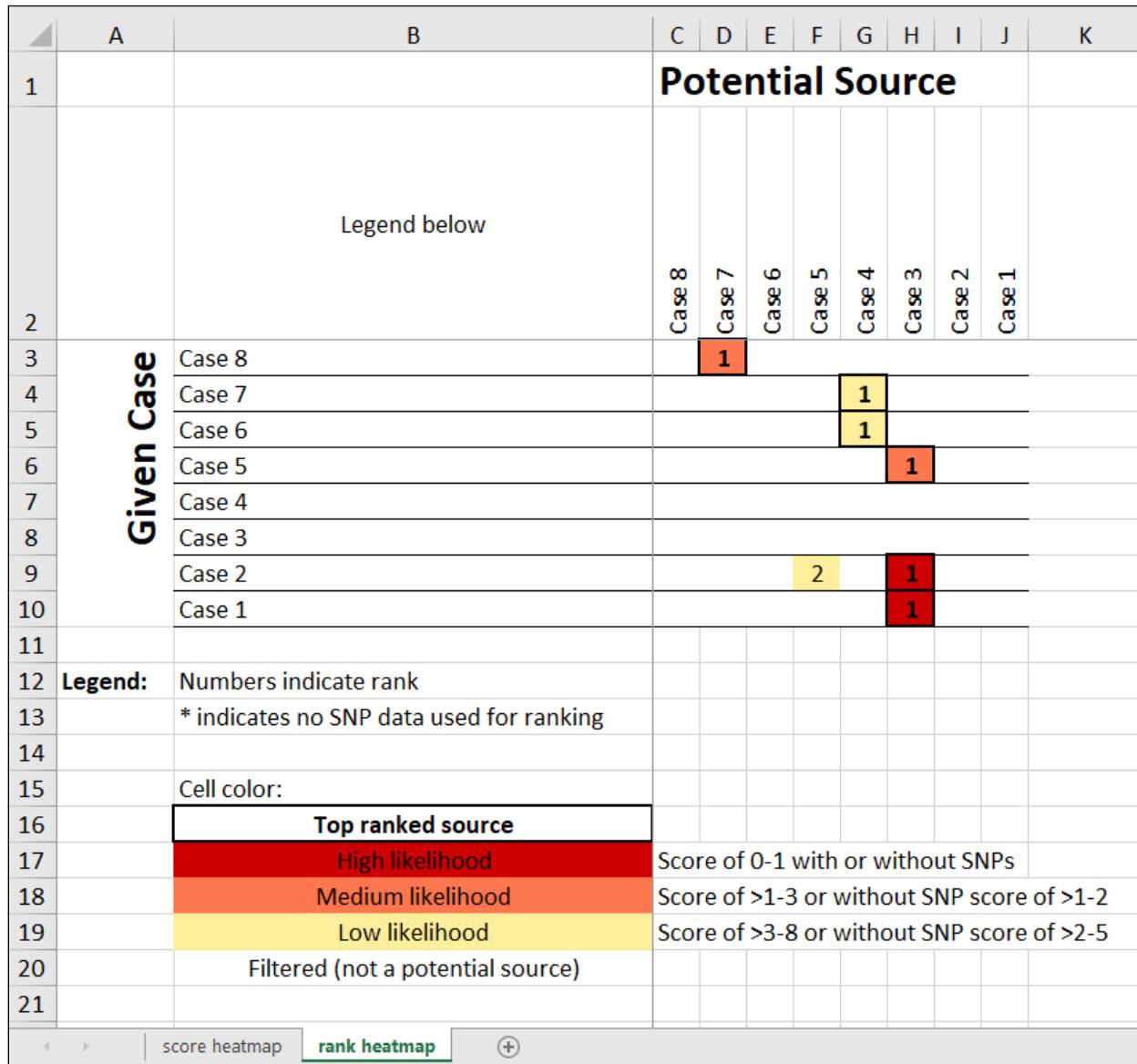
Values for variables designated with an “\*” are imported directly from the [Epi link table](#) input file by LITT:

- **Case1**: unique case identifier ([Case ID](#)) for one case in an epi linked pair of cases.\*
- **Case2**: unique case identifier ([Case ID](#)) for the other case in the epi linked pair of cases.\*
- **Epi Link Strength**: strength of the epi link between the pair of epi linked pair of cases.\* If a cell in the Epi Link Strength column of the [Epi link table](#) input file was left blank, LITT will have assigned a default value of “Probable” during the analysis.
- **Label**: text describing the relationship between the pair of epi linked cases.\*

- **SNP Distance:** the SNP distance between the specified pair of epi linked cases. These data are from the **SNP distance matrix** input file (if isolates from both specified cases are included in the **SNP distance matrix** input file).
5. **(Prefix)LITT\_Heatmap:** this file contains two worksheets, named **score heatmap** (*Figure 27*) and **rank heatmap** (*Figure 28*), which provide visual representations of values of **Score** and **Rank**, respectively, reported in the **potential sources** worksheet of the **(Prefix)LITT\_All\_Potential\_Sources** output file.



**Figure 27.** Screenshot of the **score heatmap** worksheet within the **(Prefix)LITT\_Heatmap** output file.



**Figure 28.** Screenshot of the **rank heatmap** worksheet within the **(Prefix)LITT\_Heatmap** output file.

Both worksheets present a matrix in which every case in the cluster being analyzed is listed as both a row (given case) and as a column (potential source case). Each cell in the matrix reports the value of **Score** (or **Without SNP Score**) (or **Rank** or **Without SNP Rank**) calculated by LITT for the likelihood that the

specified potential source case was the source of infection for the specified given case. The heatmaps are intended to facilitate comparisons of given cases and potential source cases with several key layout features:

- For each given case the cell associated with the potential source case with the lowest value of **Score** or **Rank** has a bold border.
- Cells of both matrices are color coded according to values of **Score** (or **Without SNP Score**) such that the lowest scores (highest likelihood of infection) are colored dark red and the highest scores (lowest likelihood of infection) are color coded pale yellow (i.e., the heat mapping).
- For case pairs with no WGS data available, **Without SNP Score** and **Without SNP Rank** are used and designated with an “\*.”

If LITT filtered out a case as a potential source of infection for a given case, the associated cell of the matrix is left blank (i.e., no value for **Score** or **Rank**, no shading).

6. **(Prefix)LITT\_Risk\_Factor\_Weights**: this file contains a single worksheet named **RFweights** (*Figure 29*). This worksheet reports all risk factors present in the **Table of risk factor weights** input file and provides the normalized weights calculated for the LITT analysis. The normalization process scales each raw weight value such that individual weight values fall between “0” and “1,” and the sum of all normalized weight values equals 1.0.

	A	B
1	variable	weight
2	Homelessness	0.166666667
3	HIV_pos	0.5
4	Inj_drug_use	0.333333333
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		

Figure 29. Screenshot of the **RFweights** worksheet within the **(Prefix)LITT\_Risk\_Factor\_Weights** output file.

Data for the following variables are included in the [\(Prefix\)LITT\\_Risk\\_Factor\\_Weights](#) output file. Values for variables designated with an “\*” are imported directly from the [Table of risk factor weights](#) input file by LITT:

- **variable**: name of the risk factor.\*
  - **weight**: numerical value assigned to the risk factor by the user to specify the importance of that factor in terms of transmission relative to other factors. These data are from the [Table of risk factor weights](#) input file and have been normalized by LITT to range between 0–1.
7. [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#): this file contains a single worksheet named [top ranked sources \(Figure 30\)](#). For each given case with at least one potential source case in the [potential sources](#) worksheet of the [\(Prefix\)LITT\\_All\\_Potential\\_Sources](#) output file, this worksheet contains information about the potential source case identified by LITT as the most likely source case (i.e., the potential source case with the lowest value of **Score** when paired with the specified given case). If more than one potential source case was associated with the lowest value of **Score** when paired with the specified given case, all of these potential source cases will be listed. This abbreviated list is provided for convenience for visualization; users are encouraged to consult the full list of given case / potential source case pairs in the [\(Prefix\)LITT\\_All\\_Potential\\_Sources](#) output file for a complete review of results of a LITT analysis.

	A	B	C	D	E	F
1	Given Case	Potential Source	Score	Without SNP Score	Label	Score Category
2	Case 1	Case 3	1		Homeless shelter A	high likelihood
3	Case 2	Case 3	1		Household	high likelihood
4	Case 5	Case 3	2		Homeless shelter A	medium likelihood
5	Case 6	Case 4	3.5		RF: Homelessness   RF: Inj_drug_use	low likelihood
6	Case 7	Case 4	7		RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use	low likelihood
7	Case 8	Case 7	3		RF: Homelessness   RF: HIV_pos   RF: Inj_drug_use	medium likelihood
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

**Figure 30. Screenshot of the top ranked sources worksheet within the (Prefix)LITT\_Top\_Rank\_Potential\_Source output file.**

Data for the following variables are included in the (Prefix)LITT\_Top\_Rank\_Potential\_Source output file:

- Given case: the value of Case ID for the given case in the case pair.
- Potential Source: the value of Case ID for the potential source case in the cluster that LITT identified as the most likely source of infection for the specified given case.
- Score: value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case. The value of Score is calculated as the sum of the values of SNP rating, Time rating, Infectious Rating, and Epi and Risk Factor Rating.
- Without SNP Score: value inversely proportional to the likelihood that the specified given case was infected by the associated potential source case. The value of Without SNP Score is calculated as the sum of the Time rating, Infectious Rating, and Epi and Risk Factor Rating.
- Label: description of the relationship between the specified given and potential source cases. This text is pulled from a) the Label variable in the data worksheet of the Epi link table input file (if data on epi

links were included), and/or b) risk factor information in the **data** worksheet of the **Case data table** input file (if data on risk factors were included).

- **Score Category:** categorical classification of the calculated value of **Score** (and/or value of **Without SNP Score**) describing the likelihood that the specified given case was infected by the associated potential source case. Categories are: “High likelihood” (**Score** or **Without SNP Score** of “0” or “1”), “Medium likelihood” (“2”  $\leq$  **Score** < 3; **Without SNP Score** = “2”), “Low likelihood” (“4”  $\leq$  **Score** < 8; “3”  $\leq$  **Without SNP Score** < 4).
8. **(Prefix)LITT\_Distance\_Matrix:** this file contains a single worksheet named **Distance Matrix** (*Figure 31*). It returns the matrix of SNP distances uploaded for the LITT analysis as the **SNP distance matrix** input file. This file is only generated when the analysis parameter box **Include distance matrix in outputs** is checked (E in *Figure 14*).

	A	B	C	D	E	F	G	H	I	J
1		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	MRCA
2	Case 1	0	0	1	20	1	21	25	26	10
3	Case 2	0	0	1	20	1	21	25	26	10
4	Case 3	1	1	0	21	2	22	26	27	11
5	Case 4	20	20	21	0	21	1	5	6	10
6	Case 5	1	1	2	21	0	22	26	27	11
7	Case 6	21	21	22	1	22	0	6	7	11
8	Case 7	25	25	26	5	26	6	0	1	15
9	Case 8	26	26	27	6	27	7	1	0	16
10	MRCA	10	10	11	10	11	11	15	16	0
11										
12										
13										
14										
15										
16										
17										
18										

**Figure 31.** Screenshot of the Distance Matrix worksheet within the **(Prefix)LITT\_Distance\_Matrix** output file.

## **B. Using analytic outputs**

While there is no single correct process by which users should review and interpret the outputs of a LITT analysis, LITT developers have found the following sequential approach to be intuitive and efficient. This suggested approach helps ensure that the analysis was completed without any errors or issues and then allows users to quickly identify key analytic results.

1. Review the [\(Prefix\)LITT\\_log](#) output file: verify that the number of cases, epi links, and risk factors are equal to expected values based on input data, and that the [SNP cutoff](#) value was set at the desired level. Analytic issues encountered by LITT during the analysis will be listed and briefly described in this output file. Users should trouble shoot the analysis based on these reported issues before moving on to review analytic results.
2. Review the [\(Prefix\)LITT\\_Heatmap](#) output file: identify potential source cases that appear to have been important in disease transmission (e.g., implicated in likely transmission to a large number of cases within the cluster or to particular given case(s) of interest) by orienting to low values of [Score](#) and/or [Rank](#) and cells shaded red and/or with a bold border.
3. Review the [\(Prefix\)LITT\\_All\\_Potential\\_Sources](#) output file: perform a more detailed, in depth review of how values calculated for the variables [SNP rating](#), [Infectious Rating](#), [Time Rating](#), and [Epi](#) and [Risk Factor Rating](#) contributed to either a) the values of [Score](#) and [Rank](#) calculated for each given case/potential source case pair (the [potential sources](#) worksheet), or b) the filtering out of cases as potential sources of infection for particular given cases (the [filtered cases](#) worksheet).
4. Review the [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#) output file: review the top ranked potential source case for each given case in the cluster, noting the value of [Score](#) and the epi links and/or risk factors associated with each pair of cases. The information in this output file is also contained within the [\(Prefix\)LITT\\_All\\_Potential\\_Sources](#) output file; this file was largely created for use in visualizing networks (see [below](#)).
5. Review and safeguard remaining files: the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#), [\(Prefix\)LITT\\_Calculated\\_Epi\\_Data](#), [\(Prefix\)LITT\\_Distance\\_Matrix](#), and [\(Prefix\)LITT\\_Risk\\_Factor\\_Weights](#) files provide a record of the input data used by LITT for the analysis. Users can review the contents of these files to confirm that the input data are consistent with those uploaded at the start of the analysis. As input data change (e.g., new or updated case data are added as an investigation proceeds), these files can serve as archived reference data associated with prior LITT analyses on the same cluster.

**Note:** The [\(Prefix\)LITT\\_Risk\\_Factor\\_Weights](#) file will contain value of **Weight** that have been standardized such that they sum to 1.0.

Following results review, users can visualize findings generated by a LITT analysis using separate, stand-alone data visualization platforms. Specifically, LITT was designed to output data in formats that can be readily imported into MicrobeTrace, a software tool developed by the CDC for rapid visualizations of networks and associated data. Like LITT, MicrobeTrace is available as an interactive webpage hosted by the CDC and can be used without installing any program files on the user's computer. MicrobeTrace can be accessed using this [\*link\*](#).

The MicrobeTrace user's manual, along with a variety of information on development and use of the software tool, can be accessed at the [MicrobeTrace GitHub](#) webpage using this [\*link\*](#). Either the Chrome or Firefox web browser must be used.

Please see [\*Appendix 3\*](#) for instructions on how to use MicrobeTrace to visualize a two dimensional TB transmission network using output files generated by a LITT analysis.

### **C. Limitations**

When interpreting outputs of a LITT analysis, users should be mindful of associated limitations. Chief among these are limitations related to data completeness and quality. LITT has no way of detecting missing cases (i.e., cases that are in fact part of the cluster but were not included in the analytic dataset); the algorithm may identify one or more potential source cases for a given case even if the true source case was erroneously omitted from the analytic dataset. Additionally, while complete data on SNP distances, epi links, and risk factors are not required to complete a LITT analysis, these data can strongly influence and improve the accuracy of LITT outputs. For example, in instances where input data are limited in quality, completeness, and/or breadth, LITT may identify one or more potential source cases for a given case even if the given case was erroneously included in the analytic dataset (i.e., not actually part of the cluster). Conducting thorough, systematic, and rigorous epidemiologic and contact investigations, and carefully managing associated data, represent the best approaches to overcoming these limitations. LITT outputs are intended to complement, not replace, traditional methods of epidemiologic field investigations and local knowledge about clusters and associated disease transmission.

## Training datasets

To help users develop familiarity with performing and interpreting output from a LITT analysis, and to empirically elucidate some of the sensitivities and limitations of LITT, one or more training datasets have been created. These training datasets can be downloaded from the [TB molecular epidemiology GitHub](#) webpage, which can be accessed by clicking on the [Help](#) (“?”) button in the LITT online user interface ([Figure 14](#)) or by using this [link](#). Each training dataset includes the following:

- One of each of the input files for an analysis populated with hypothetical data (e.g., [Case data table](#), [Epi link table](#), [SNP distance matrix](#), and [Table of risk factor weights](#)).
- Visual representations of the case data including a diagram showing epi links and risk factors and a phylogenetic tree (constructed using information in the [SNP distance matrix](#) input file).

Training dataset input files are formatted so that they can be uploaded without any modification to perform a LITT analysis. As such, training datasets can be used most simply to confirm that LITT is working properly and that users are uploading files and selecting analytic options/settings correctly. For example, by performing a LITT analysis using the input files associated with [LITT training dataset 01](#), users can replicate all of the analytic outputs and associated screenshots presented in this user’s manual.

Training datasets can also be used to explore sensitivities and limitations of LITT. For example, in addition to running a full analysis using all four input files included with each training dataset, users can run subset analyses in which one or more of the optional input files ([Epi link table](#), [SNP distance matrix](#), [Table of risk factor weights](#)) is excluded. By comparing results of full and subset analyses, users can investigate the sensitivity of LITT analyses to exclusion of different data types by noting changes in the outputs. Similarly, users can alter or delete values in input files (e.g., change risk factor weights, delete one or more epi links from the [Case data table](#) and [Epi link table](#) input files, delete one or more risk factors from the [Case data table](#) and [Table of risk factor weights](#) input files) to explore resulting changes in LITT outputs. Finally, users can delete one or more cases from the [Case data table](#), [Epi link table](#), and [SNP distance matrix](#) input files to demonstrate how LITT outputs can change during the course of a public health investigation as additional cases in a cluster are identified.

## Contacts and additional resources

Users should direct all LITT-related questions or comments to their DTBE point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov). Note that CDC technical assistance may not always be available for all user analyses and that the availability of these support services may be discontinued at any time.

Users interested in reviewing or running the LITT R code can obtain it on the [TB molecular epidemiology GitHub](#) webpage, which can be accessed by clicking on the [Help \("?"\)](#) button in the LITT online user interface (I in [Figure 14](#)) or by using this [link](#). Running LITT on a local (e.g., health department) computer gives users more flexibility to use personally identifiable information (e.g., names of cases, contacts, or locations) since input data are not transmitted to other computers or servers as part of a local analysis.

## Citations

1. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
2. Winglee et. al., Logically Inferred Tuberculosis Transmission (LITT): A Method to Integrate Whole Genome Sequencing, Clinical, and Epidemiologic Data to Identify and Rank Potential Tuberculosis Sources Cases (*manuscript in preparation*).
3. Lonetti A, Fontana MC, Martinelli G, Iacobucci I (2016). Single nucleotide polymorphisms as genomic markers for high-throughput pharmacogenomic studies. *Atlas of Science*. March 22, 2016. <http://atlasofscience.org/single-nucleotide-polymorphisms-as-genomic-markers-for-high-throughput-pharmacogenomic-studies/>
4. National Tuberculosis Controllers Association and Centers for Disease Control and Prevention (CDC) (2005). Recommendations from the National Tuberculosis Controllers Association and CDC. Guidelines for the investigation of contacts of persons with infectious tuberculosis. *MMWR Recommendations and reports*. 54:1-47
5. Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I (2016). Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC medicine*. 14:21.

6. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet infectious diseases*. 13(2):137-146.

## Acknowledgments

Contributions from the following persons to the conceptualization, development, testing and/or evaluation of LITT are gratefully acknowledged: Martin Cilnis, Lauren Linde, and Tambi Shaw (California Department of Public Health, Tuberculosis Control Branch); Wendy Pernal and Shameer Poonja (Los Angeles County Department of Public Health, Tuberculosis Control Program); Jillian Knorr and Jeanne Sullivan Meissner (New York City Department of Health and Mental Hygiene, Bureau of Tuberculosis Control).

Online hosting of LITT was made possible by staff of OAMD.

Funding for the development of LITT and supporting documentation, and for training of state and county health department staff on use of LITT, was provided by CDC.

## Appendix 1 — Accessing the OAMD web portal

In general, LITT users can be classified into one of four types depending on CDC employment status, CDC identification credentialing, and location of access:

**User type 1: non-CDC users who do not have CDC Secure Access Management Services (SAMS) credentials**

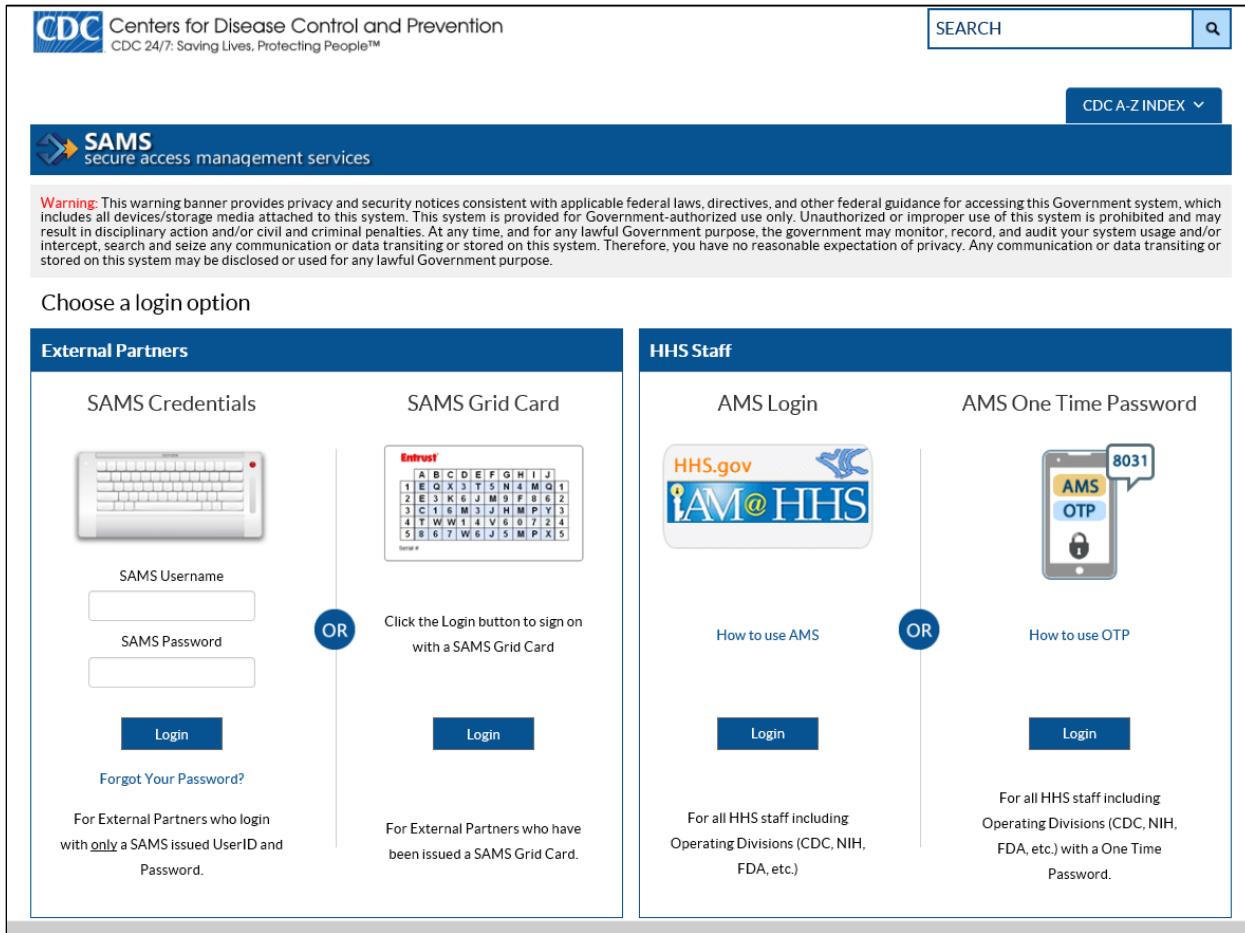
**User type 2: non-CDC users who already have SAMS credentials (e.g., for TB GIMS)**

**User type 3: CDC users accessing from outside the CDC network**

**User type 4: CDC users accessing from within the CDC network**

**User type 1: non-CDC users who do not have SAMS credentials:**

1. Non-CDC users will use non-CDC hardware (i.e. health department computer) to access the OAMD web portal via CDC's SAMS web portal. Access to the SAMS web portal requires credentialing. To initiate the SAMS credentialing process, users must send an email to [dtbesupport@cdc.gov](mailto:dtbesupport@cdc.gov) with the subject line “OAMD TB GIMS new user”; background information, problems, or questions related to the request can be detailed in the body of the email. Some non-CDC users will already have SAMS credentials as part of other work with CDC. Notably, non-CDC users with access to TB GIMS already have the SAMS credentials necessary to access LITT and should consult the instructions for User type 2.
2. Users will receive an email invitation to register with SAMS from [sams-no-reply@cdc.gov](mailto:sams-no-reply@cdc.gov) within approximately two business days. The email will have the subject line “U.S. Centers for Disease Control: SAMS Partner Portal – Invitation to Register” and contain further instructions on how to complete the SAMS registration process. Additional information about the SAMS registration process can be obtained using this [link](#). Additional information about the SAMS identity verification process can be obtained using this [link](#).
3. Once SAMS credentialing is complete, users will need to have the OAMD web portal added to their approved list of accessible online CDC applications. To initiate this addition, users must send an email to their DTBE point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov), with the subject line “OAMD portal access request.” Background information, problems, or questions related to the request can be detailed in the body of the email. Users will be advised via email when the OAMD portal has been added to their approved list of accessible applications.
4. With SAMS credentialing complete and the OAMD web portal added to their approved list of accessible online CDC applications, users can begin the process of accessing LITT by logging into the SAMS web portal via the SAMS login webpage ([Figure 32](#)) using this [link](#). Users will log in using the “SAMS Credentials” option on the left side of the login webpage. Users experiencing problems accessing the SAMS web portal via the SAMS login webpage should send an email to [samshelp@cdc.gov](mailto:samshelp@cdc.gov) to request assistance.



**Figure 32. Screenshot of the login webpage for CDC's Secure Access Management Service (SAMS) web portal.**

5. After successfully logging into the SAMS web portal (*Figure 33*), users will see their approved list of accessible online CDC applications. The OAMD web portal should be listed as hyperlinked text reading “SciComp SAMS Portal” under the heading “OAMD Gateway” (A in *Figure 33*); click on this hyperlinked text to navigate to the OAMD web portal (*Figure 11*). If this hyperlink is not present or working properly, users should send an email to *TBGenotyping@cdc.gov* to request assistance.

The screenshot shows the CDC SAMS web portal. At the top, there is a navigation bar with the CDC logo and a search bar. Below the navigation bar is a blue header bar with the text "SAMS secure access management services". On the left side, there is a sidebar with a "Menu" section containing "My Profile" (with a user icon), "Logout", and a "Links" section with "SAMS User Guide", "SAMS User FAQ", and "Identity Verification Overview". The main content area is titled "My Applications" and lists two items: "OAMD Gateway" and "Tuberculosis Genotyping Information Management System". Under "OAMD Gateway", there is a link "SciComp SAMS Portal" followed by a red annotation "A". Under "Tuberculosis Genotyping Information Management System", there is a link "TB GIMS". At the bottom of the main content area, there is a note "\* Strong credentials required."

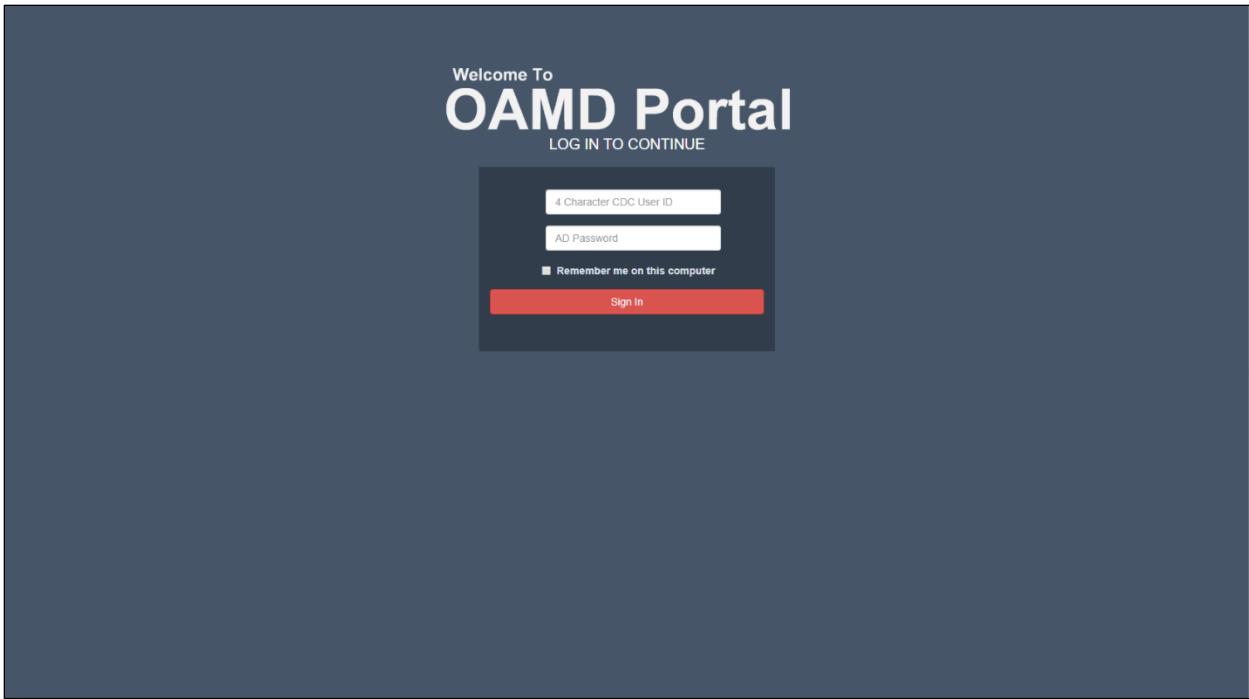
**Figure 33. Annotated screenshot of the CDC’s Secure Access Management Service (SAMS) web portal. Annotation appears in red.**

**User type 2: non-CDC users who already have SAMS credentials (e.g., for TB GIMS):**

1. Non-CDC users will use non-CDC hardware (i.e., health department computer) to access the OAMD web portal via CDC’s SAMS web portal. Access to the SAMS web portal requires credentialing. Some non-CDC users will already have SAMS credentials as part of other work with CDC. Notably, non-CDC users with access to TB GIMS already have the SAMS credentials necessary to access LITT.
2. Users with SAMS credentials need to have the OAMD web portal added to their approved list of accessible online CDC applications. To initiate this addition, users must send an email to their DTBE point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov), with the subject line “OAMD portal access request.” Background information, problems, or questions related to the request can be detailed in the body of the email. Users will be advised via email when the OAMD portal has been added to their approved list of accessible applications.
3. With SAMS credentialing complete and the OAMD web portal added to their approved list of accessible online CDC applications, users can begin the process of accessing LITT by following instructions in User type 1, Step 4 and in all subsequent steps.

**User type 3: CDC users accessing the CDC network indirectly using VPN or CITGO**

1. In order to remain compliant with terms of the National Tuberculosis Surveillance System Assurance of Confidentiality, CDC users must use CDC hardware (i.e., CDC laptop) to access LITT; non-CDC hardware (e.g., health department computer or personal laptop, using a smart card reader) may not be used. Users will first need to be granted access to the OAMD web portal. To request this access, users must send an email to their DTBE point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov), with the subject line “OAMD portal access request.” Background information, problems, or questions related to the request can be detailed in the body of the email. Users will be advised via email when access to the OAMD portal has been granted. After receiving access, users will set up an OAMD web portal password (the user’s ID will be the same four characters of their CDC user ID).
2. Once access to the OAMD web portal has been granted and a password set, users can access LITT by remotely logging into the CDC network via CDC Virtual Private Network (VPN) or via CDC Information Technology on the Go (CITGO). Details about how to remotely log into the CDC network via these methods can be obtained from the CDC Information Technology Services Office (ITSO) by phone (404-639-6000 in Atlanta, 1-888-647-3375 outside Atlanta) or email ([ITSOServicedesk@cdc.gov](mailto:ITSOServicedesk@cdc.gov)). After successfully logging into the CDC network, users can access the OAMD login webpage (*Figure 34*) using this *link*. Note that when using this option, only input files saved to a location on the CDC network (that has been remotely accessed) will be available for LITT analyses; input files saved to the local computer will not be available. Users experiencing problems accessing the OAMD web portal via the OAMD login webpage should send an email to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov) to request assistance.



**Figure 34. Screenshot of the login webpage for CDC's OAMD web portal.**

3. After successfully logging into the OAMD login webpage (*Figure 34*) using their CDC user ID and password, users will be directed to the OAMD web portal (*Figure 11*).

**User type 4: CDC users accessing the CDC network directly through physical connect in a CDC office**

1. Users will first need to be granted access to the OAMD web portal. To request this access, users must send an email to their DTBE point of contact for LITT analyses or, if they do not know who this person is or have not been assigned a point of contact for LITT analyses, to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov), with the subject line “OAMD portal access request.” Background information, problems, or questions related to the request can be detailed in the body of the email. Users will be advised via email when access to the OAMD portal has been granted. After receiving access, users will set up an OAMD web portal password (the user’s ID will be the first four characters of their CDC user ID).
2. Once access to the OAMD web portal has been granted and a password set, users can access the OAMD web portal (*Figure 34*) using this *link*. Users experiencing problems accessing the OAMD web portal via the OAMD login webpage should send an email to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov) to request assistance.

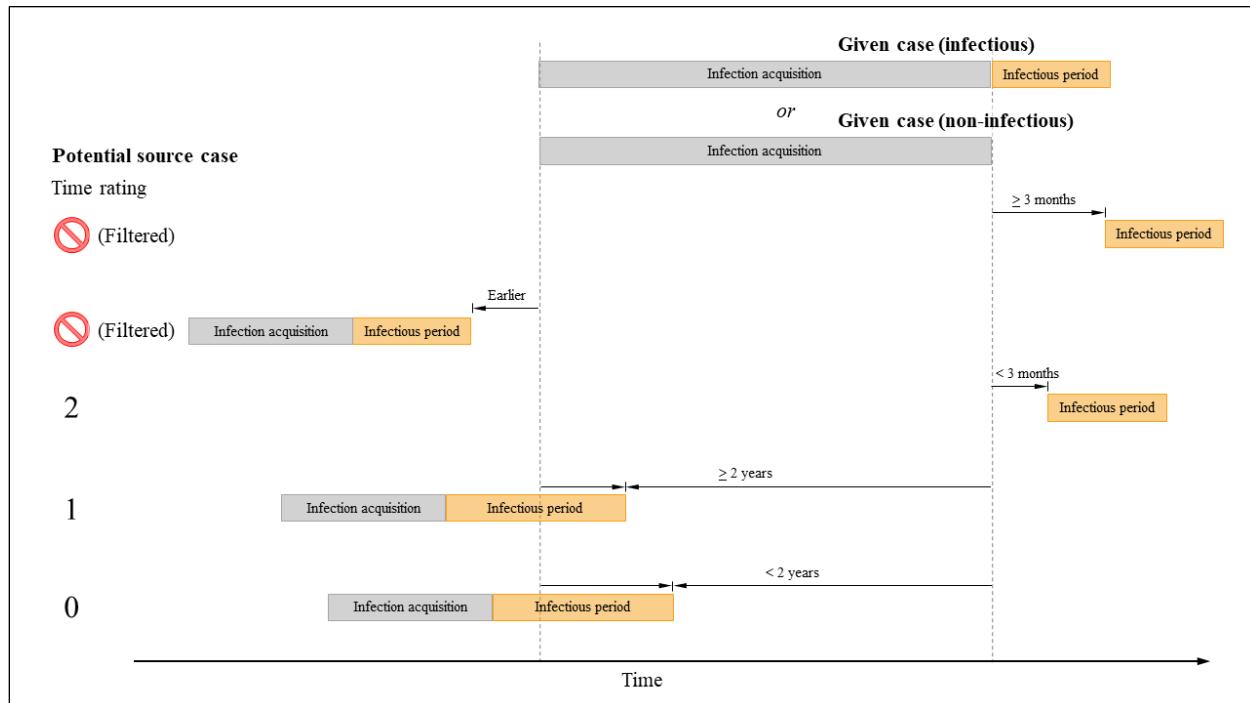
- After successfully logging into the OAMD login webpage (*Figure 34*) using their user CDC user ID and password, users will be directed to the OAMD web portal (*Figure 11*).

#### Note about internet browsers:

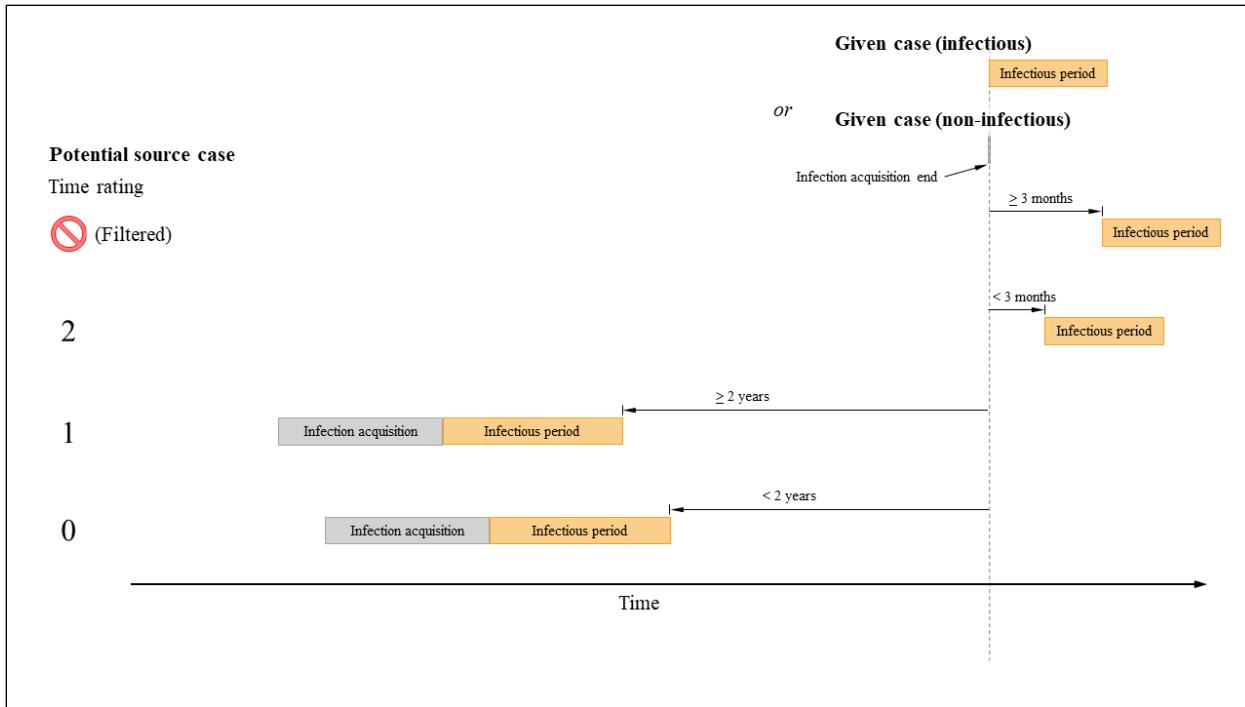
It is possible that on occasion users may experience difficulty logging into SAMS, the OAMD portal, or the TB molecular epidemiology webpage due to idiosyncrasies of their network connection or web browser. If problems are encountered users are encouraged to reboot their computer and attempt to log in using a different browser. LITT seems to consistently perform with few if any issues when used in the Chrome web browser. If the problem(s) persist, users should take screenshots to illustrate the issue(s) they are encountering and attach these to an email sent to [TBGenotyping@cdc.gov](mailto:TBGenotyping@cdc.gov).

## Appendix 2 — Calculation of time rating scores

The following figures visually illustrate how LITT scores the time rating based on the infection acquisition and infectious period start and end dates of given and potential source cases.



**Figure 35. Scoring of time rating when infection acquisition start of given case is known.**

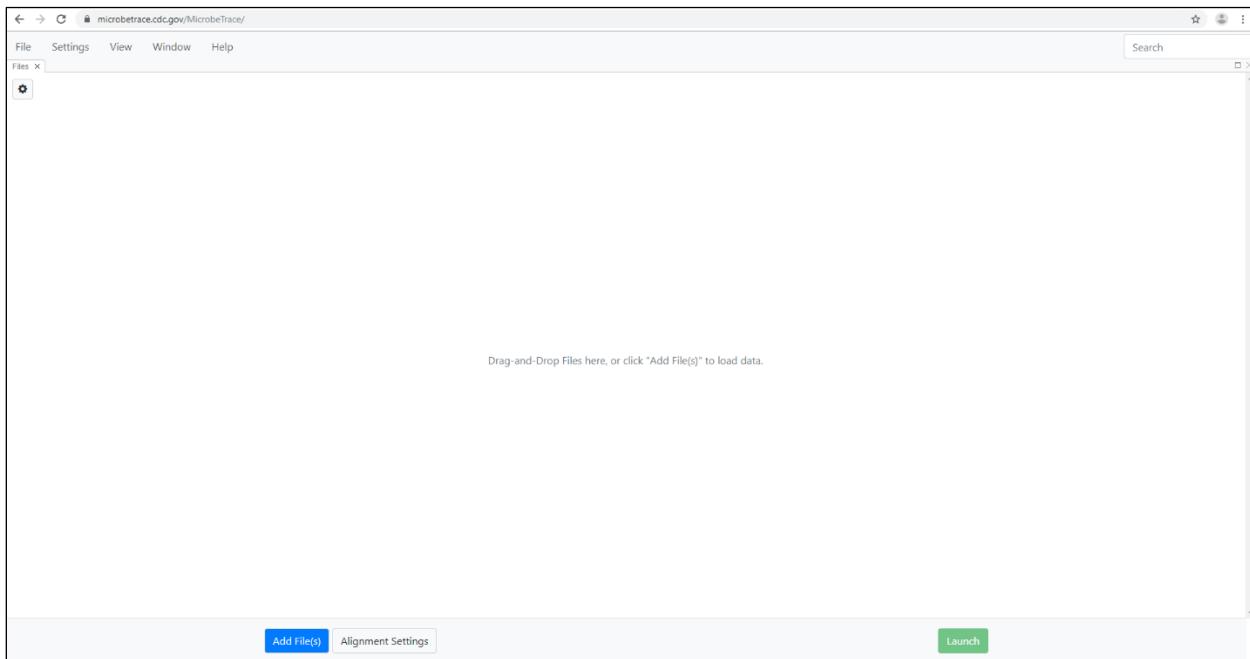


**Figure 36. Scoring of time rating when infection acquisition start of given case is not known.**

## Appendix 3 — Using MicrobeTrace to visualize transmission networks

MicrobeTrace is a software tool created by staff of the Centers for Disease Control and Prevention (CDC) to visualize transmission networks and associated data as part of epidemiologic investigations and public health activities. Within MicrobeTrace, networks can be created using data on epidemiologic links among cases and/or transmission links (with directionality) and be annotated using symbols and/or colors to illustrate risk factors, locations, and case characteristics. Like LITT, MicrobeTrace is available as an interactive webpage hosted by the CDC and can be used without installing any program files on the user's computer. The following provides an example of how to create and annotate a two dimensional TB transmission network in MicrobeTrace using the output files of a LITT analysis based on [LITT\\_training\\_dataset\\_01](#). For additional information about this and numerous other analytic and data visualization functionalities available in MicrobeTrace, please see the MicrobeTrace user's manual posted on the [MicrobeTrace GitHub](#) webpage, which can be accessed using this [link](#).

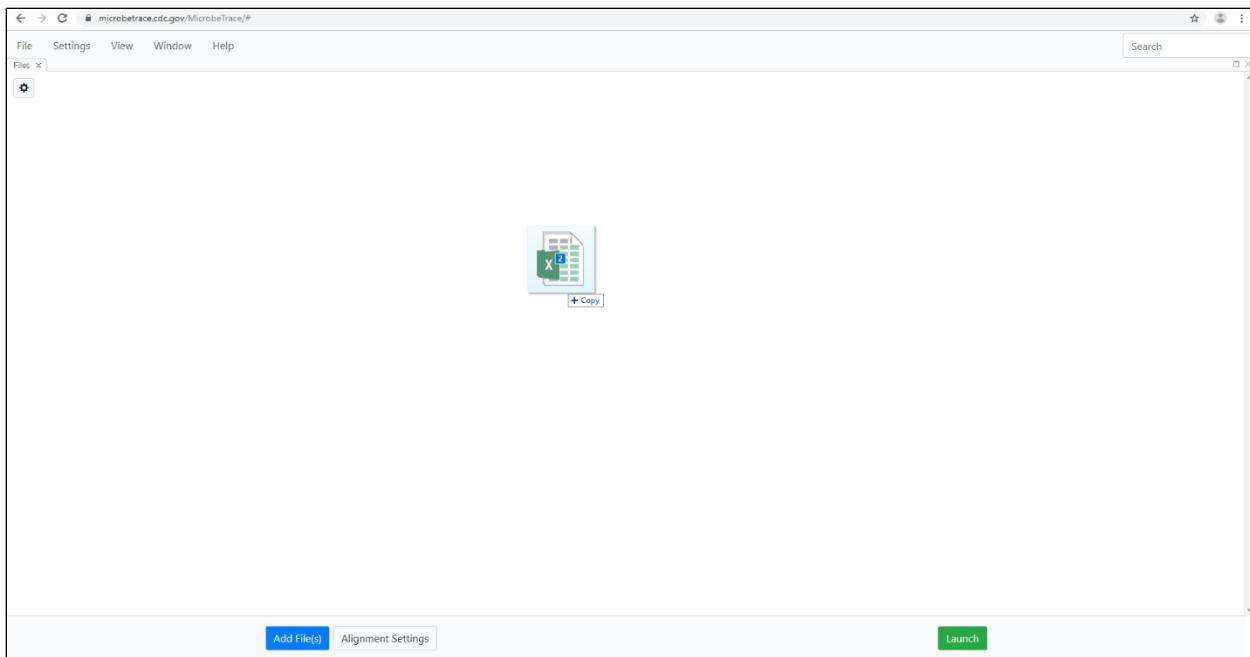
1. Log into the MicrobeTrace data submission webpage (*Figure 37*). MicrobeTrace can be accessed using this [link](#). Either the Chrome or Firefox web browser must be used.



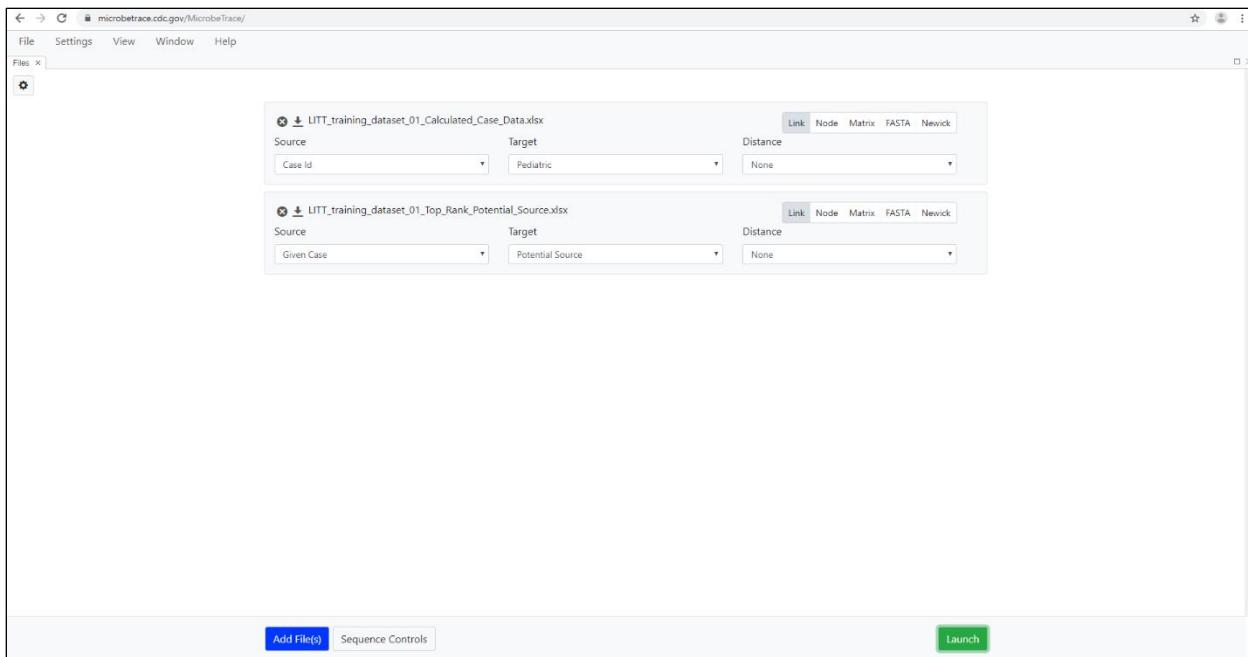
**Figure 37. Screenshot of the MicrobeTrace data submission webpage.**

2. Load the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) and [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#)

LITT output files either by dragging them into the window (*Figure 38*) or using the [Add File\(s\)](#) button. Once successfully loaded, both files will appear as windows (*Figure 39*).

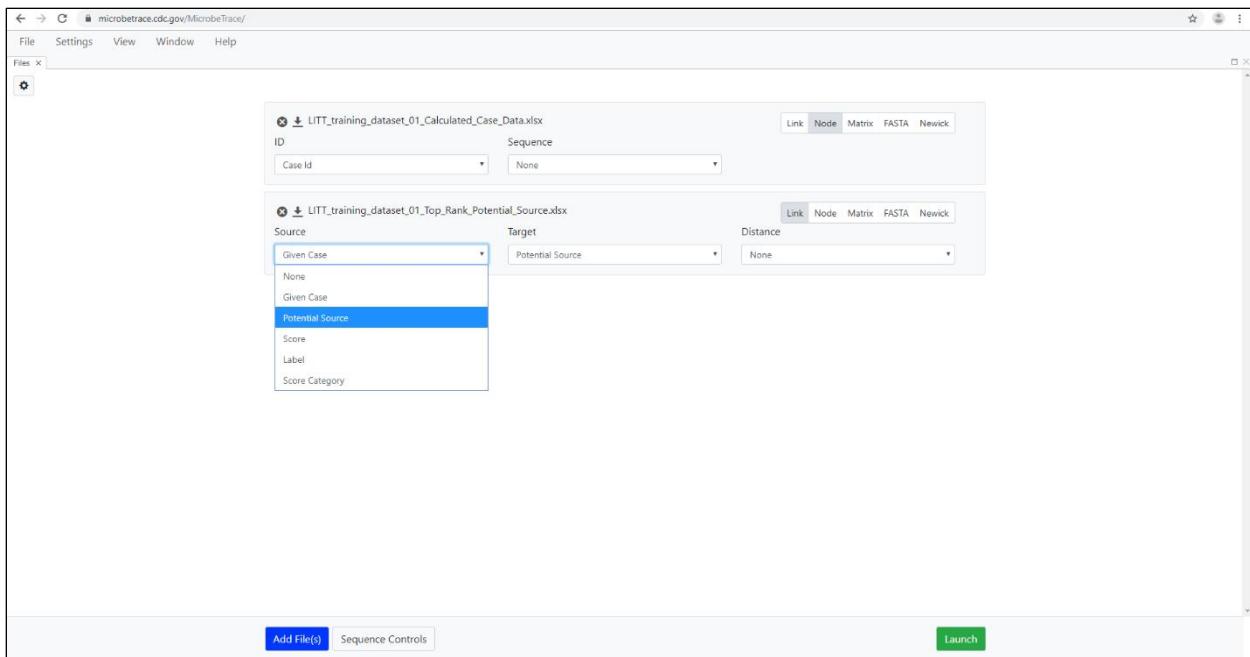


**Figure 38. Screenshot showing the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) and [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#) LITT output files being dragged into the MicrobeTrace data submission webpage.**

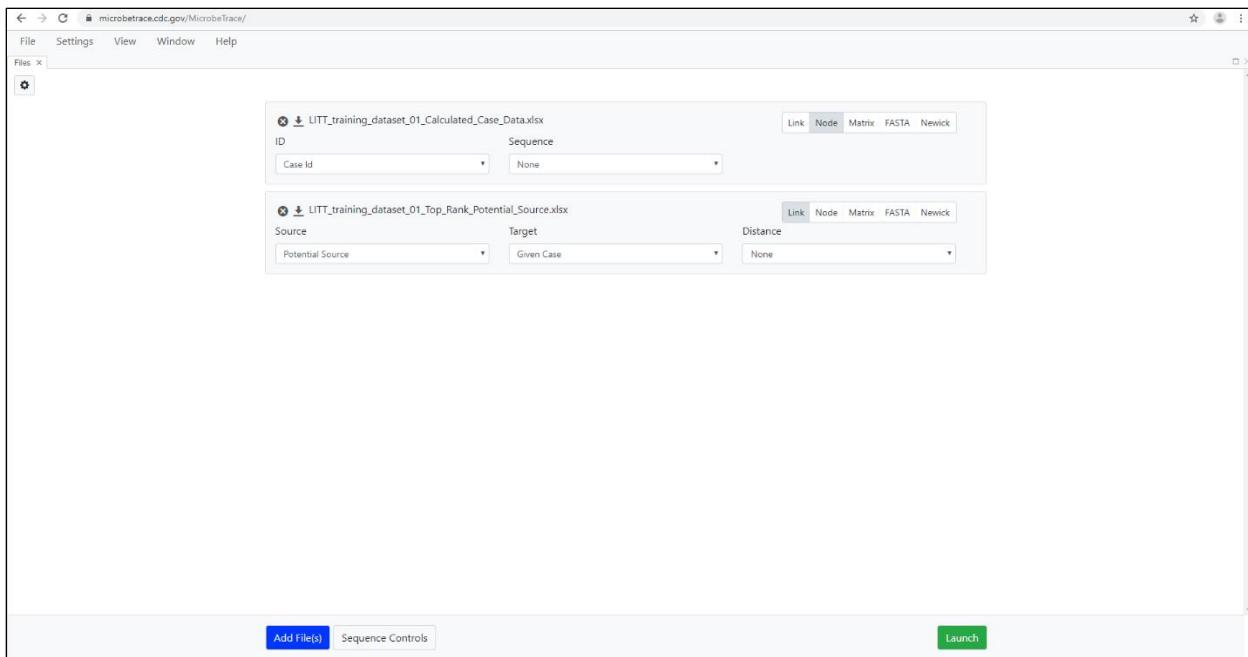


**Figure 39. Screenshot showing windows for the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) and [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#) LITT output files after they were successfully loaded into the MicrobeTrace data submission webpage.**

3. Specify the correct settings for each file. For the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) file, select “Node” from the choices on the right, then use the drop-down menus to set [ID](#) to “Case Id” and set [Sequence](#) to “None” (*Figure 40*). For the [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#) file, select “Link” from the choices on the right, then use the drop-down menus to set [Source](#) to “Potential Source,” [Target](#) to “Given Case,” and [Distance](#) to “None” (*Figure 41*).

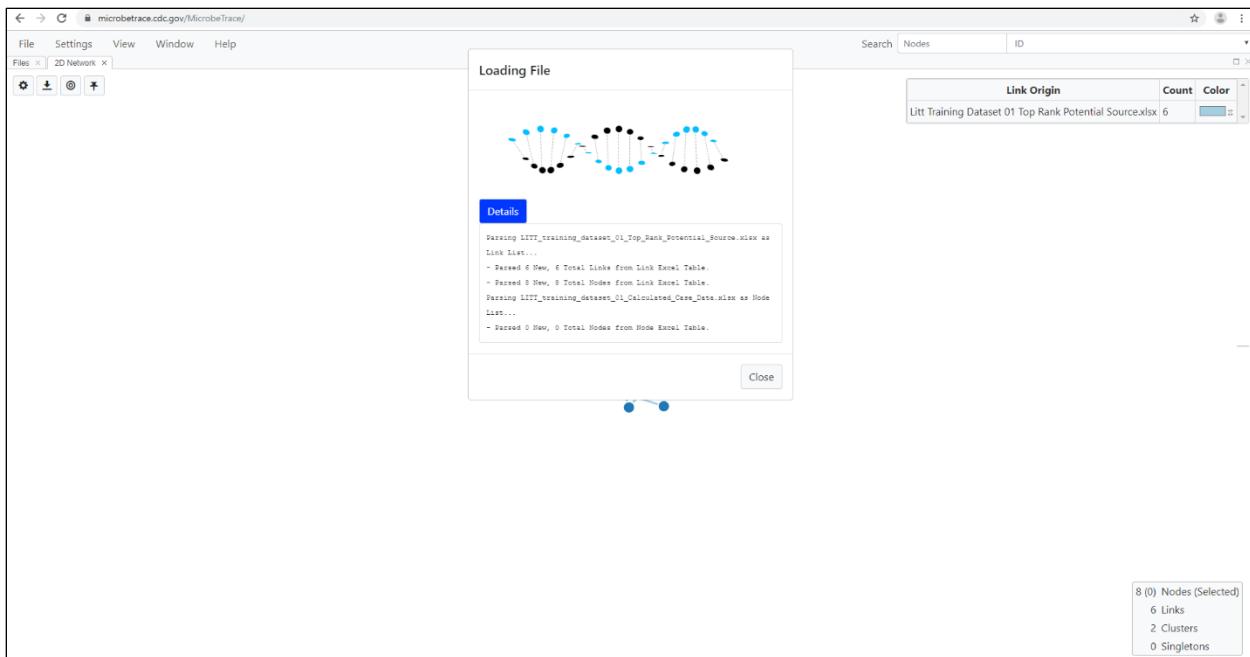


**Figure 40. Screenshot showing correct file settings being selected for the (Prefix)LITT\_Calculated\_Case\_Data and (Prefix)LITT\_Top\_Rank\_Potential\_Source LITT output files using drop-down menus in the MicrobeTrace data submission webpage.**



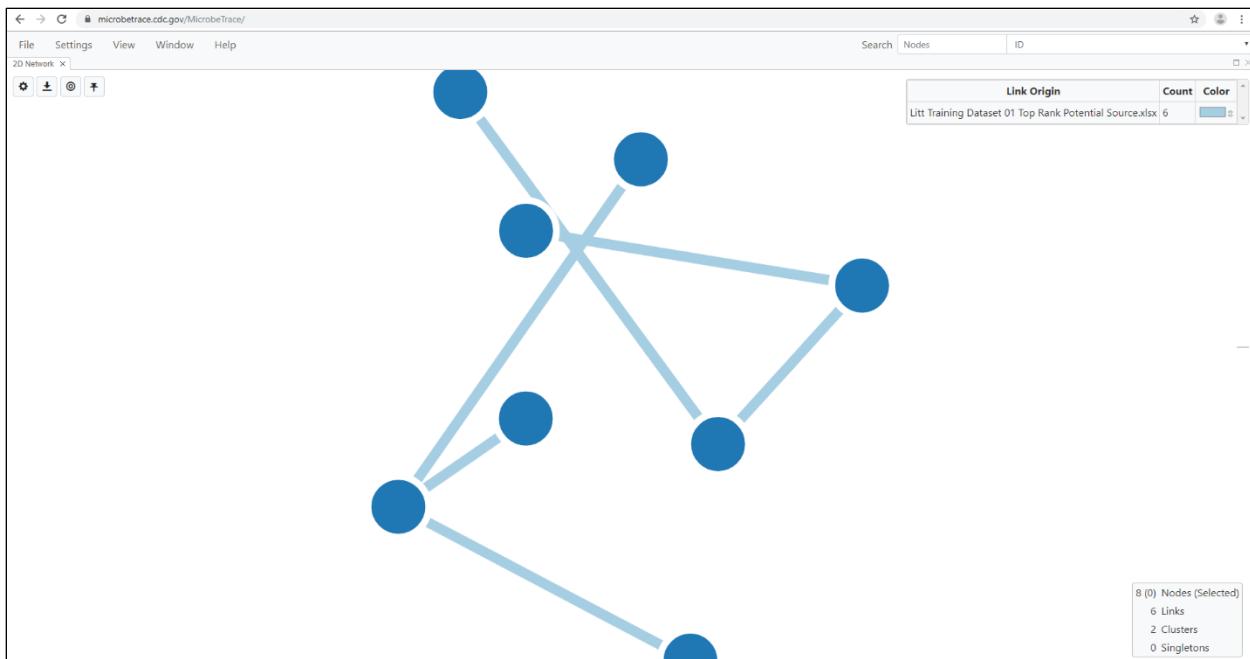
**Figure 41.** Screenshot showing correct file settings selected for the [\(Prefix\)LITT\\_Calculated\\_Case\\_Data](#) and [\(Prefix\)LITT\\_Top\\_Rank\\_Potential\\_Source](#) LITT output files in the MicrobeTrace data submission webpage.

4. Click on the green [Launch](#) button at the bottom right hand corner of the screen to initiate the analysis and generate a 2D network. An animated progress window will briefly appear (*Figure 42*).



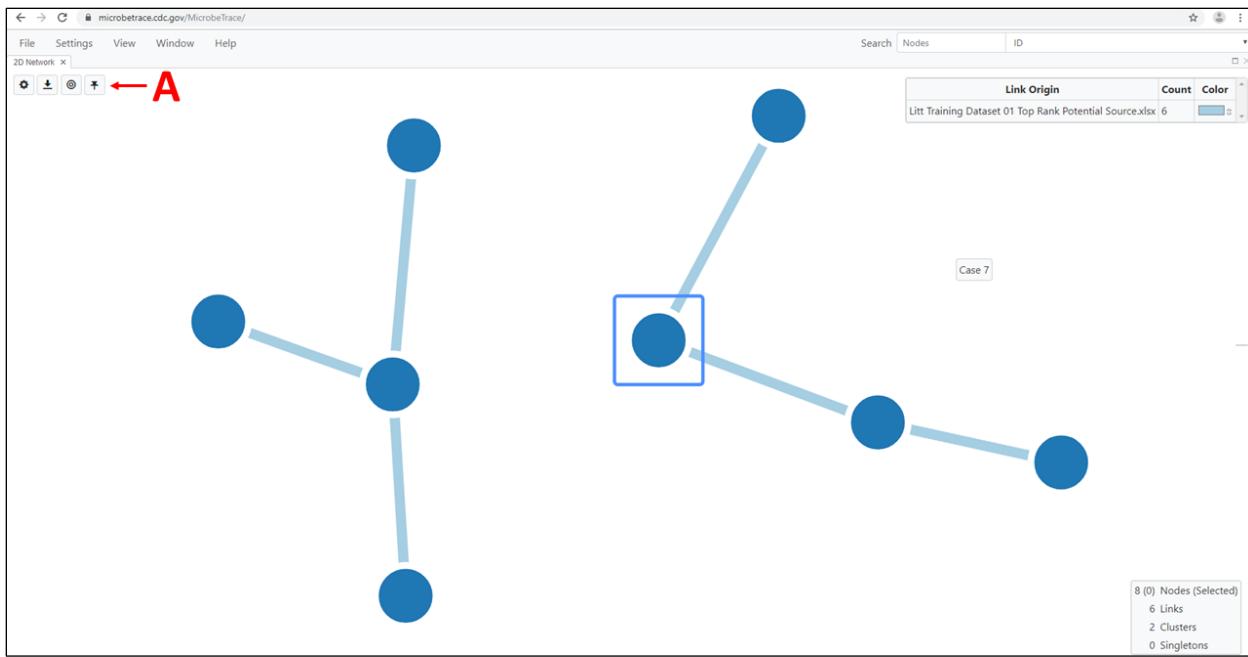
**Figure 42.** Screenshot showing animated progress windows that appears following initiation of analysis to generate a 2D network in the MicrobeTrace data submission webpage.

5. When the analysis is complete, the generated 2D network will be shown (*Figure 43*).



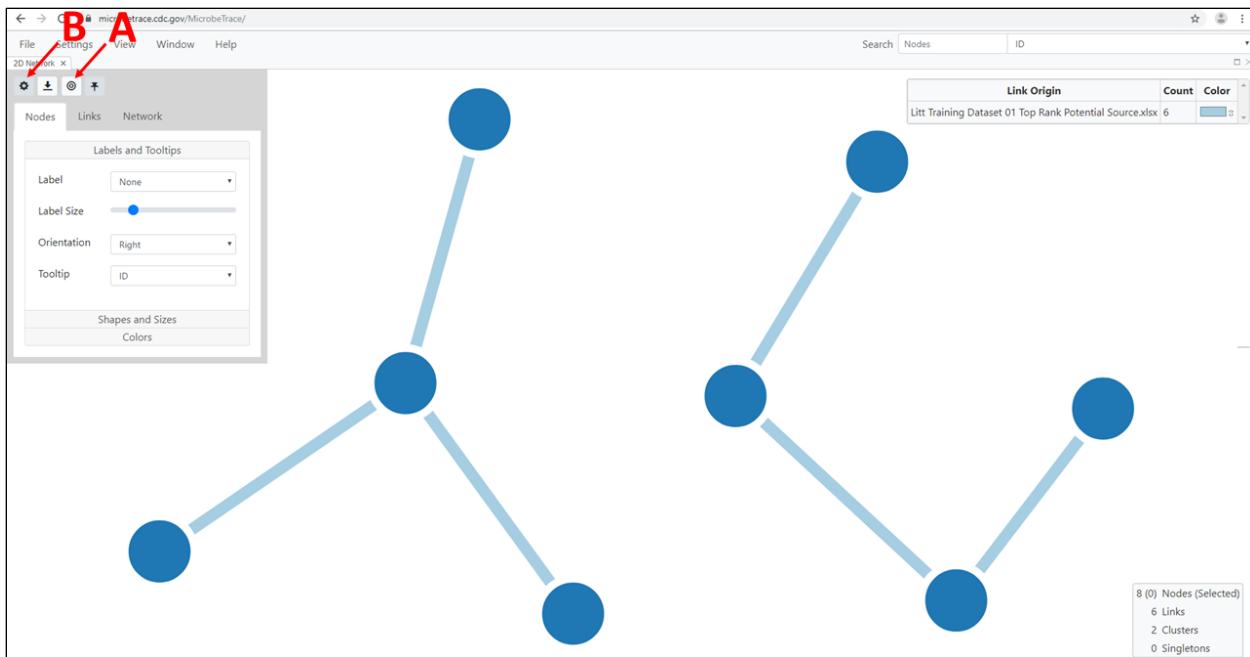
**Figure 43. Screenshot showing a 2D network in the MicrobeTrace data submission webpage.**

6. To change the visual appearance and/or orientation of the network, use the mouse to select (grab) any node and drag it around in the window to relocate it. When this is done, other nodes and links will automatically move to ensure visual buffering. Click on the Pin button in the upper left hand corner of the screen (A in *Figure 44*) to pin the network, deactivating this visual buffering and allowing the locations of nodes and length of links within the window to be manually set.



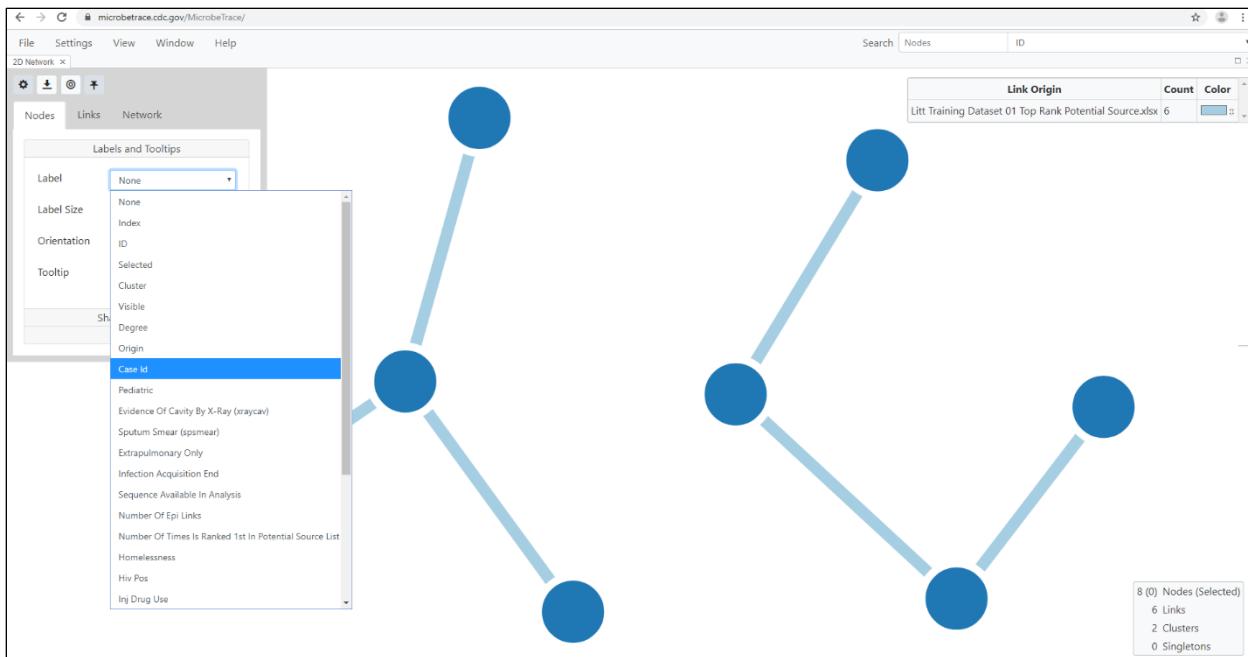
**Figure 44.** Annotated screenshot showing a selected node of a 2D network in the MicrobeTrace data submission webpage and the location of the pin icon (A). Annotation appears in red.

7. Once the appearance of the network is satisfactory, the window can be visually re-centered on the network by pressing the [Target](#) button in the upper left hand corner of the screen (A in [Figure 45](#)). The network can then be annotated. Clicking on the [Gear](#) button in the upper left hand corner of the screen (B in [Figure 45](#)) will open a labelling drop-down menu with three tabs: [Nodes](#), [Links](#), and [Network](#). A few of the many network customization and annotation options are described below.

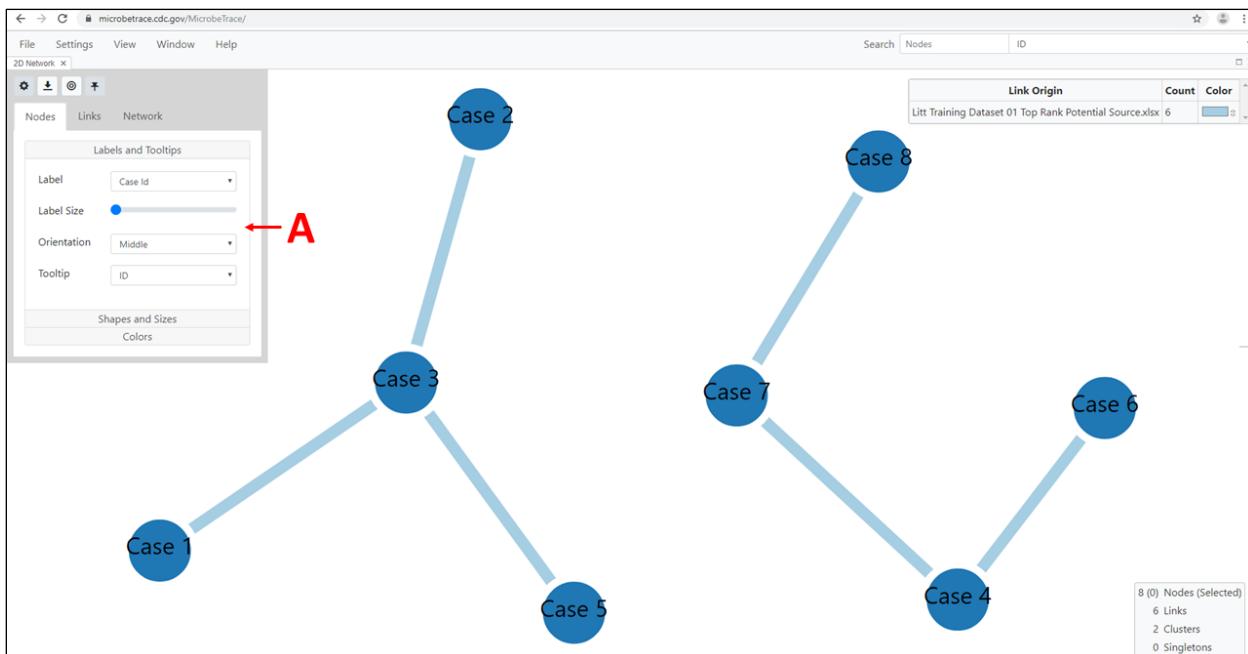


**Figure 45.** Annotated screenshot showing a window visually centered on a pinned 2D network in the MicrobeTrace data submission webpage and the locations of the target (A) and gear (B) icons. The labelling drop-down menu opened by clicking on the gear icon is also shown. Annotation appears in red.

8. Nodes of the network can be customized. For example, to label the nodes with their corresponding values of Case ID used in the LITT analysis, click on the [Nodes](#) tab and then the [Labels and Tooltips](#) subtab, then select “Case Id” from the [Label](#) drop-down menu (*Figure 46*). The label size and orientation can also be customized using the [Label Size](#) slider and [Orientation](#) drop-down menu (A in *Figure 47*).

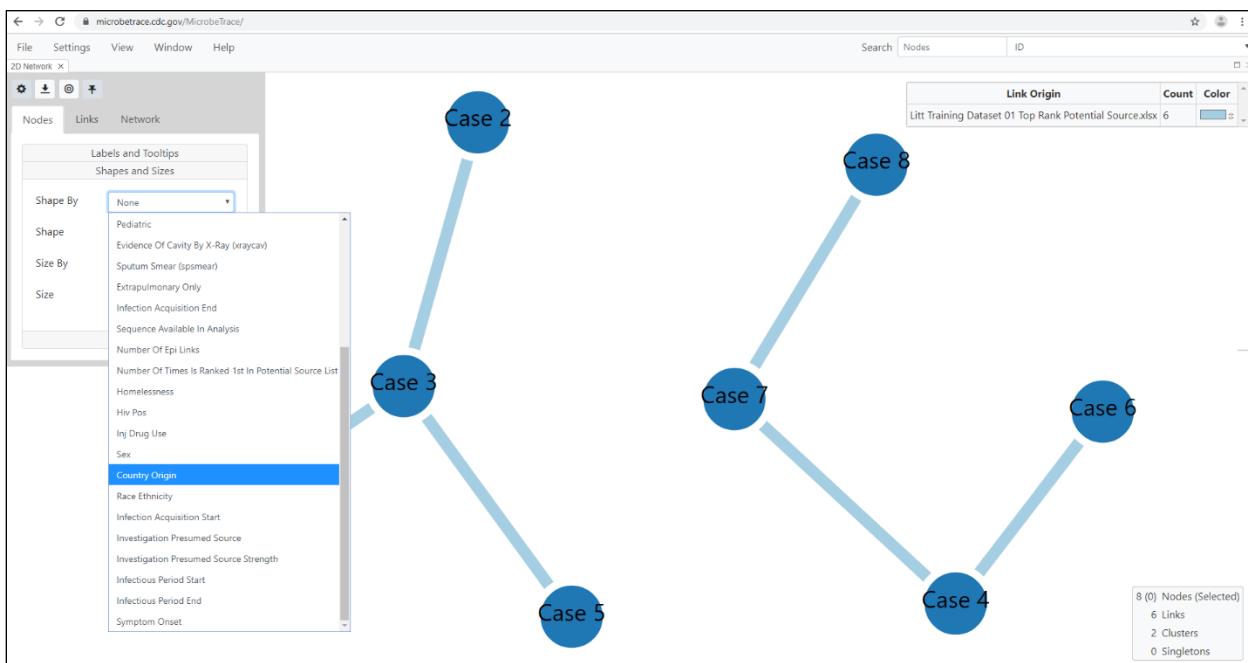


**Figure 46.** Screenshot showing labels being selected for nodes using a drop-down menu in the MicrobeTrace data submission webpage.



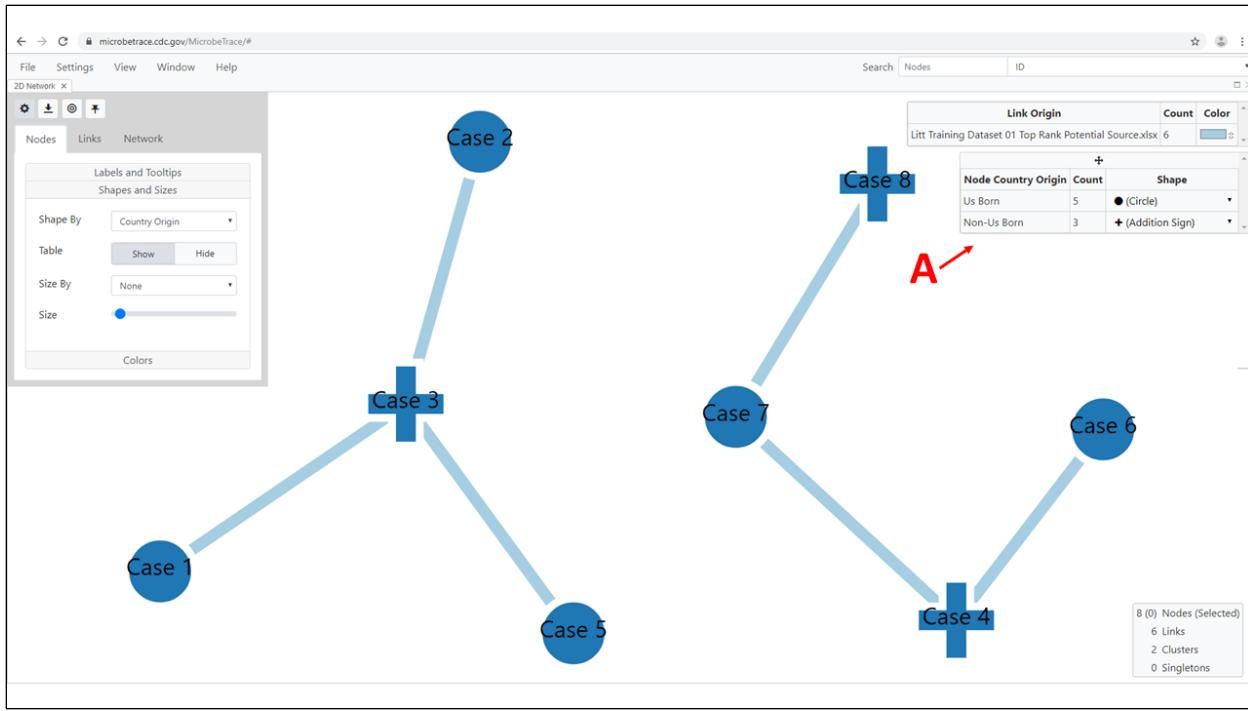
**Figure 47.** Annotated screenshot showing options to customize the size and orientation of node labels (A) in the MicrobeTrace data submission webpage. Annotation appears in red.

9. Nodes of the network can be further customized in various ways. For example, to make the shape of the nodes correspond to values of a risk factor used in the LITT analysis, click on the **Nodes** tab and then the **Shapes and Sizes** subtab, then select the risk factor of interest (e.g., “Country origin”) from the **Shape By** drop-down menu (*Figure 48*).



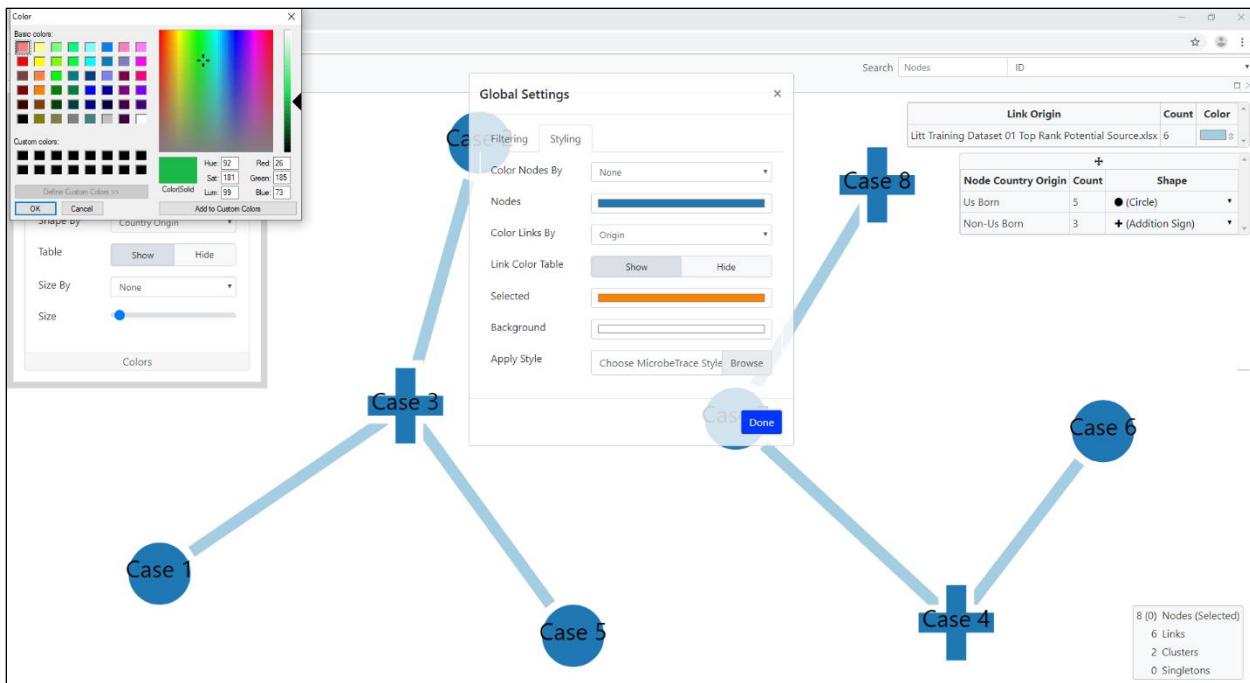
**Figure 48.** Screenshot showing node shapes being customized using a drop-down menu in the MicrobeTrace data submission webpage.

10. To customize the specific shapes of the nodes, click on the shapes in the menu in the top right corner of the screen (A in *Figure 49*).



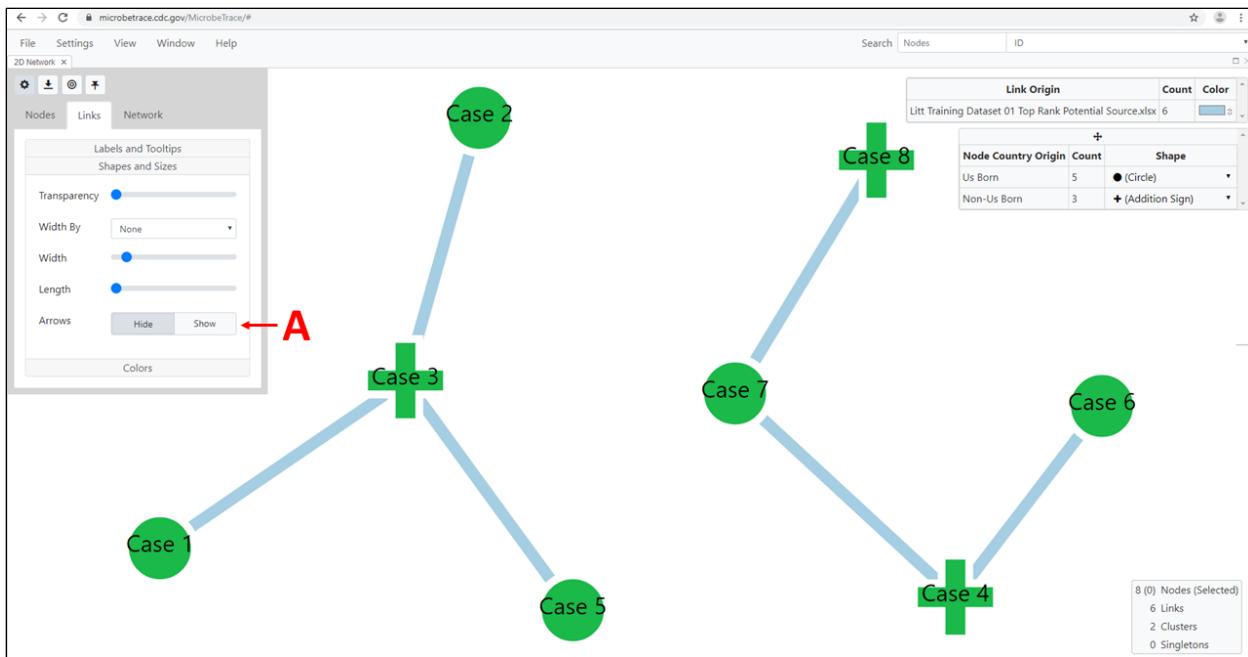
**Figure 49.** Screenshot showing node shapes being customized using a menu in the top right corner of the MicrobeTrace data submission webpage.

11. To customize the color of the nodes, click on the [Nodes](#) tab and then the [Colors](#) subtab, then in the [Global Settings](#) popup window click on the [Nodes](#) drop-down menu and pick a color from the pop up window ([Figure 50](#)).



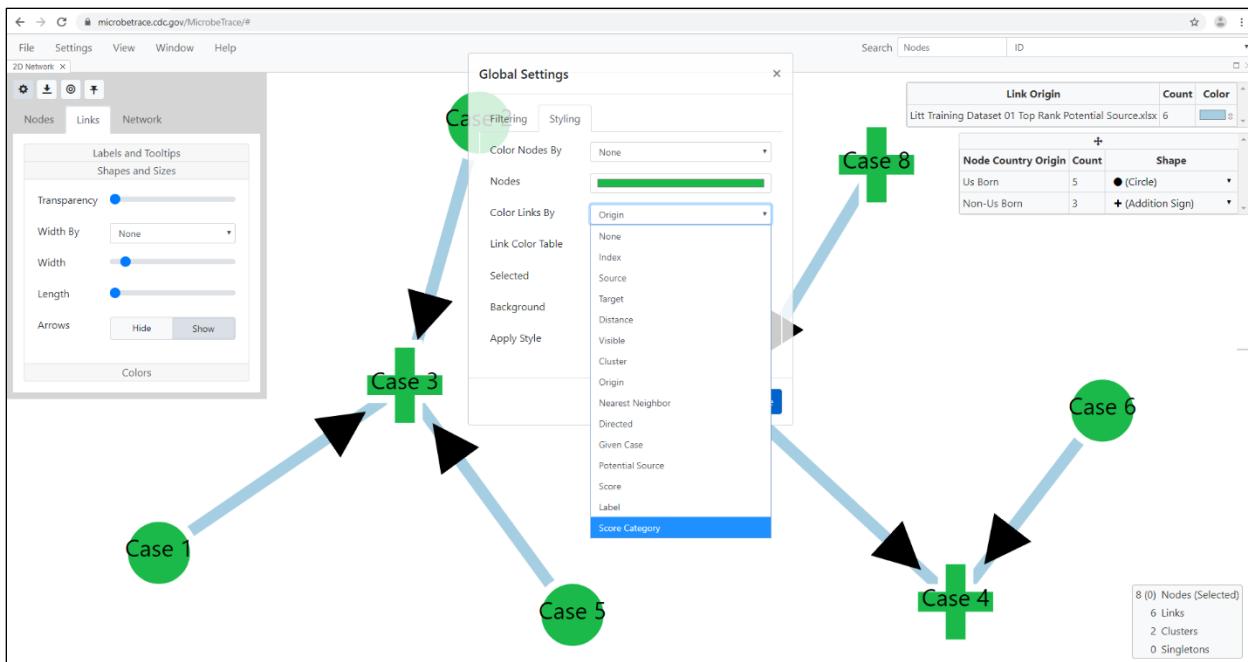
**Figure 50.** Screenshot showing node colors being customized using a popup menu in the MicrobeTrace data submission webpage.

12. Links of the network can be customized. For example, to label the links with arrowheads corresponding to the directionality of possible and probable transmission pathways within the focal case cluster identified in the LITT analysis, click on the [Links](#) tab and then the [Shapes and Sizes](#) subtab, then click on the [Show](#) button associated with [Arrowheads](#) (A in [Figure 51](#)).



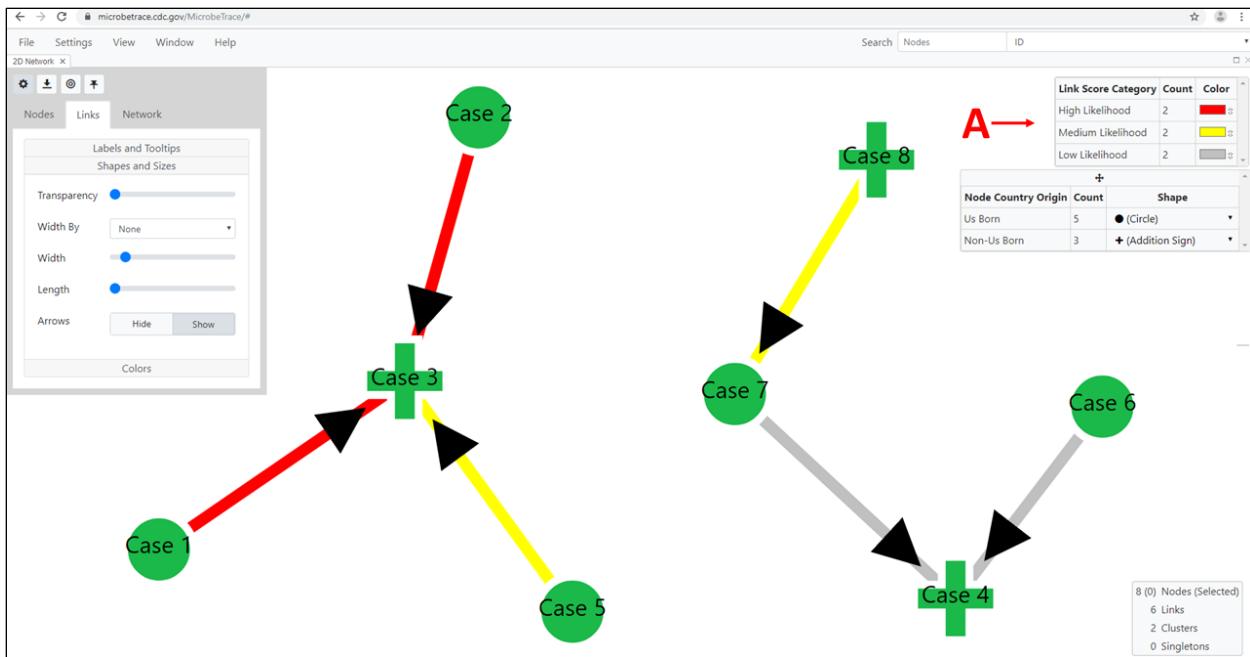
**Figure 51.** Screenshot showing links being customized to show arrowheads in the MicrobeTrace data submission webpage.

13. Links of the network can be further customized in various ways. For example, to color the links to correspond with the likelihood of transmission within the focal case cluster identified in the LITT analysis, click on the [Links](#) tab and then the [Colors](#) subtab, then in the [Global Settings](#) popup window, click on the [Styling](#) tab and then select [Score Category](#) from the [Color Links By](#) dropdown menu (*Figure 52*).



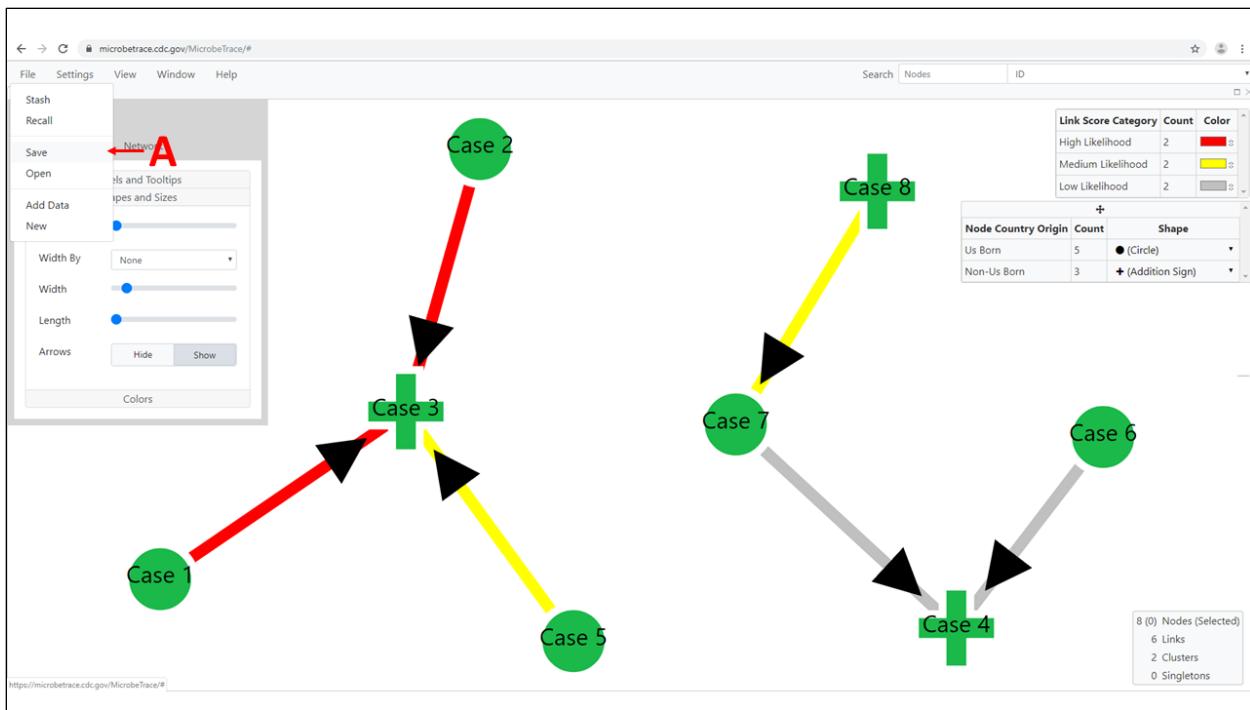
**Figure 52.** Screenshot showing link colors being customized using a dropdown menu in the MicrobeTrace data submission webpage.

14. The color of links associated with each value of **Score Category** can then be changed using the window in the top right hand corner of the screen (*Figure 53*).



**Figure 53.** Screenshot showing link colors being customized using a menu in the top right corner of the MicrobeTrace data submission webpage.

15. Once done customizing the appearance of the 2D network, the associated MicrobeTrace project file can be saved by clicking [Save](#) from the dropdown menu opened when [File](#) is selected in the screen header (A in *Figure 54*).



**Figure 54.** Screenshot showing the 2D network created in the MicrobeTrace data submission webpage being saved.