



# 用户手册

版本 0.6.2

2021 年 5 月



**U.S. Department of  
Health and Human Services**  
Centers for Disease  
Control and Prevention

## 软件免责声明

本软件中包含的材料按“原样”提供给您, 不提供任何明示、暗示或其他形式的保证, 包括但不限于任何关于适用于特定目的的保证。在任何情况下, 疾病控制和预防中心(CDC)或美国(U.S.)政府均不对您或任何其他承担任何直接、特殊、偶然、间接或后果性损害或其他任何损害, 包括但不限于利润损失; 使用、储蓄或收入的损失; 或第三方的索赔, 无论 CDC 或美国政府是否已被告知此类损失的可能性, 因拥有、使用或执行本软件而引起或与之相关的任何责任理论。

请在 <https://microbetrace.cdc.gov/MicrobeTrace/> 访问 MicrobeTrace

如果您在出版物中使用 MicrobeTrace, 请引用 <https://github.com/CDCgov/MicrobeTrace/wiki>

# 内容

软件免责声明 II

简介 ..... 1

MicrobeTrace 用户 ..... 2

MicrobeTrace 帮助 ..... 2

专业术语表 ..... 3

    网络术语..... 3

    遗传分析..... 4

系统要求 ..... 6

下载样本数据..... 6

创建并导入文件..... 8

    可能的文件输入组合..... 12

访问 MicrobeTrace 和加载文件..... 13

    主菜单..... 13

    加载文件..... 16

    加载 FASTA 文件..... 17

    加载节点列表和/或边列表 ..... 22

    加载[边列表](#)而非 FASTA 文件..... 24

数据可视化 ..... 25

平铺不同的视图 ..... 26

2D 网络视图..... 28

    网络配置..... 29

    过滤 ..... 29

节点和链接设置..... 32

    节点属性 ..... 33

    链接属性 ..... 36

    网络属性..... 38

3D 网络视图..... 40

直方图视图 ..... 42

表视图 .....	44
表设置 .....	46
聚合视图 .....	46
交叉表视图 .....	49
气泡视图 .....	50
气泡视图设置 .....	51
流程图视图 .....	53
流程图视图设置 .....	54
散点图视图 .....	56
瀑布视图 .....	57
地图视图 .....	60
地图视图设置 .....	61
地球视图 .....	68
变更全球视图选项 .....	70
数据选项卡 .....	70
Gantt 视图 .....	74
Epi 曲线 .....	76
热图视图 .....	77
序列视图 .....	79
系统发育视图 .....	81
系统发育设置 .....	83
树选项卡 .....	84
操作子选项卡 .....	87
分支选项卡 .....	88
叶选项卡 .....	88
节点选项 .....	90
故障排除 .....	93
参考文献 .....	99
致谢 .....	99

## 简介

在本节之后将提供一个**术语表**，以详细了解网络构建和分析中常用的术语。当您浏览手册时，您会发现许多术语或参考文献都以蓝色文本显示；点击这些超链接将带您到词汇表部分中的相关词或网站以获取更多信息。

MicrobeTrace 是一种软件工具，可实现[网络](#)和相关数据的快速可视化。MicrobeTrace 可让用户将他们的数据特征映射到网络元素的可视屏幕特征（例如颜色、大小、形状）。除了网络可视化，MicrobeTrace 还提供其他分析工具（例如，表、过滤器、地理地图、直方图、3D 网络、系统发育树构建、Gantt 图、时间播放器、冲积图和散点图），以探索和背景化核苷酸序列和其他数据。这些方法已在流行病学中广泛采用，尤其是在应对结核病、HIV、HCV 和 COVID 暴发时，但在分子生物学和社会学等方面都有广泛的应用。

对于核苷酸序列，在使用 TN93 (Tamura-Nei, 1993) 核苷酸替换模型计算遗传距离后构建遗传网络，该模型根据每个位点序列之间的核苷酸差异计算两个序列之间的距离。使用经验确定的遗传距离截止值来识别各个序列之间的潜在联系。对于 HIV 序列，TN93 是使用的核苷酸替换模型，1.5% 核苷酸替换/位点的遗传距离是检查数据集中遗传关系的良好初始截止值，尽管较小的距离（例如 0.5%）可能会提高近期传播的特异性。对于其他病原体，MicrobeTrace 允许导入使用其他核苷酸替代模型确定的距离矩阵或具有单核苷酸多态性 (SNP) 数据的病原体的汉明距离。您还可以导入 Newick 系统发育树文件，MicrobeTrace 将使用从树分支长度计算的遗传距离生成一个网络，使用[共祖距离](#)算法。

MicrobeTrace 还可以使用联系人追踪或合作伙伴服务数据生成社交网络图表。所有网络均可以根据可用的补充数据源（人口统计、临床、流行病学等）进行定制，并且可以使用包含的最小跨度方法确定最可能的传播途径等数学推论。

MicrobeTrace 是一种高响应性的可视化序列分析工具，它可以减少数据收集和分析之间的差距，并帮助您发现、理解和交流个体（表示为网络中的[节点](#)）之间的关系（以线或[边](#)表示）。虽然它使用网络浏览器的功能，但 MicrobeTrace 可以在您的笔记本电脑上的某个位置运行，而非在以网络为基础的服务器上，并且可以部署在没有互联网访问的位置，从而降低启动成本和分析时间和精力，同时保持数据安全。使用敏感数据时，数据安全至关重要，在使用 MicrobeTrace 时应给予最高考虑。**使用 MicrobeTrace 时请遵守您所在机构的数据安全政策。**

本用户手册是关于如何使用软件的分步指南，从加载数据开始，使用每个步骤的文本描述以及相应的屏幕截图生成许多可视化。我们从网络分析中常用的术语表以及与 MicrobeTrace 一起使用的各种文件的描述开始。本手册可作为独立指南提供。但是，我们很乐意根据需求和我们的日程安排安排培训或帮助用户浏览 MicrobeTrace。我们在[支持和问题](#)部分提供技术支持详细信息。

## MicrobeTrace 用户

MicrobeTrace 是一种多功能、功能强大的开源数据可视化工具,可能对来自广泛学科的研究人员很有价值。调查活跃的微生物传播集群的州和地方公共卫生工作者以及进行传播网络分析的研究人员(学术界和政府)会发现 MicrobeTrace 特别有用。尽管该软件最初是为 HIV 和接触者追踪传播分析而设计的,但 MicrobeTrace 与病原体无关,可用于任何病原体或疾病数据。

## MicrobeTrace 帮助

可通过选择菜单选项右侧的“帮助”链接获得 MicrobeTrace 帮助(如图 1 所示)。帮助包含非常有用的信息和分步过程,可帮助您使用该工具。

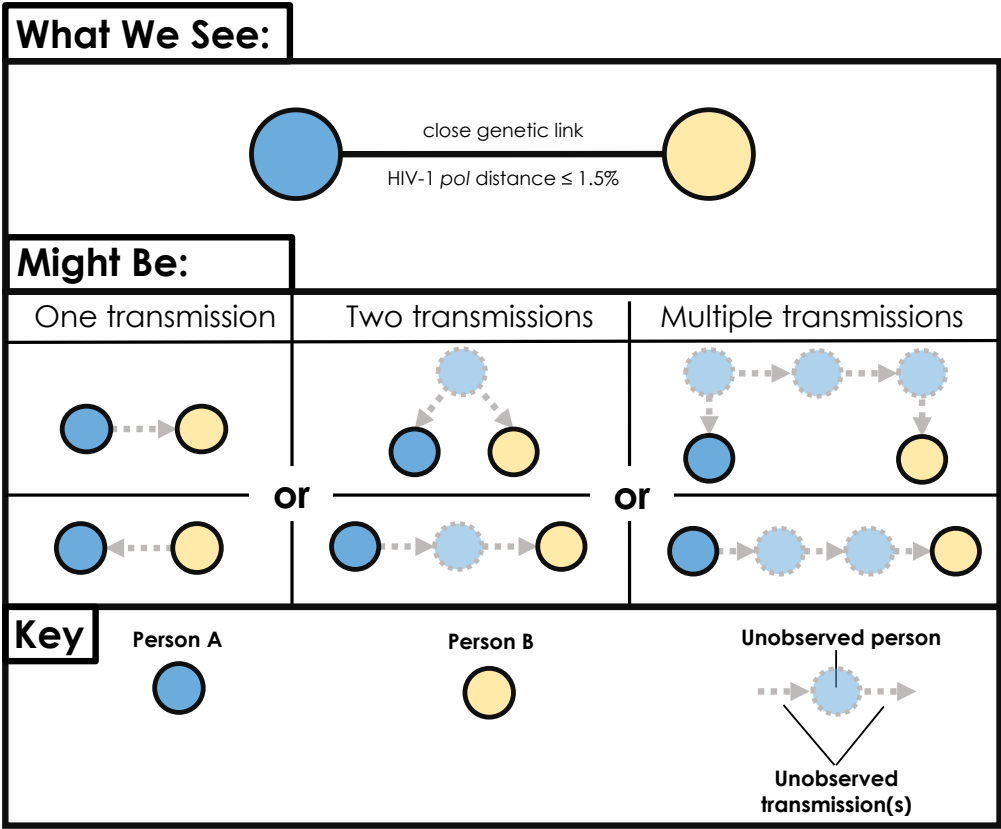
- ✦ **网站:** <https://microbetrace.cdc.gov/MicrobeTrace/> — Follow this link to launch MicrobeTrace
- ✦ **GitHub:** <http://github.com/cdcgov/microbetrace> — GitHub 是托管 MicrobeTrace 的软件存储库。您可以找到软件及其功能的简要说明,以及各种组件的代码。这也是一个报告您遇到的任何问题或错误的平台。
- ✦ **支持和问题:** [microbetrace@cdc.gov](mailto:microbetrace@cdc.gov) 向我们发送有关 MicrobeTrace 的任何问题。我们的团队成员会很乐意回应。

# 专业术语表

## 网络术语

**边** — 网络中连接两个节点的链接或线路称为边。一条边由两个连接节点的独特 ID 组成, 通常标记为“源”和“目标”。边可以是数据中两个序列之间的密切遗传相关性, 但不能推断这两个序列之间传输的方向性。

图 1. 节点(人)之间可能的传输关系



HIV-1 聚合酶(pol)序列(距离  $\leq 1.5\%$ )之间的密切遗传联系可以代表许多可能涉及至少一个未观察到的人的实际传播情形。上面显示了六种潜在的传输情形。

**边属性** — 与边关联的数据字段(即边的特征), 可以是分类值或数值。例如, 可以计算由一条边连接的个体之间年龄的绝对差异。上方图 1 概述了节点(人) A 和节点(人) B 之间的各种可能的传播情形, 这些情形在不包含额外流行病学信息的情况下难以确定方向性。

**边列表** — 所有边和相关信息(例如, 遗传距离和/或接触类型数据)仅出现一次的列表。边列表也称为链接列表。对于 MicrobeTrace, 这些数据包含在 CSV (逗号分隔值) 文件或 Microsoft Excel 文件中。可以通过将元数据存储在 Excel 文件中, 以准备 CSV 文件, 然后将其另存为 CSV 文件。请注意, 倒数边(个人 A  $\rightarrow$  个人 B 和个人 B  $\rightarrow$  个人 A)被认为是独特的。下方是边列表的示例。

来源	目标	遗传距离	接触方式
个人 A	个人 B	0.004	性方式
个人 B	个人 A	0.004	社交

**元数据** — 提供有关其他数据的信息的数据。边和/或节点都可以存在元数据：例如，新病例的记录输入日期或与链接关联的高风险联系人的类型。对于 MicrobeTrace，此数据包含在 CSV (逗号分隔值) 文件中。可以通过将元数据存储在 Excel 文件中，以准备 CSV 文件，然后将其另存为 CSV 文件。

**节点** — 网络中的离散对象，通常代表来自感染者的人 (如在接触者追踪网络中) 或病毒核苷酸序列 (如在遗传距离网络中)。

**节点属性** — 与节点关联的数据字段 (即，节点的特征，如人的年龄)，可以是分类值或数值。

**节点列表** — 每个节点及其相关信息 (例如人口统计和行为细节) 仅出现一次的列表。对于 MicrobeTrace，此数据包含在 CSV 文件中。可以通过将元数据存储在 Excel 文件中，以准备 CSV 文件，然后将其另存为 CSV 文件。下面是一个节点列表的例子。

ID	性别	年龄 (岁)	种族/民族
个人 A	男性	26	白种人
个人 B	女性	23	黑人

**网络** — 就本软件和手册而言，有社交或联系网络和遗传距离网络。社交和接触网络是根据调查期间合作伙伴服务收集的行为数据确定的。遗传网络是从正在研究的病原体的微生物核苷酸序列中推断出来的。这两种网络类型都应包括在分析中，从而可以对传播网络进行最佳流行病学理解，其中传播网络是一个结合了来自社会/接触者追踪和遗传网络的数据的网络。

**源** — 边开始的节点。例如，如果 **个人 A** 命名 **个人 B** (**个人 A** → **个人 B**)，那么来源就是 **个人 A**。请注意，在此上下文中，源并不意味着传输的来源。

**目标** — 边结束的节点。例如，如果 **个人 A** 命名 **个人 B**，则目标是 **个人 B**。请注意，在此上下文中，目标并不意味着传输的目标。

## 遗传分析

**集群** — 集群被定义为一组节点，其中每个节点都可以直接或间接地从任何其他节点到达。可以使用分子序列和/或接触者追踪数据来识别集群。如果两个节点之间不存在路径或边，则认为它们在不同的集群中，或者它们是单例。**需要注意的是，集群的识别将根据您选择的遗传距离阈值或接触者追踪数据中提供的信息的不同而改变。**

**Dyad** — 仅包含两个节点的集群。

**FASTA 文件** — 一种以文本为基础的文件格式，用于表示由标准 [IUPAC 单字母字符组成的核苷酸或氨基酸序列的文件格式](#)。FASTA 文件的文件扩展名可以是 .FASTA、.FA、.FAS，甚至可以保存为文本文件 (.TXT)。FASTA 文件扩展名无需大写。FASTA 文件中的第一行以“>” (无撇号的大于号) 开头，包括用于序列、标本或人员名称的代码或文本。FASTA 文件的下一行包含使用单字母 IUPAC 代码的实际核苷酸序列。**如果您要上传序列文件以及包含人口统计数据 (CSV 文件) 的相应节点列表，则 FASTA 文件中用于序列的 ID 必须与 CSV (或 Excel) 文件中的 ID 完全匹配，并且还必须是独特的。**多序列 FASTA 文件将

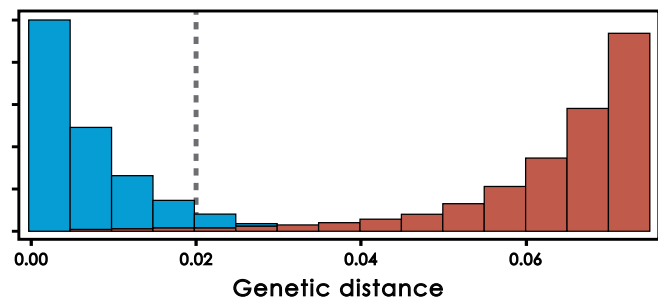


包含独特序列名称及其对应序列的多次迭代。空行不必分隔多序列 FASTA 文件中的第一个和后续序列。下面是一个包含三个不同短序列的多序列 FASTA 文件的内容示例。

**遗传距离阈值** — 在分析核苷酸序列时, 所有可能的序列对之间的遗传距离使用核苷酸替代模型 (MicrobeTrace 使用 TN93 模型) 确定。因此, 必须选择分析的遗传距离阈值或截止值以确定潜在的传播联系。对于 HIV 传播, 对应于 0.015 个核苷酸替换/每个位点的 1.5% 遗传距离阈值被用作连接密切相关病毒的起点。作为比较, 两个 HIV pol 序列之间 1.0% 的距离代表单个感染 HIV-1 亚型 B 的个体中大约 10 年的病毒进化。但是, 用户需要根据情况 (例如, 最近与遥远的进化历史) 和正在研究的特定病原体选择适用的阈值。您分析的潜在截止值可以通过确定一个阈值来确定, 该阈值最能区分通常存在于您的序列中的遗传距离的双峰分布。

下面的图 2 是遗传距离的频率分布示例, 蓝色框表示与已知病原体传播病例的遗传距离, 红色框表示与无病原体传播证据的病例的遗传距离。对于该数据集, 0.02 个核苷酸替换/位点的遗传距离可能最能区分与传播相关和未传播的病毒/病原体的遗传距离 (参见图中的虚线)。较低的阈值将导致更少的已识别联系, 但会增加对近期传播的特异性。

图 2. 遗传距离直方图



```
> 序列 ID 1
ATCGATCGATCGATCGATCG
> 序列 ID 2
ATCGATCGATCGGGGGGGGG
> 序列 ID 3
ATCGATCGATCGCCCGGGTT
```

**共祖距离算法** — 链接树中两个节点的分支长度的总和。

**Singleton** — 一个孤立的节点。例如, 一个不链接到网络中的任何其他节点的节点。

**SNP** — 单核苷酸多态性。这是发生在基因组特定位置的两个序列之间的单核苷酸差异, 通常称为基因突变。SNP 更常用于细菌病原体。

**网络可视化参数** — 网络可视化使用物理属性(包括电荷、摩擦力和重力)进行优化, 以确定网络中节点的密集程度。

**电荷** — 在网络可视化中节点相互排斥以保持分离, 因此所有节点都是可见的。要更改节点相互排斥的程度, 可以根据需要修改此设置。

**摩擦** — 节点可在计算机屏幕上的网络视图中移动的速率。高摩擦意味着节点不会过度移动。

**重力** — 节点被引致计算机屏幕网络视图的中心, 与重力常数成比例。低重力意味着节点将向边浮动, 而高重力将确保它们在屏幕上紧密聚集。

**工具提示** — 当鼠标指针悬停在对象上时显示信息的软件可视化工具。

## 系统要求

MicrobeTrace 能够在任何使用 Google Chrome、Firefox 或 Microsoft Edge 的计算机上运行。

**\*请注意\* MicrobeTrace 与任何版本的 Microsoft Internet Explorer 均不能兼容。**

## 下载样本数据

[点击这里](#)下载示例文件, 包括示例节点列表、边列表、FASTA 文件和可导入 MicrobeTrace 的 SecureHIVTrace 文件。SecureHIVTrace 是一个私人生物信息学工具, 供美国注册的公共卫生管辖区使用 pol 序列执行 HIV 集群检测。下面列出了示例文件的名称。

**节点列表:**Demo\_outbreak\_NodeList.csv

**边列表:**Demo\_outbreak\_EdgeList.csv

**序列文件(FASTA):**Demo\_outbreak\_Sequences.fas

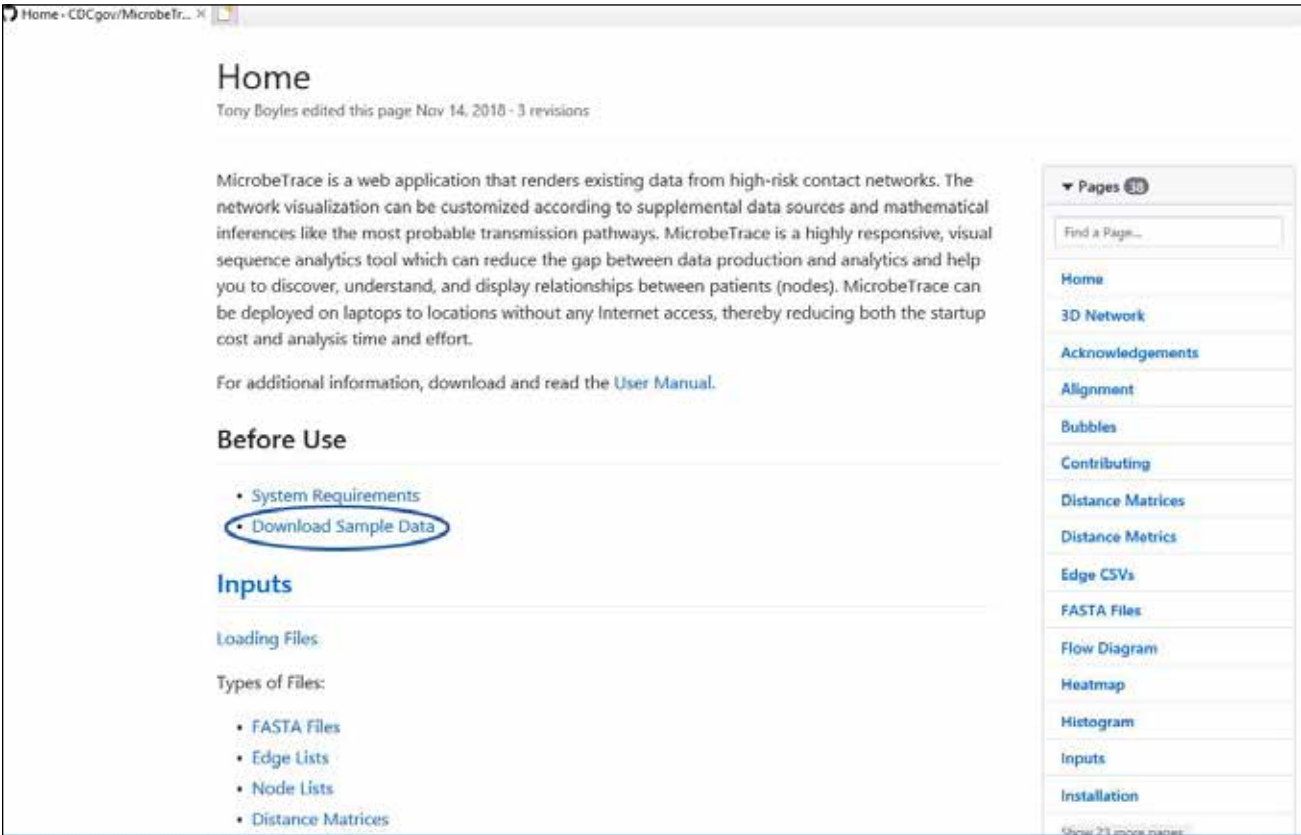
**距离矩阵:**Demo\_outbreak\_DistanceMatrix.csv

**Newick(树)文件:**Demo\_outbreak\_PhylogeneticTree

有关每种文件类型的详细信息, 请参阅下方的说明。

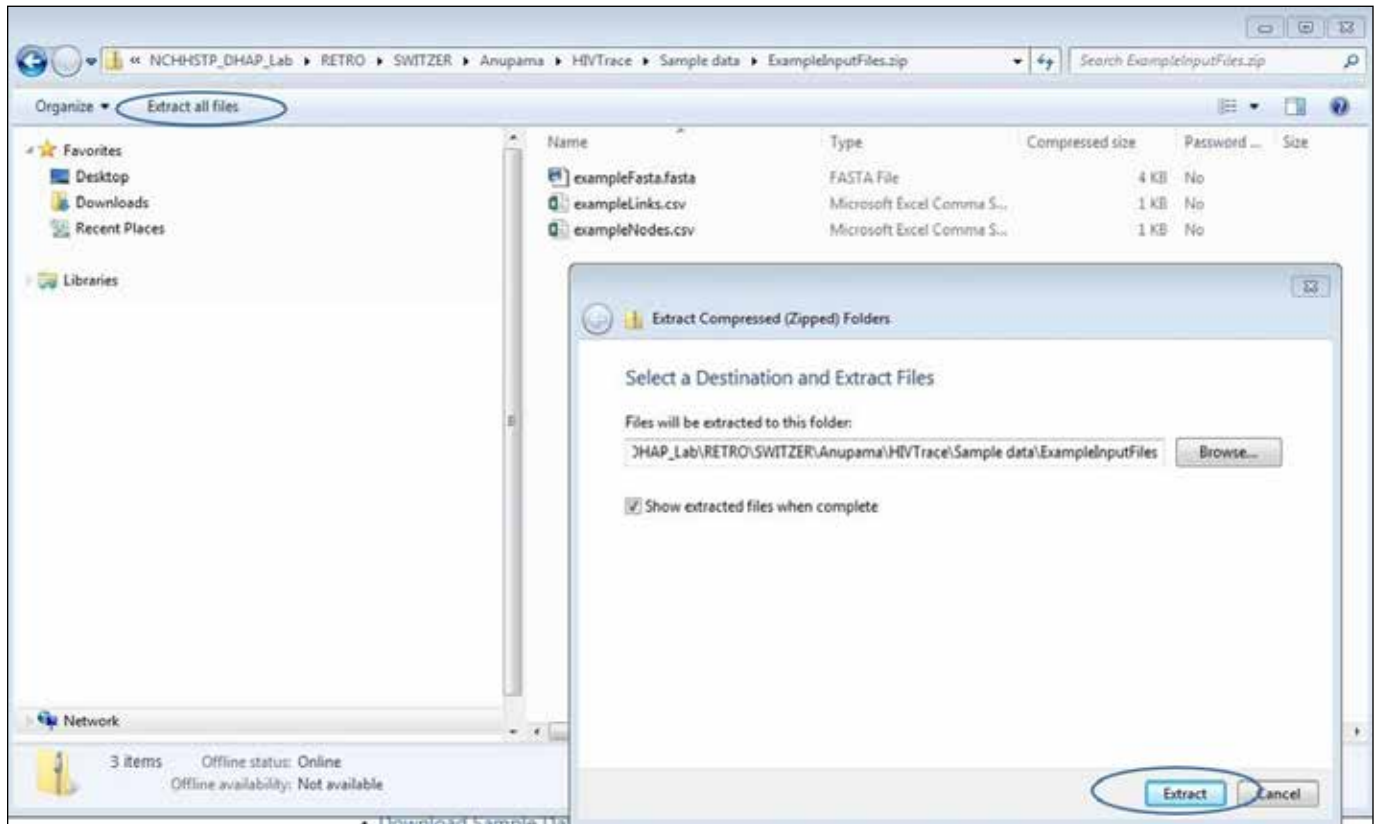
当您点击 MicrobeTrace 网站上的下载样本数据时, 系统会提示您保存或打开文件。选择**另存为(Save As)**, 并将压缩文件保存到您选择的位置(图 3)

图 3.a. 从 MicrobeTrace 网站下载示例文件



在您的计算机上打开文件位置, 您将看到三个压缩文件。当您点击**提取所有文件(Extract all Files)**时, 会打开一个对话框; 点击**提取(Extract)**。

图 3.b. 从压缩文件夹中提取 MicrobeTrace 示例文件



您可以使用这些解压缩的示例文件来探索 MicrobeTrace。

## 创建并导入文件

MicrobeTrace 接受以下文件格式：

- ✦ 在 [FASTA 文件](#) 格式中的核苷酸序列
- ✦ 标准 Microsoft Excel 文件，或逗号或空格分隔的文件（.CSV 文件）。这些文件可以是 [边列表](#) 或 [节点列表](#)（具有序列/患者 ID 和相应数据的文件，例如年龄、性别、风险类型、传播方法、诊断日期、序列子类型）。一个节点可以代表许多事物，但在合作伙伴服务（例如，接触者追踪）的上下文中，它们通常代表受感染者或其高风险合作伙伴。在遗传距离网络中，节点代表出现在 [FASTA 文件](#) 中的病原体序列。FASTA 文件中的 ID 显示为“>”之后和每个序列中任何空格之前的文本。

虽然 [节点属性](#) 并非可视化网络所必需的，但它们是表征和探索网络的重要组成部分。要将节点属性与网络相关联，节点列表 CSV 或 Excel 文件中的每个节点 ID 必须与其在提供的边列表中的相应 ID 完全匹配；或如果序列作为 FASTA 文件单独加载，则与序列 IDS 完全匹配。所有可能有助于网络分析的可用 [元数据](#) 应作为附加列附于节点列表 CSV 或 Excel 文件中的节点 ID 列之后。

**\*注意：如果节点列表文件包含具有序列的列，MicrobeTrace 将在分析中包含这些序列，并且您将拥有与单独的序列文件相关联的所有功能。或者，序列名称可以存储在 CSV 或 Excel 文件中的任何列中，列标题名为“ID”。如果存在多个“ID”列，将使用文件中最左边的列\***

**\*请注意:**在 CSV 或 Excel 文件中, 具有相同节点名称的行会导致删除重复的行。您必须确保 ID 名称是独一无二的。

**边列表:**作为对 FASTA 文件的替代输入, 可以提供边列表, 以指示节点 CSV 或 Excel 文件中定义的节点之间的连接。这称作“边 CSV”文件, 是在接触者追踪期间确定的典型的人与人之间的联系。在 MicrobeTrace 中, 这也称为链接文件或链接列表或链接 CSV 文件。来自边 CSV 或链接文件的网络被称作联系人追踪或社交网络。这是边文件中的数据示例。

源 ID	目标 ID	边属性
John	Jacob	高危接触
John	Mary	高危接触

可以通过将数据列添加到边列表来可视化其他边属性(或数据)。边属性可以是进一步定义两个节点之间关系的任何特征。重要的是要注意边属性应该引用由边连接的两个节点。边属性的一些示例包括两种病原体之间的遗传距离、两个人之间发生的高风险接触类型或两个人之间的年龄差异。在接触者追踪网络中, 一条边代表两个人之间的流行病学联系。在遗传距离网络中, 一条边代表两种病原体之间的遗传关系。边可以有向的或无向的。有向边由节点之间的箭头表示。**\*请注意:**默认情况下, 箭头是关闭的。我们强烈建议使用此默认设置, 除非使用额外的流行病学信息有很强的信心支持方向性(有关更多信息, 请参阅[方向性](#))。

- ✦ **MicrobeTrace 会话文件和样式文件:**如果您正在处理数据集并需要保存数据和设置以供以后使用, 此功能非常有用。该文件将以文件扩展名 .microbetrace 保存。然后, 您可以将保存的 .microbetrace 文件直接加载到 MicrobeTrace 中。当您保存会话时, 您还可以选择按集群下载数据(图 4)。

图 4. 保存文件:按集群下载节点和边列表

Save

ClusterDownloadTest

session

☒ Compress?


































☒ By Cluster

Cancel

Save

如果您选择此选项, MicrobeTrace 将在您选择的位置创建一个压缩文件夹。此文件夹将依次包含每个集群的子文件夹, 其中将包含该集群的节点列表和边列表(图 5)。

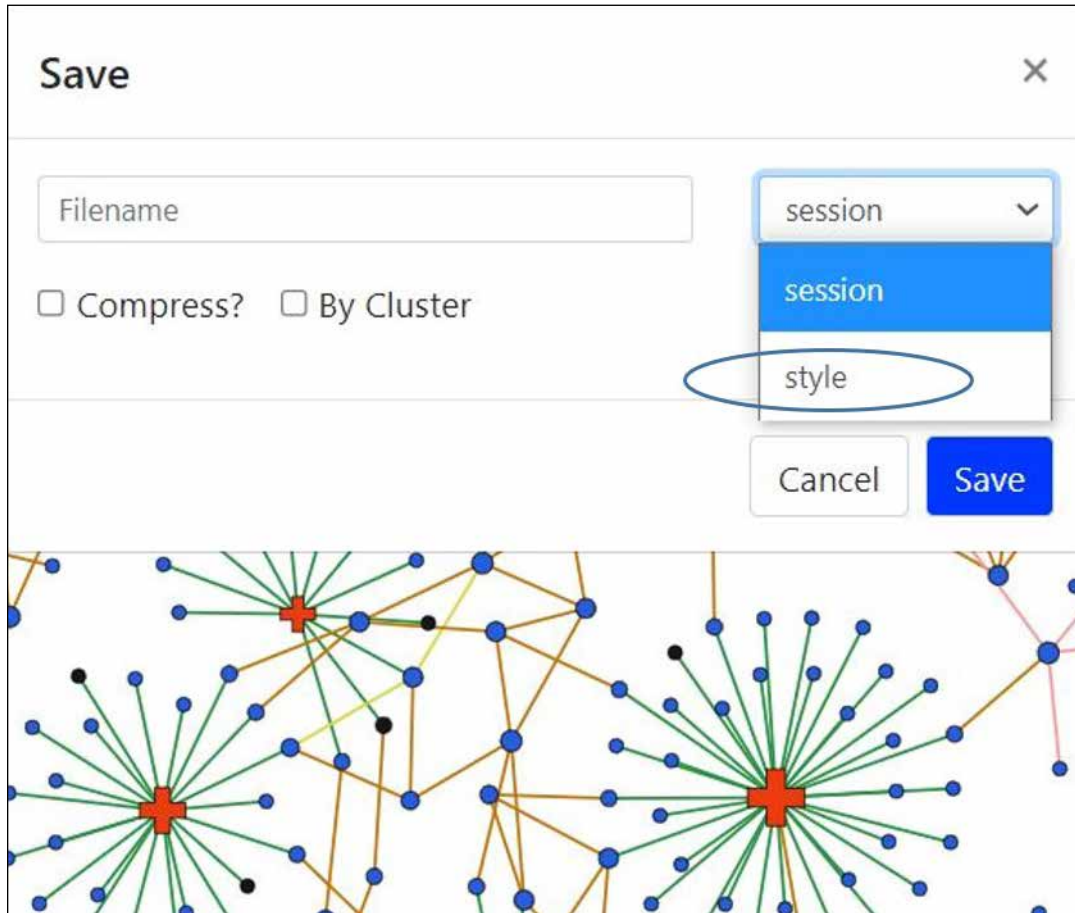
图 5. 集群级数据下载。每个文件夹都包含节点列表和边列表, 其中包含该集群内节点的数据。

<input type="checkbox"/> Name	Status	Date modified	Type
 cluster-0		5/26/2021 5:20 PM	File folder
 cluster-1		5/26/2021 5:20 PM	File folder
 cluster-2		5/26/2021 5:20 PM	File folder
 cluster-3		5/26/2021 5:20 PM	File folder
 cluster-4		5/26/2021 5:20 PM	File folder
<input type="checkbox"/> cluster-6		5/26/2021 5:20 PM	File folder
 cluster-9		5/26/2021 5:20 PM	File folder
 cluster-10		5/26/2021 5:20 PM	File folder
 cluster-12		5/26/2021 5:20 PM	File folder
 cluster-16		5/26/2021 5:20 PM	File folder
 cluster-17		5/26/2021 5:20 PM	File folder
 cluster-18		5/26/2021 5:20 PM	File folder
 cluster-19		5/26/2021 5:20 PM	File folder
 cluster-20		5/26/2021 5:20 PM	File folder
 cluster-21		5/26/2021 5:20 PM	File folder
 dyads		5/26/2021 5:20 PM	File folder
 singletons		5/26/2021 5:20 PM	File folder

同样, 如果您更改了节点和链接设置(通过各种属性着色或调整大小), 您可以将它们保存为样式文件(下图 6), 然后, 您可以通过样式设计选项卡(Styling tab)下的[全球设置\(Global Settings\)](#)菜单将此样式文件与另一个数据集一起使用。重要的是, 当您在新会话中应用样式文件时, 您必须记住使用相同的文件类型(即节点与链接), 否则保存的样式将无法正确应用。



图 6. 保存样式文件



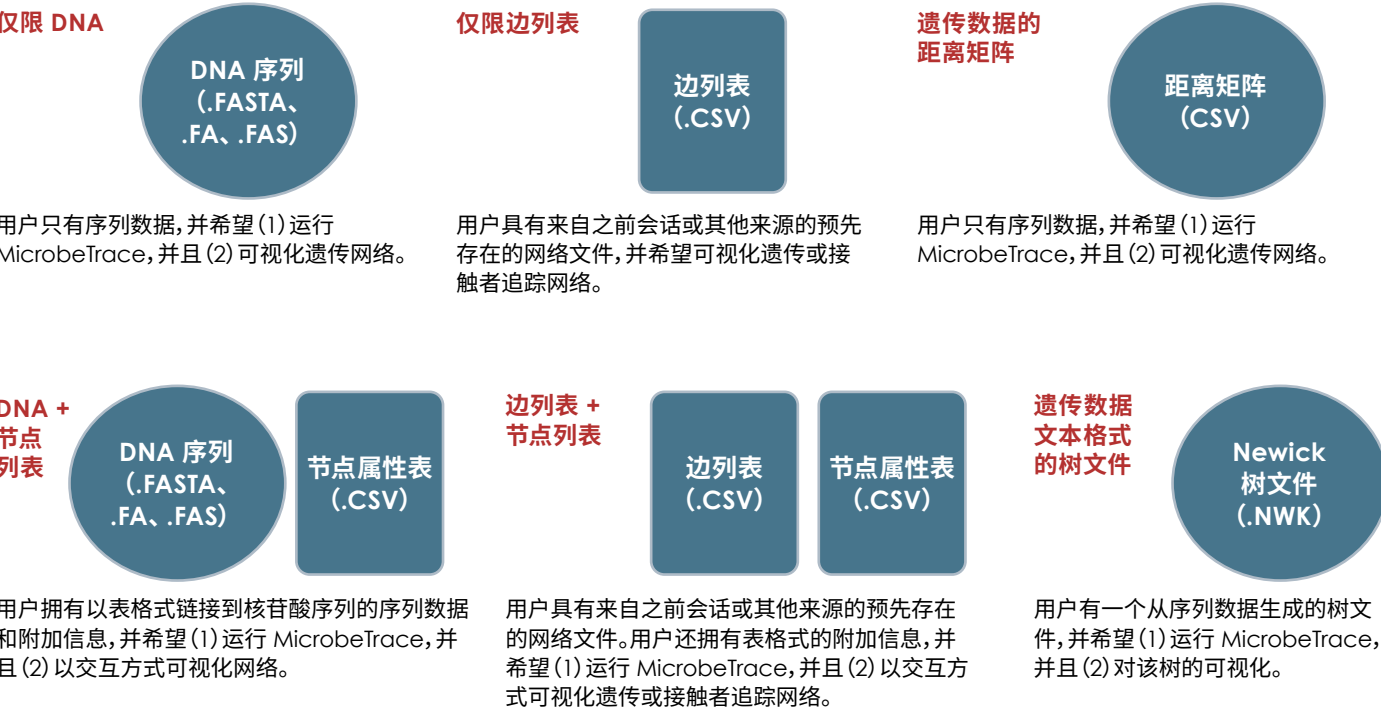
- ✦ **距离矩阵文件(.csv)**。如果您计划从大型序列数据集中分析遗传距离网络, 则此文件类型会特别有用。如果预先计算遗传距离, 且将生成的距离矩阵文件导入 MicrobeTrace, 则在 MicrobeTrace 中的处理时间会显着减少。
- ✦ **Newick(.nwk)树文件**。如果您有一个从外部系统发育工具生成的树, 并且希望将其可视化 MicrobeTrace 中的网络以及该文件中分类群的其他相关数据, 则此文件类型很有用。MicrobeTrace 使用**共祖距离**方法从您的 Newick 文件生成网络图。如果您可能无法访问原始序列数据, 但可以从诸如 Nextstrain (<https://nextstrain.org/>) 或如果您想将系统发育树转换为网络, 然后添加相关的元数据以进行进一步探索的话, 该功能特别有用。

**\*重要通知\***: FASTA 文件中用于序列的 ID 必须与 CSV 文件中的 ID 完全匹配, 并且必须是独一无二的。为获得最佳实践, FASTA 文件中不应使用重复的 ID。如果检测到重复 ID, 则具有相同 ID 的序列会自动修改为下划线和连续数字 (例如, PersonA\_1、PersonA\_2) 以使其具有独特性。新的独一无二的 ID 将传播到所有数据可视化层。

# 可能的文件输入组合

如图 7 所示, 数据文件的组合可以输入到 MicrobeTrace, 具体取决于所需的特定分析或网络可视化。图 7 所示的例子只是一些常用的例子。您还可以加载两个边列表或两个序列文件, 以覆盖来自多种病原体的网络, 或比较从接触者追踪数据和序列数据生成的网络。这些功能将在稍后的手册(覆盖网络)中进行描述。

图 7. 可能的文件组合以创建新网络或可视化先前创建的网络



MicrobeTrace 生成网络的速度取决于数据文件的数量、数据的数量和类型以及您的特定计算机配置。下表为您提供了处理遗传数据的平均时间的估计值。这些结果包括距离矩阵的计算和网络计算。

序列数	持续时间估计
≤ 50	1 秒
100	2 秒
150	2 秒
200	3 秒
250	4 秒
350	8 秒
500	10 秒
750	15 秒
1000	30 秒
1500	75 秒
2000 年	140 秒



# 访问 MicrobeTrace 和加载文件

打开 Chrome 浏览器窗口并导航到地址 <http://microbetrace.cdc.gov>。加载后, 将显示以下 MicrobeTrace 主屏幕(图 8), 您可以选择和加载文件。

图 8. 用于选择和加载数据文件的 MicrobeTrace 主屏幕

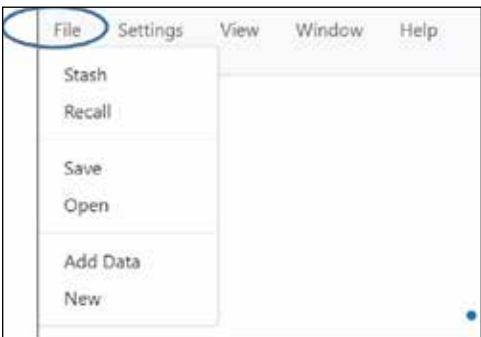


## 主菜单

从主菜单, 您可以访问各种选项(图 9)。

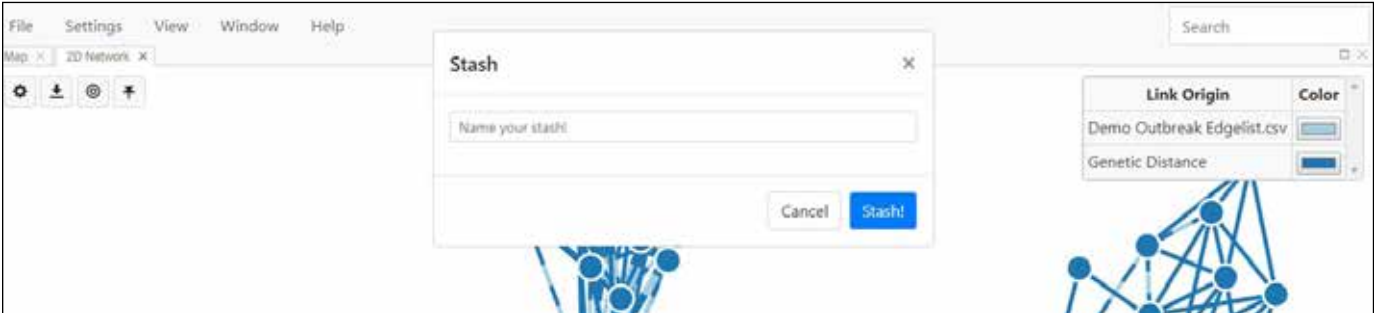
**文件:**允许您打开一个新文件, 将当前会话保存为 .microbetrace 文件, 或将数据文件添加到当前分析。

图 9. 文件菜单



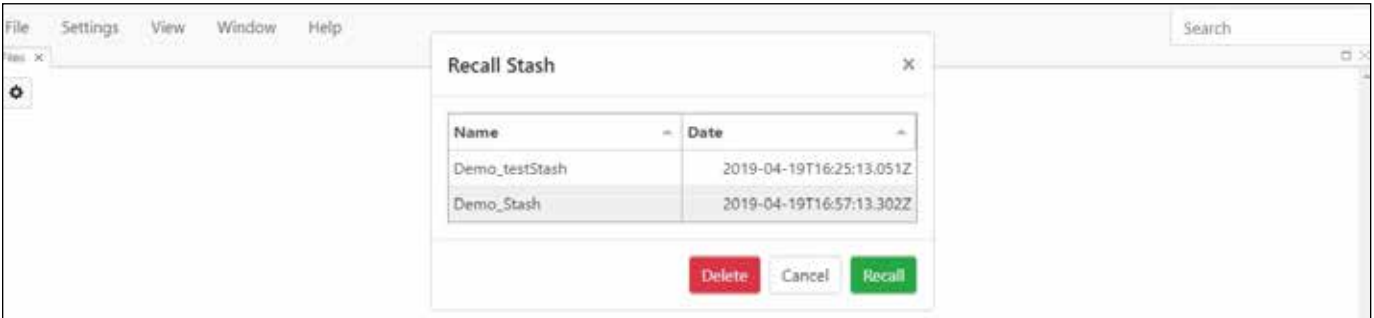
选择 **隐藏(Stash)** (图 10) 将允许 MicrobeTrace 保存您的会话, 以便如果 MicrobeTrace 意外关闭, 您可以重新启动程序并使用文件菜单下的 **召回(Recall)** 功能来恢复您正在处理的数据。这对于保存应用于数据的设置特别有用。

图 10. 隐藏您的 MicrobeTrace 会话



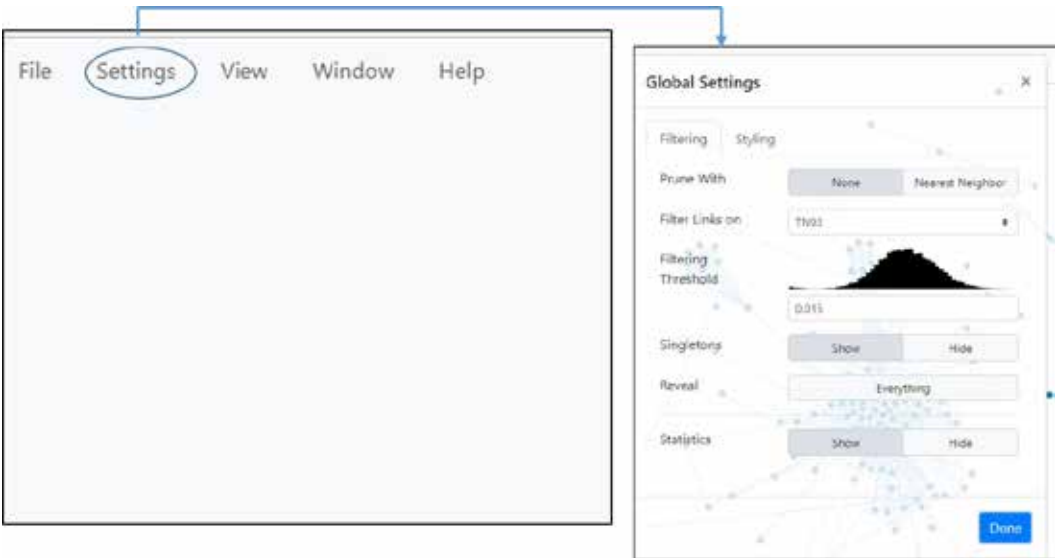
当您从下拉菜单中选择 Recall 时, 会打开一个包含隐藏会话列表的对话框(图 8)。当您从下拉菜单中选择 Recall 时, 会打开一个包含隐藏会话列表的对话框(图 11)。选择您想要的, 然后点击 Recall。

图 11. 召回一个隐藏的会话。



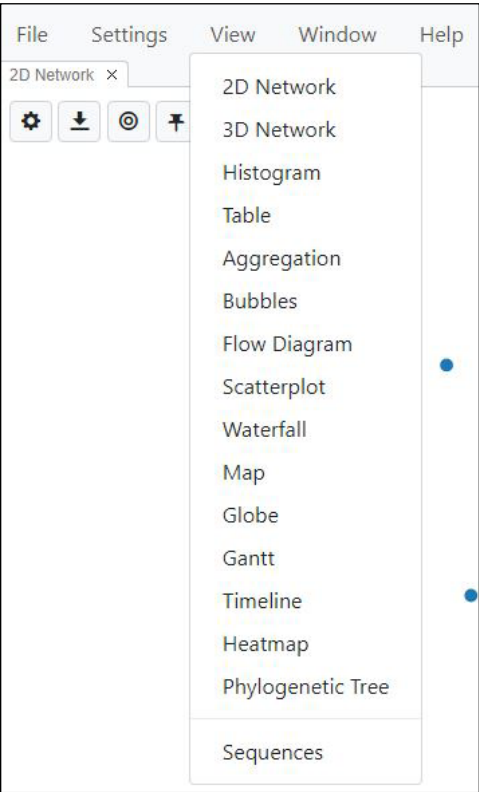
**设置:** 打开全球设置对话框。允许自定义链接和节点样式设置以及遗传距离截止值。每个设置的描述都在 [网络配置\(Network configuration\)](#) 部分。

图 12. 全球设置菜单



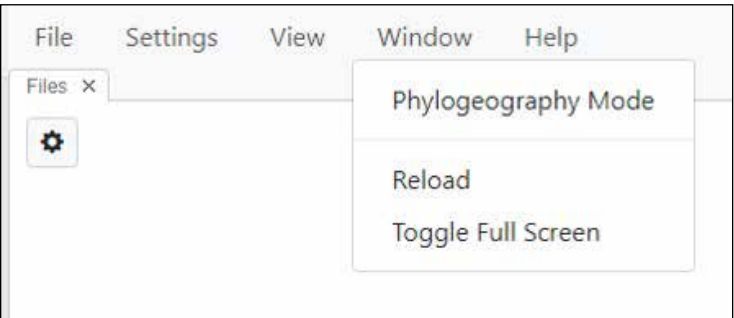
**视图:**为您提供可供选择的数据可视化视图列表, 每个视图将在下面的部分中进行详细描述。如果您想先阅读特定视图, 可以点击目录中的相关标题, 直接进入该视图的描述。

图 13. 查看菜单



**窗口:**此菜单为您提供重新加载数据的选项, 并允许您在全屏视图之间进行切换。

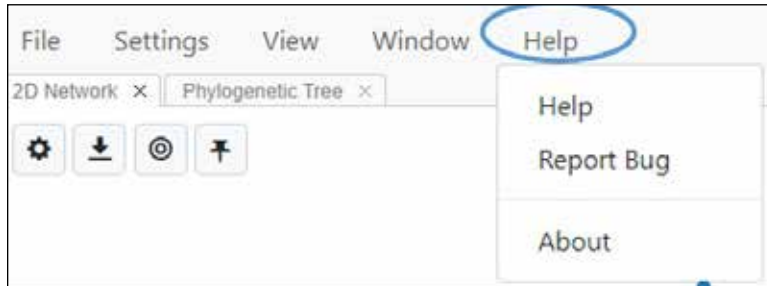
图 14. 窗口菜单



选择重新加载会退出会话, 并返回到“添加文件”屏幕。切换全屏 (Toggle Full Screen) 在全屏视图之间切换。

**帮助:**提供对 MicrobeTrace GitHub 站点上有关 MicrobeTrace 信息的访问(<https://github.com/CDCgov/MicrobeTrace/wiki>), and allows you to report any issues encountered(<https://github.com/CDCgov/MicrobeTrace/issues>)。

图 15. 在 MicrobeTrace 中选择帮助选项



## 加载文件

根据您的计划分析的数据类型, 您可以同时加载多个数据文件。您还可以在 MicrobeTrace 会话期间添加文件。以下是您将使用的相应文件类型的一些常见情况。每种文件类型的详细说明如下所述。

1. 您只有序列数据, 且要构建一个遗传距离网络: 加载 **FASTA 文件**
2. 您有序列数据, 还有一个节点列表, 其中包含与序列相关的案例的人口统计信息或其他信息: 加载 **FASTA 文件** 以及包含列中节点信息的**节点列表 CSV** 文件。
3. 您有一个节点列表, 其中内含人口统计信息或其他信息, 包括带有序列的列, 以及一个联系人追踪文件 (边列表): 加载节点 **CSV** 和 **边 CSV**。这将为提供两个相重叠的网络 - 一个从边计算 (接触追踪网络), 一个从节点 CSV (遗传距离网络) 中的序列计算。


## 加载 FASTA 文件

**第 1 步** 选择添加文件 (Add Files) 按钮 (图 16)。系统显示一个标准的 Windows 资源管理器页面, 用于导航到所需的文件。您还可以将文件从计算机拖放到 MicrobeTrace 窗口中。

图 16. 将文件加载到 MicrobeTrace 中



**第 2 步** 导航到示例 FASTA 文件, 并双击文件以将其添加到分析中, 或选择文件并选择打开 (Open)。这会将 FASTA 文件加载到 MicrobeTrace 中。

**第 3 步** 选择指标 (图 17) - 点击屏幕左上角的设置按钮 。这将显示两个选项卡: 文件和实验性。

文件选项卡允许您自定义以下设置

- ✦ 距离度量 (TN93 或 SNP)
- ✦ 处理歧义 (下方的选项)
  - » 平均 (将匹配的任何分辨率计算为完美匹配)
  - » 解决 (平均所有可能的解决方案)
  - » 跳过 (跳过所有有歧义的位置)
  - » GapMM (将字符间隙位置计算为 4 路不匹配, 否则与平均值相同)
  - » HIV-Trace-g (任何不超过选定比例 [0-1] 的序列都将解决其歧义 [如可能], 并且具有较高比例的序列中的歧义将被平均化。这减轻了由于高度模棱两可的序列造成的虚假联系。)

- ✦ 链接阈值设置(截止以确定两个节点是否链接)
- ✦ 打开视图以启动(2D 网络、3D 网络、气泡、表、系统发育树)

如果您想查看网络图, 请选择 2D 或 3D 网络。如果您希望查看没有链接的节点(在敏感数据的情况下), 请选择气泡。如果您想以表格式而非网络图查看数据, 请选择表。如果您要上传序列或 Newick 文件, 并希望从树视图开始, 请选择系统发育树。请注意, 这仅设置打开视图。然后, 您可以以所有这些格式以及更多格式将您的数据进行可视化。

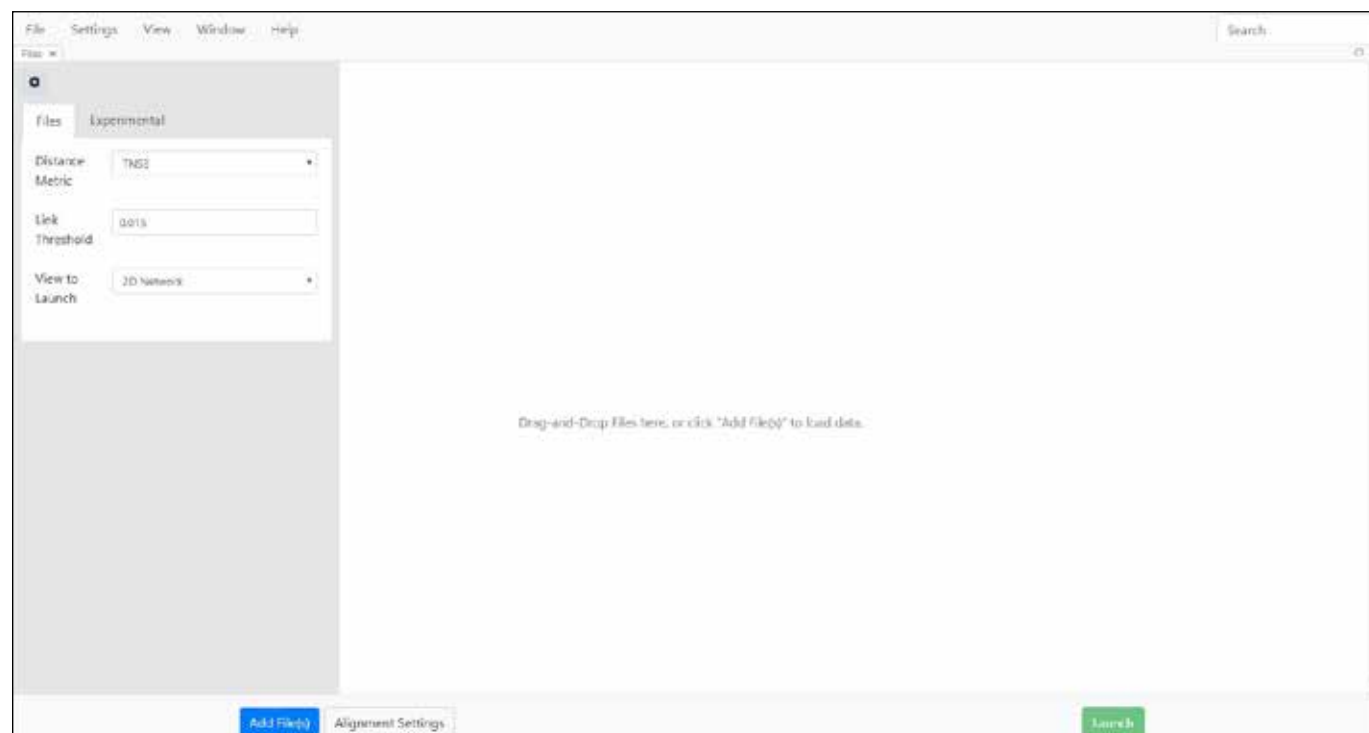
### 默认设置为:

**距离度量:**TN93, **歧义:**平均, **链接阈值:**0.015

**查看启动:**2D 网络

更改这些首选项后, 它们将被保存为您的默认设置, 并在您打开 MicrobeTrace 时应用。您还可以在 MicrobeTrace 会话期间随时更改这些设置。

**图 17. 选择距离度量、阈值和打开视图的首选设置。**



### 实验性标签:

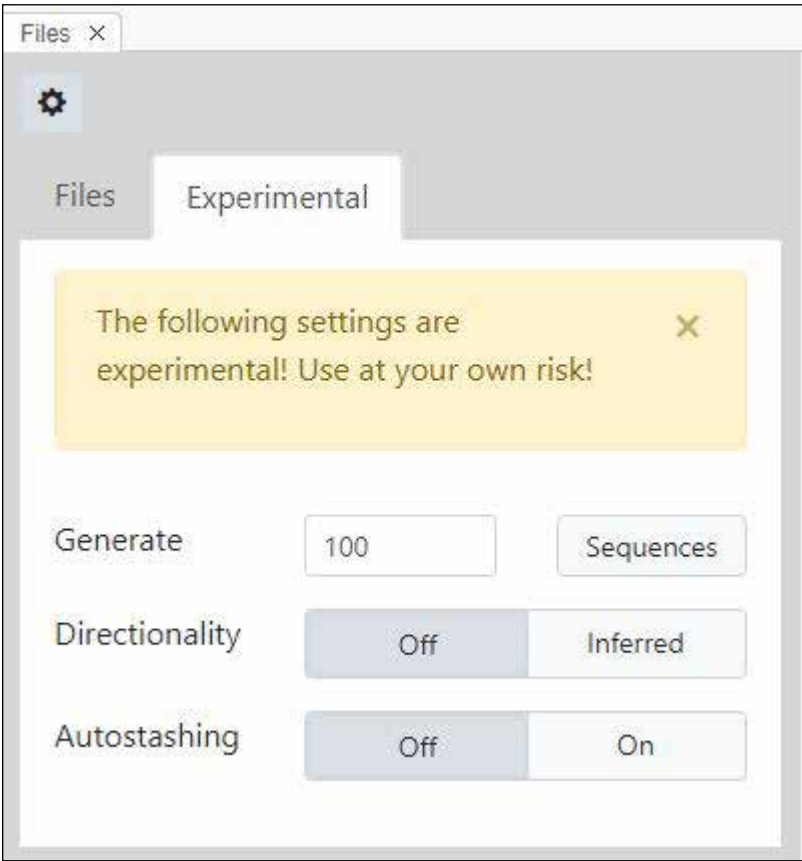
这些选项纯粹是“实验性的”, 允许您尝试其中一些不太常用的功能。您可以生成一组随机序列来探索 MicrobeTrace, 并选择您希望在此数据集中所有的序列数。

您可以选择是否要打开方向性或自动隐藏。**\*重要的提示:通过序列预测方向性是不可靠的, 除非使用额外的流行病学信息(有关更多详细信息, 请参阅[方向性](#))\*得到强有力的支持**

打开**自动隐藏(Autostashing)**允许 MicrobeTrace 自动保存您的会话以供以后重新加载。打开后, 会话每分钟都会保存到本地存储。

这两个参数默认关闭。

图 18. 实验性标签设置



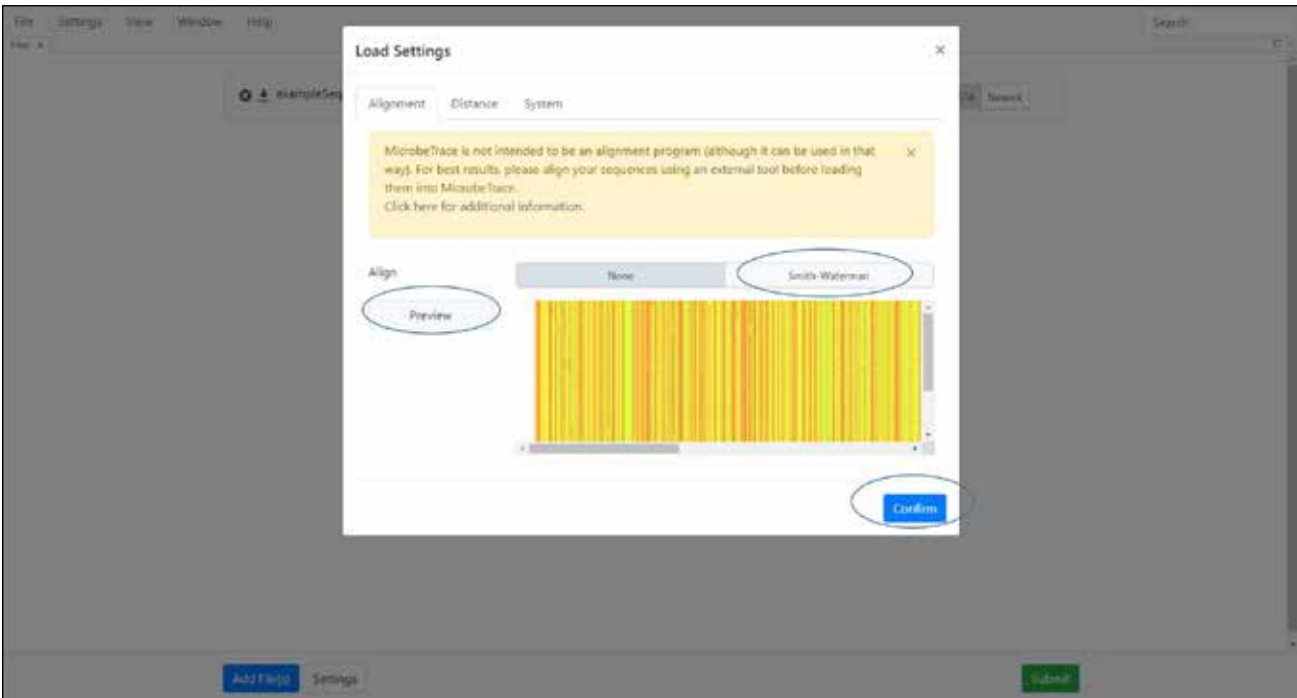
**第 4 步 设置(Settings)**按钮可以让您选择将您的序列与参考序列进行比对。

**\*重要提示\***:默认选项是指您的序列不会被比对。请确保您的序列在网络分析之前比对;在 MicrobeTrace 中(以下步骤)或使用您选择的比对软件。

**\*重要提示\***:如果您使用的是预先比对的序列, 请直接进入提交(Submit)步骤, 这将使用默认设置提交序列。如您要求 MicrobeTrace 对已比对的序列进行比对, 它可能会给您错误的网络。

如果您希望 MicrobeTrace 比对您的序列, 请选择**设置(Settings)**。序列文件加载时默认**比对(Align)**按钮设置为无。

**图 19. 加载设置视图以分析序列并选择序列预览**



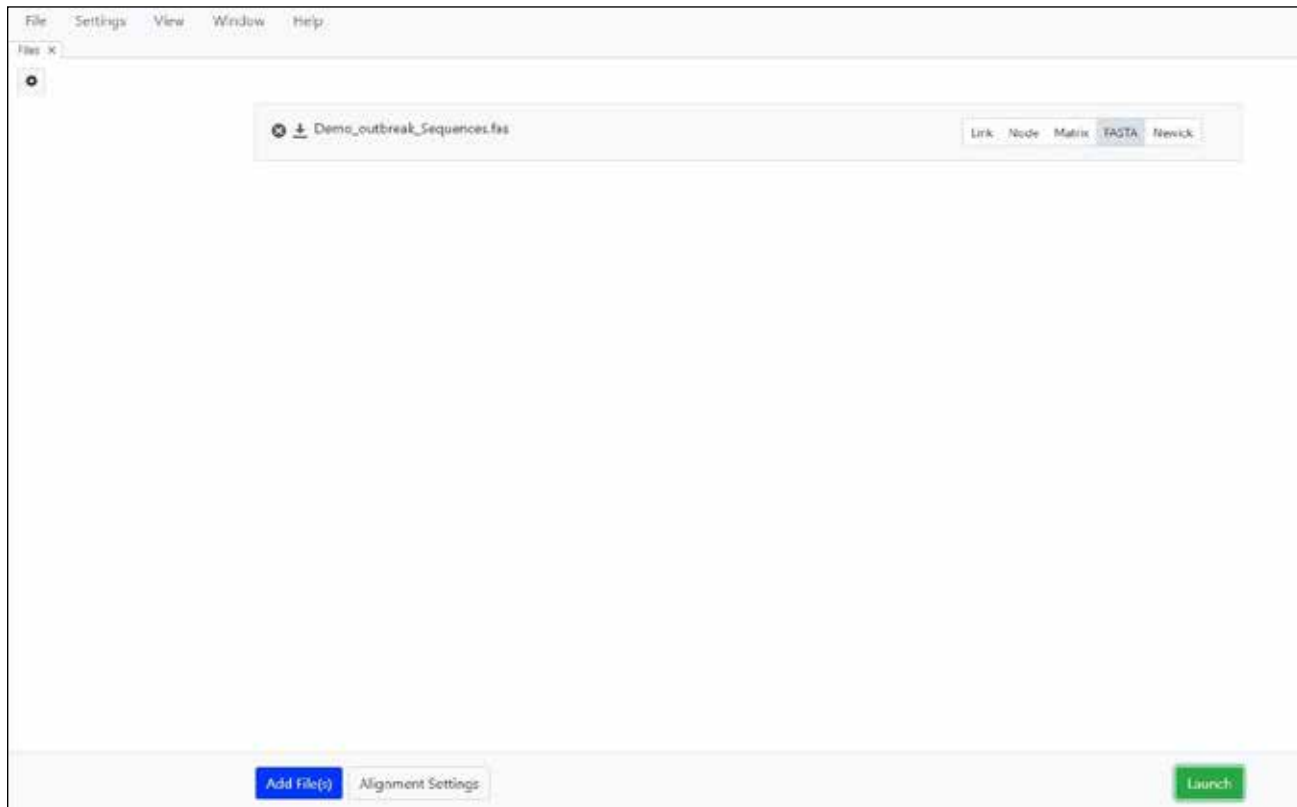
如果您需要比对序列, 请选择 [Smith-Waterman](#) (图 16)。如果您需要比对序列, 请选择 Smith-Waterman (图 16)。默认设置是将您的序列与 HIV-1 HXB2 pol 区域参考 (HIV-1\_HXB2 GenBank 登录号 K03455) 进行比对。如果您希望将您的序列与您选择的参考进行比对, 请使用 **浏览(Browse)** 按钮导航, 并加载您的参考。

您将在窗口中看到 HIV (HXB2.pol)。选择**确认(Confirm)**按钮, 以接受此选项。如果您希望加载特定的参考序列文件, 请选择**浏览(Browse)**按钮, 并从计算机上的适当位置加载文件, 然后点击**确认(Confirm)**。

点击**确认(Confirm)**后, 您将看到以下屏幕 (图 20)



图 20. 选择用于上传/分析序列的比对选项后显示



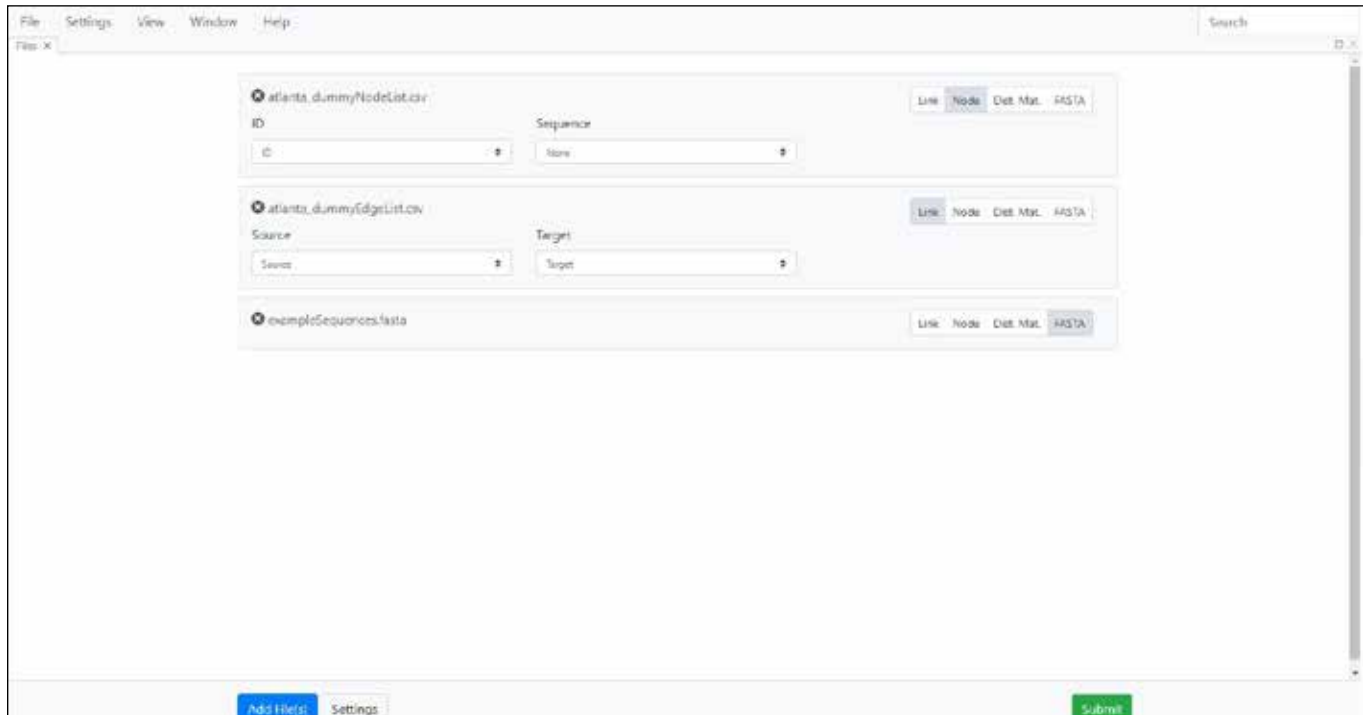
**第 5 步** 如果您不需要加载节点 CSV 文件, 且只想将遗传距离网络进行可视化, 请选择立即**启动(Launch)**。如果您想在节点文件中加载人口统计/epi 数据, 请继续下一部分。

**\*重要提示\***:我们强烈建议您对非 HIV 病原体使用预先比对的序列, 因为 MicrobeTrace 主要不是针对其他病原体的比对程序。MicrobeTrace 当前配置用于确定 FASTA 文件中仅 HIV-1 *pol* 序列与嵌入式比对算法中的参考 HIV *pol* 序列(HIV\_HXB2 GenBank 登录号 K03455)之间的遗传距离。对于其他病原体, 我们建议加载**预先比对**的核苷酸序列文件作为 MicrobeTrace 所需的 FASTA 文件输入, 并跳过比对选项步骤并使用默认设置(无比对)。

## 加载节点列表和/或边列表

如果您有与网络中的节点关联的其他数据,也可以将这些数据导入 MicrobeTrace(图 21)。此数据必须以 CSV 或 Excel 文件格式准备,并包含一个 ID 列,其值与 Edge CSV 文件的源或目标列和/或 FASTA 文件中的序列 ID 匹配。如果存在多个 ID 列,则将使用该文件中最左侧(第一)的列。要加载节点或边文件,请选择添加文件(Add File)按钮,或像以前一样拖放到 MicrobeTrace 主窗口中。图 21. 显示加载了三种文件的屏幕。MicrobeTrace 尝试自动确定已加载的文件类型(如,节点、链接、距离矩阵或 FASTA),并在文件上传菜单的右侧对其进行选择,然后它会变灰以显示所选。请经常检查以确保选择正确。

图 21. 加载了三种类型的文件(节点、链接、FASTA),并准备好进行分析的屏幕视图

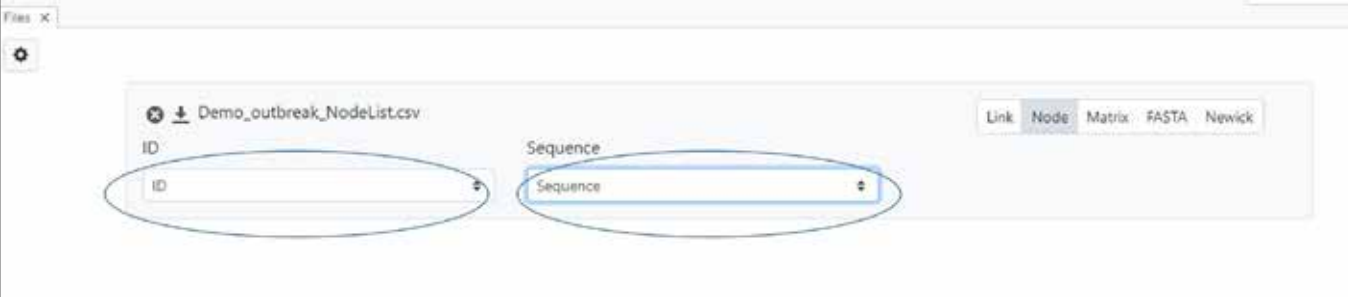


**\*注意\*** 如果您的节点列表文件包含带有序列的列, MicrobeTrace 将自动识别和分析这些序列,您将拥有与单独的序列文件相关的所有功能。在这种情况下,无需加载单独的 FASTA 文件。

在节点文件的 **ID** 列下拉菜单中,节点名称的默认选择是 ID;然而,如果您的节点文件有一个不同的列标题来表示您希望使用的 ID,您可以使用下拉菜单选择该列标题。

如果节点列表中有包含序列的**列**, 请确保在加载节点列表后从下拉菜单中选择它, 以便将其包含在分析中 (图 22)。

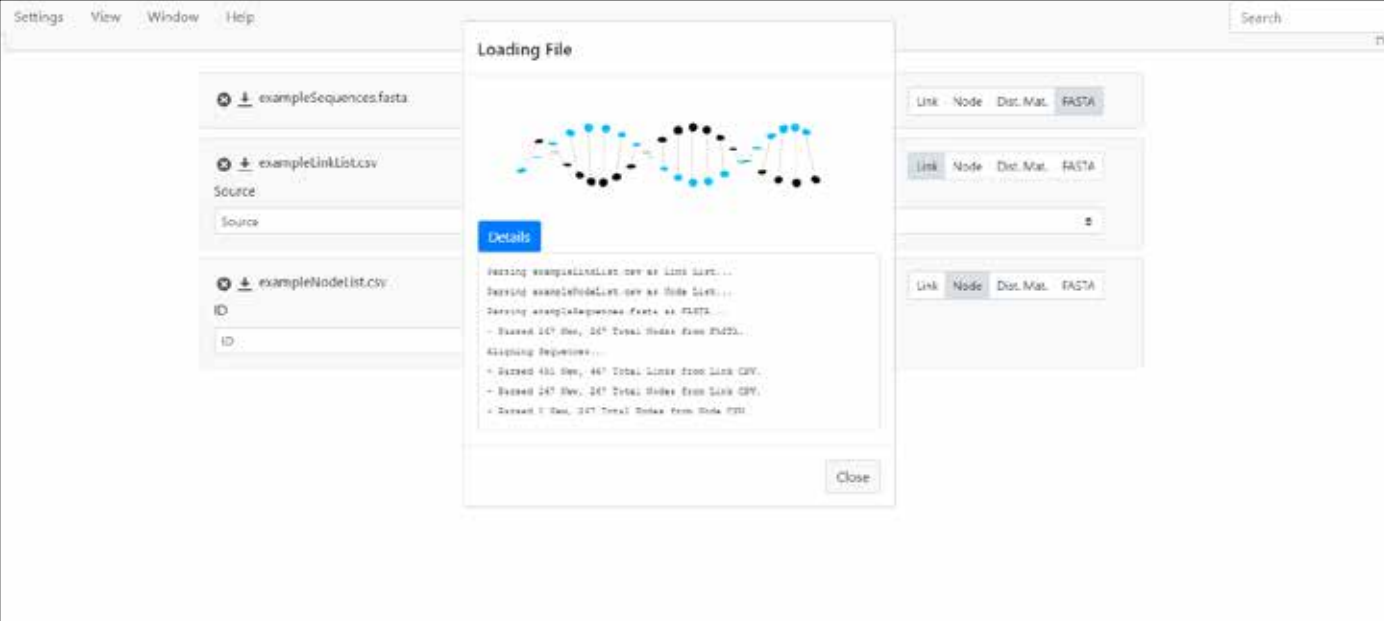
图 22. 加载带有序列列的节点列表。



**\*重要提示\***:节点列表中具有相同节点 ID 的行会导致之前具有相同 ID 的行被覆盖。请确保节点 ID 是独一无二的。

选择**提交(Submit)**, 以开始分析。随着文件的加载和遗传和/或社交网络的推断, 将出现一个状态栏来显示分析的进度(图 23)。

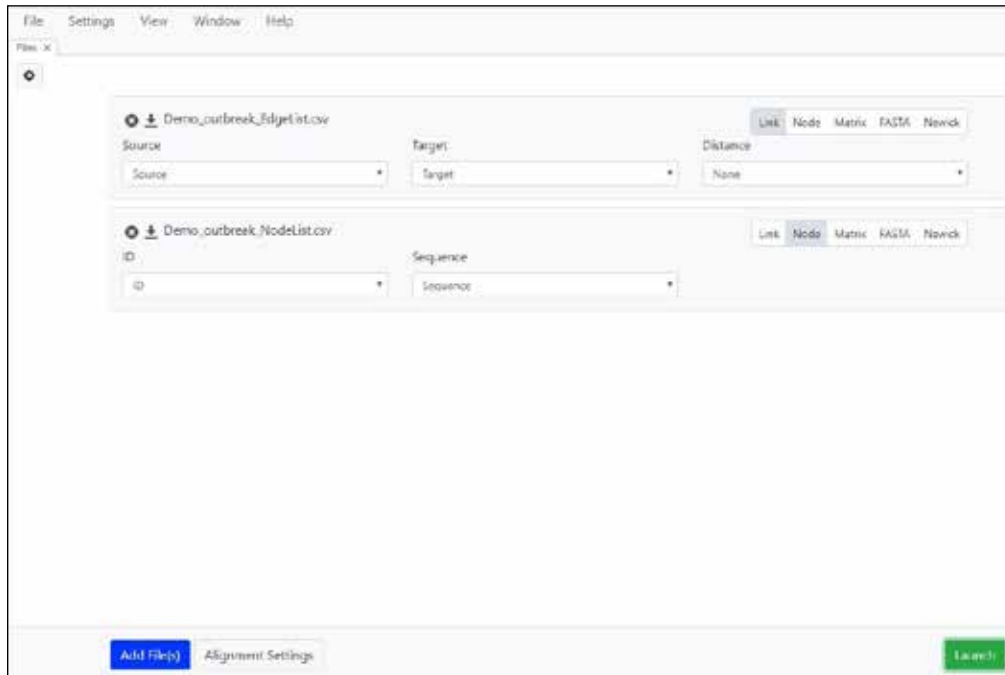
图 23. 显示文件加载和数据分析进度的状态窗口



## 加载边列表 而非 FASTA 文件

如果您已经准备了一个边列表(CSV 或 Excel 格式),那么您可以将该文件直接加载到 MicrobeTrace 中(图 24)。链接文件必须包含源(例如感染者)和**目标**(例如感染者或与感染者的接触者)列。任何附加边属性都可以作为附加列包含在边列表文件中。选择添加文件(Add File)按钮,导航到您的 CSV 文件并双击该文件,或者选择该文件,然后选择打开。加载文件后,您可以点击提交(Submit)以运行分析,或者如果您有与研究中的人员相关联的人口统计和 Epi 数据,则可以添加节点文件。为此,请参阅前面关于加载可选节点数据的部分。

图 24. 加载链接和节点文件后的屏幕视图

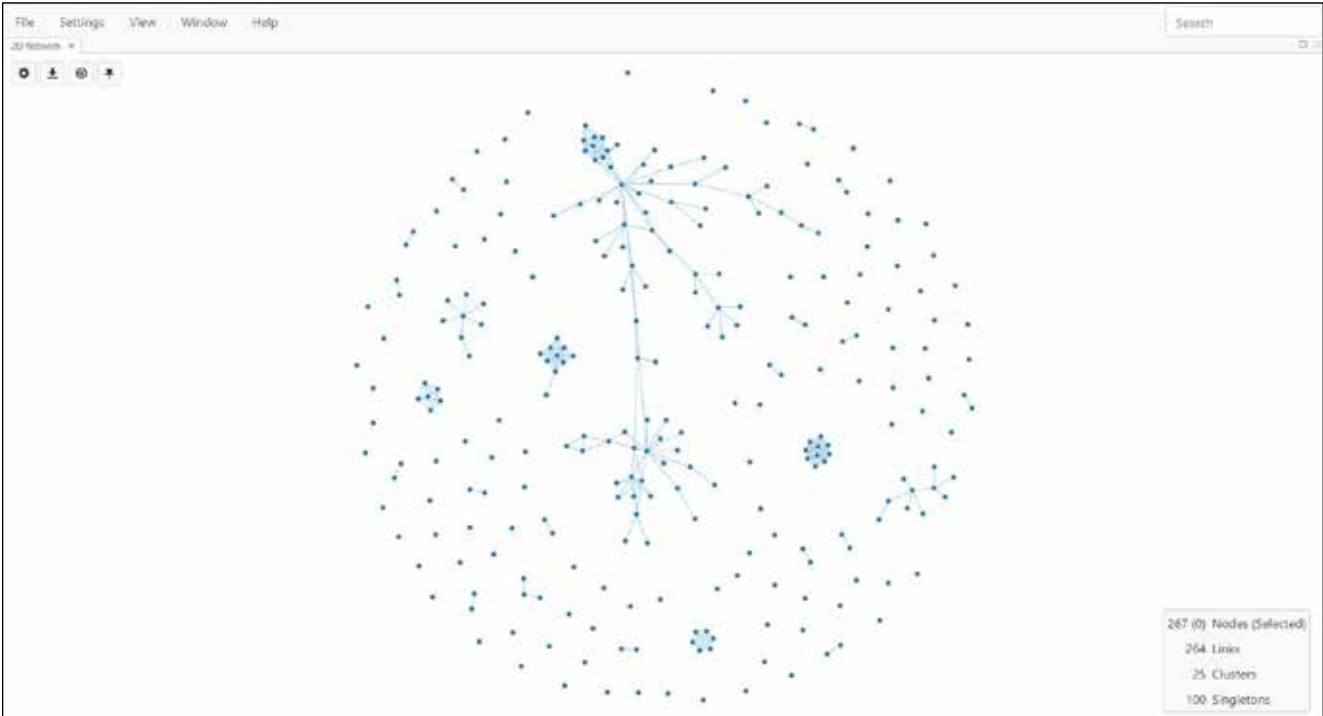


文件加载和数据分析完成后,软件会在页面右下角显示网络图和汇总统计表(图 25)。此表显示节点数、链接(边)数、**集群数**和**单例数**。

# 数据可视化

默认的数据可视化和探索方法是二维(2D)网络视图(图 25)。您可以从视图(View)菜单中选择多种不同的数据可视化, 并将数据显示为表、流(也称为冲积)图、直方图、成对遗传距离矩阵的热图、核苷酸序列的比对、核苷酸序列遗传关系的系统发育树, 或显示节点位置的地图。其中一些查看选项特定于文件类型。例如, 直方图、序列视图、热图或系统发育树需要核苷酸序列数据。以下部分介绍了其他数据可视化视图。每次选择视图时, 它都会在新选项卡中打开, 因此您可以在视图之间轻松移动。您还可以点击选项卡, 并将其拖动以创建并排窗口, 以使您能够比较视图。将两种或多种不同格式的数据一起可视化通常很有帮助。您将在稍后的描述各个视图的部分中看到此功能的示例。

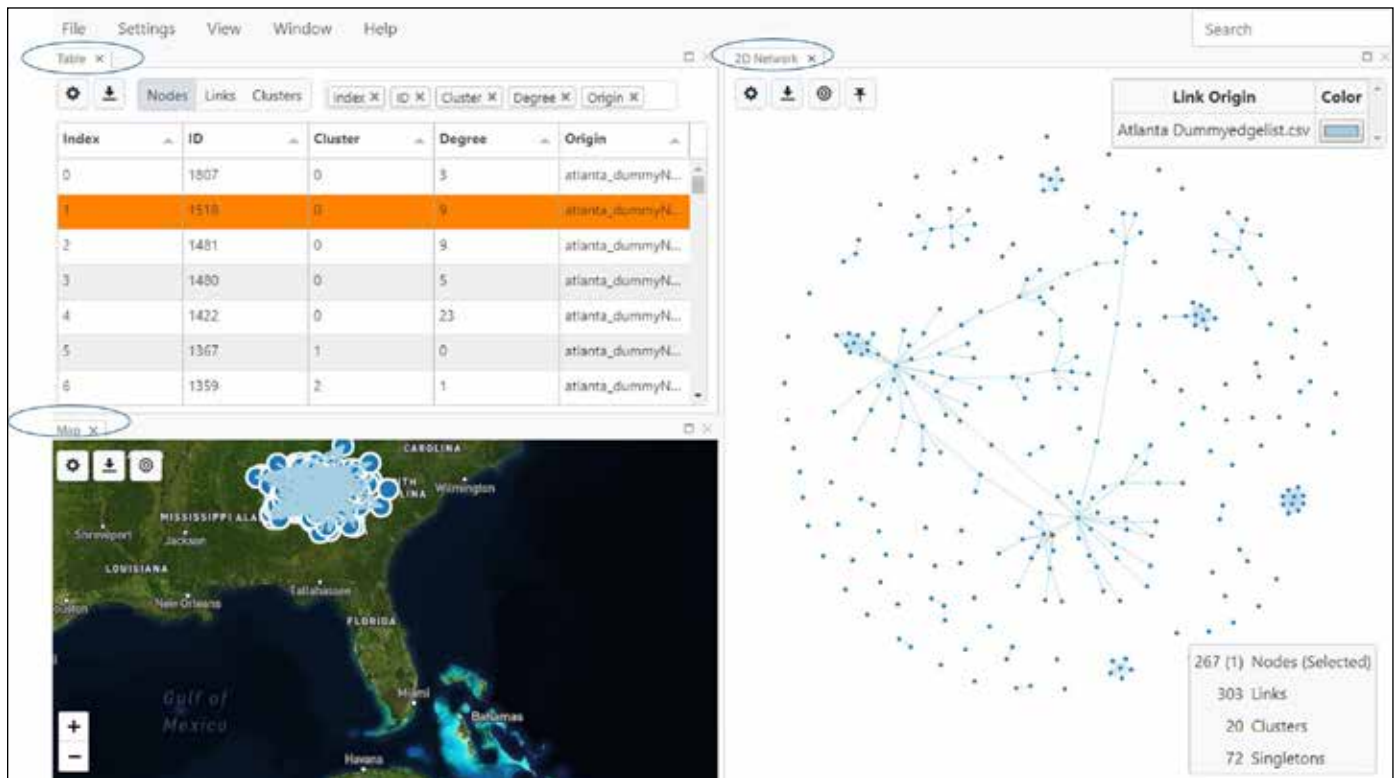
图 25. 默认 视图 - 2D 网络



## 平铺不同的视图

每次从主下拉菜单中选择视图时,都会打开一个新选项卡。您可以通过在选项卡之间移动来在视图之间切换,就像在浏览器中一样。此外, MicrobeTrace 还允许您并排显示多个视图,或者在同一窗口中平铺排列。您可以通过点击视图选项卡并同时将其拖动到窗口中当前打开的视图旁边或下方的位置,来执行此操作。在下面的示例中(图 26),打开了一个 2D 网络窗口,然后在新选项卡中打开了表和地图视图。转到 2D 网络窗口,点击表视图(Table View)选项卡,然后将其拖动到网络视图的左侧。现在点击地图视图(Map View)选项卡,并将其拖到两个打开的窗口下方。您可以根据需要平铺尽可能多的视图。

图 26. 通过选择多个视图平铺视图,并将它们排列在同一个窗口中,以便它们都可见。







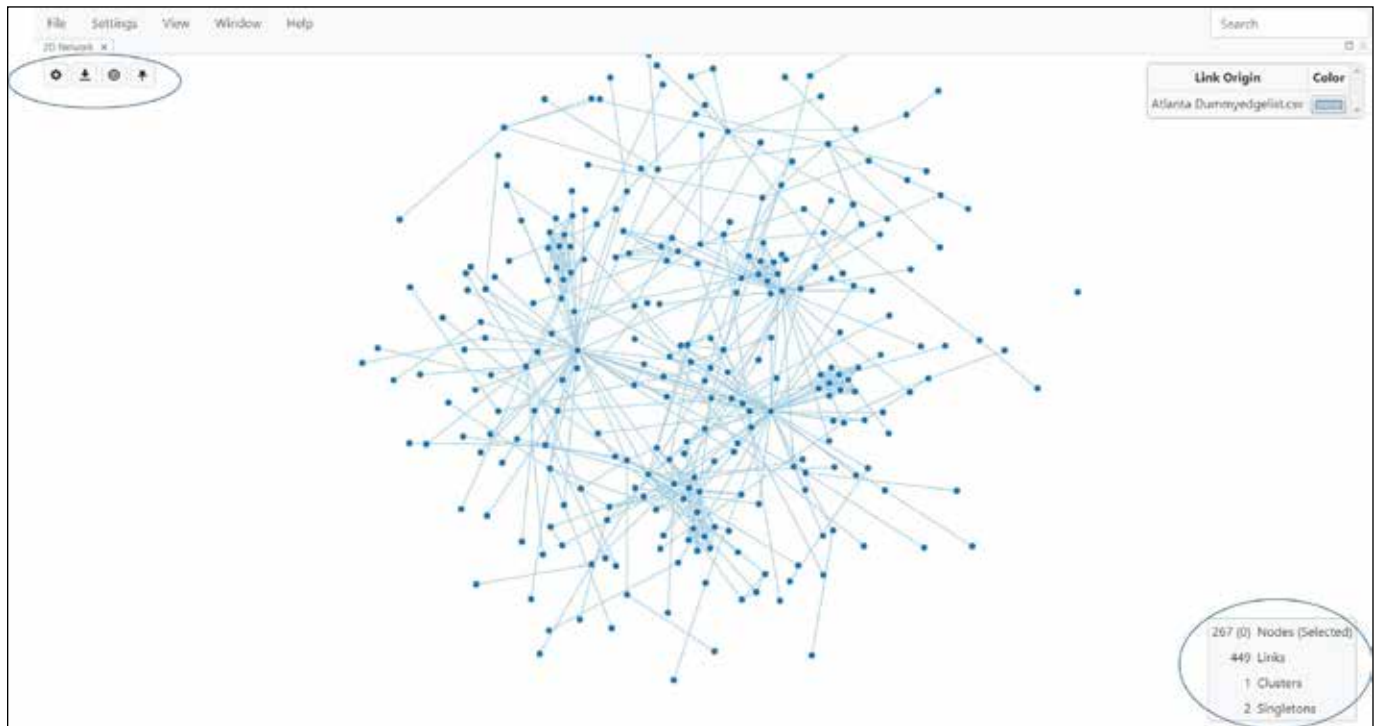




所有视图都在屏幕的左上角包含这四个图标中的两个或多个     (图 27)。将鼠标悬停在每个图标上将为您提供每个图标的描述。

图 27. 2D 网络视图, 设置按钮位于窗口左上角



1. **切换设置:**  允许您调整节点、链接和网络本身的各种参数, 如形状、大小、颜色和物理(电荷、摩擦和重力)。下面的网络配置部分提供了详细信息。在其他视图中, 此按钮可让您自定义该特定视图的设置。选择切换设置 (Toggle Settings) 按钮一次会显示菜单, 并再次选择它会隐藏菜单, 因此您可以在显示和隐藏菜单之间切换。
2. **输出:**  允许您将当前视图导出为图像或其他适用的文件类型。您可以使用下拉菜单选择文件类型。如果要导出为图像, 高级设置允许您根据应用程序调整大小。还有一个 .svg 选项可让您在不损失分辨率的情况下增加图像大小。这对于海报和出版物特别有用。下载的图像将包含 MicrobeTrace 徽标作为水印。您可以使用导出设置中的滑块调整 MicrobeTrace 水印的不透明度。
3. **中心和比例:**  您可以使用鼠标滚轮放大和缩小网络。此按钮重新缩放并将其居中返回到屏幕上的原始大小。
4. **固定所有节点:**  网络是一个动态结构, 您会看到它在渲染时略微浮动。此按钮将“冻结”网络。您可以在拖动节点后执行此操作, 以便为您的数据提供最佳视觉效果(请参阅下面的节点部分)。



## 2D 网络视图

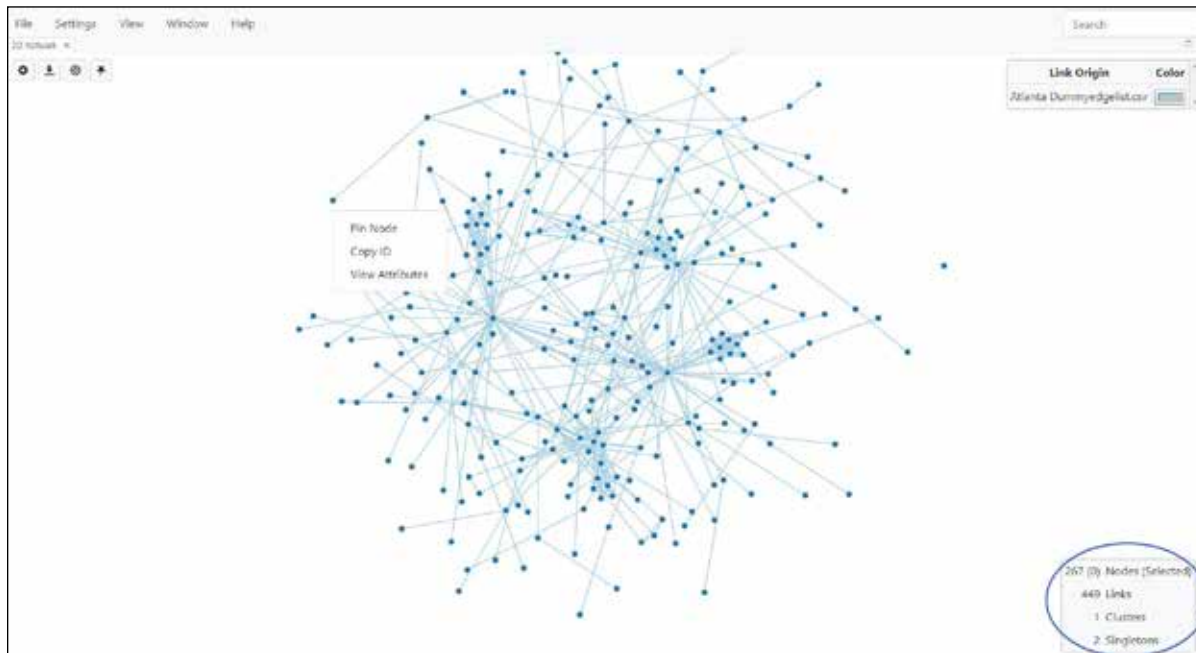
这是 MicrobeTrace 处理数据后您将看到的默认的可视化窗口(图 26)。在网络视图中, 您可以:

- ✦ 通过点击窗口中的任意位置进行平移、拖动节点、或使用鼠标上的滚轮放大或缩小
- ✦ 选择或取消选择单个节点(使用 Ctrl + 点击选择单个节点;使用 Shift + 点击选择多个节点)。您可以选择多个节点, 并在不同的视图中对其进行查看, 这在下面更详细描述的表面视图中是非常有用的)。此节点选择功能允许您仅能查看所选节点的所有特征。您可能只想查看集群中的节点, 或您在网络视图中感兴趣的任何节点的流行病学数据。
- ✦ 右键点击节点, 以查看各种选项(图 28)
  - » **固定节点:**固定节点允许您将其拖出, 以便使该集群中的节点具有更好的可视化, 并探索集群。
  - » **复制 ID:**复制节点 ID
  - » **查看属性:**允许您查看节点属性(与该节点关联的元数据)

请注意, 每次拖动节点时网络的渲染都会更新, 因此您会看到节点在进入静态位置之前四处移动。每次更改任何网络设置时, 您都会看到此渲染(有关设置中的选项, 请参见下文)。您可以使用**固定所有节点(Pin All Nodes)**图标来冻结网络。



图 28. 通过右键点击节点查看和探索节点选项



## 网络配置

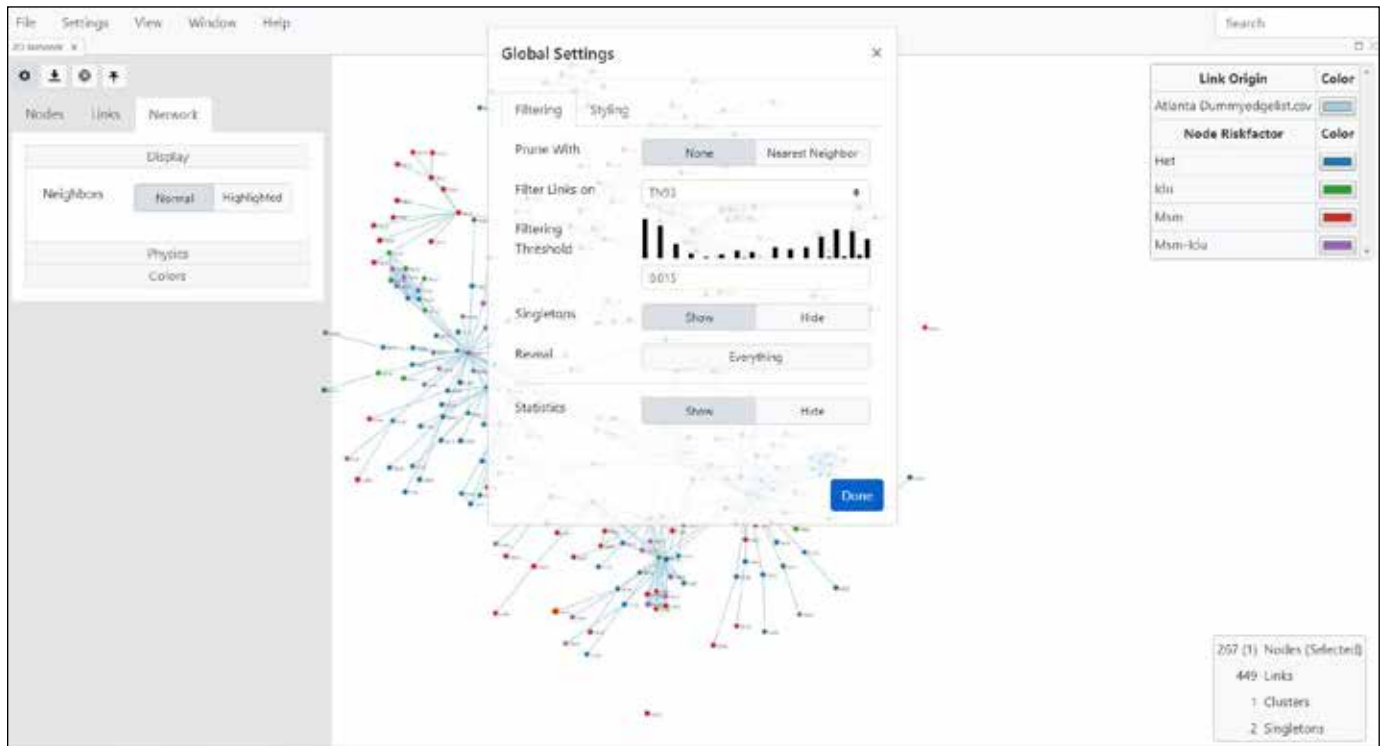
您可以修改所显示网络的多个视觉特征。首先,我们将描述全球设置(Global Setting)菜单,然后描述特定于节点、链接和网络的设置。

通过选择左上窗口主菜单栏上的设置来访问全球设置(Global Setting)菜单。此菜单选项具有三个选项卡:**过滤(Filtering)**选项卡,用于修改网络的截止值和其他特征,**样式(Styling)**选项卡用于为节点、链接和背景着色(**样式(Styling)**选项卡在稍后节点部分中描述),和**时间轴(Timeline)**选项卡,它为您提供了一个时间播放器来追踪集群随时间的增长。

## 过滤

此选项卡允许您设置距离阈值,使用最近邻算法修剪链接,并设置用于定义集群的最小节点数的阈值。此功能允许您按集群大小进行过滤,这在某些爆发情形中可能很有用,尤其是当您想从网络视图中删除所有单例时。此选项卡还允许您显示或隐藏在分析期间自动生成,并显示在窗口右下角的框中的网络统计信息(参见图 29 中的圆圈)。

图 29. 全球设置, 过滤选项卡。



**最小集群大小:**允许您设置用于可视化集群的节点数。默认设置为 1, 这意味着也显示单例。要隐藏单例或未连接到任何其他节点的节点, 您应该将此数字设置为 2 (将显示大小为 2 或更大的集群)。

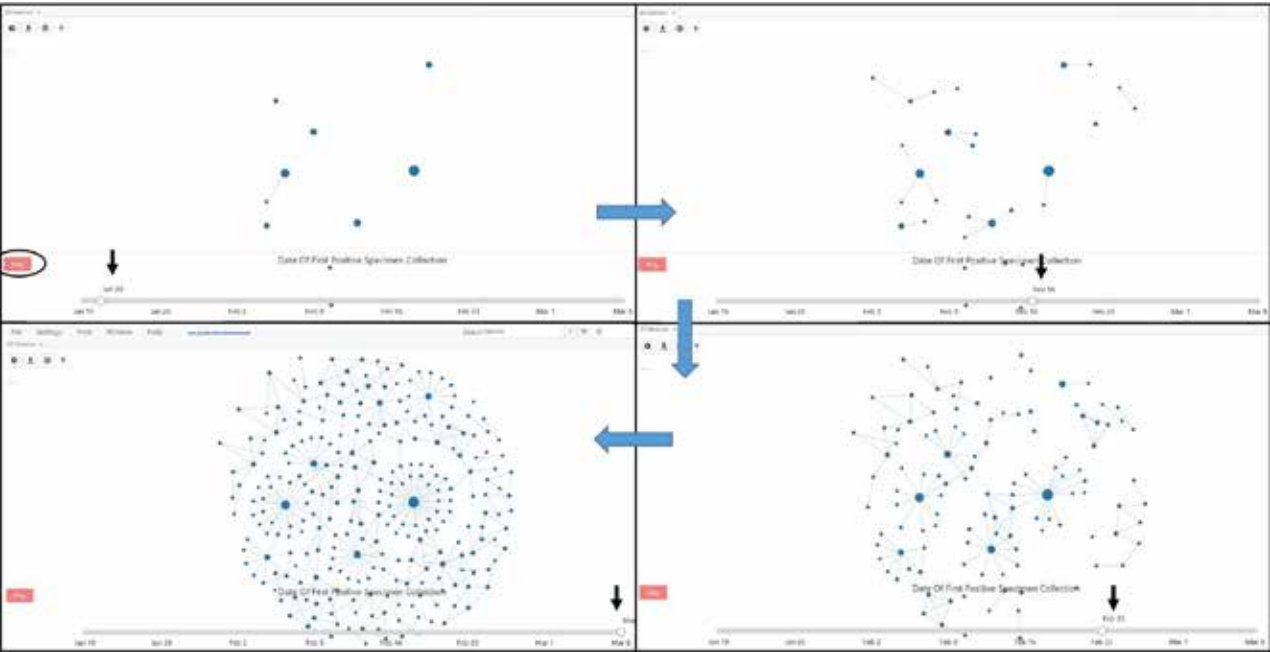
**时间轴选项卡:**时间轴选项卡允许您监测网络及其集群随时间的增长。它提供了数据日期范围内任何给定时间点的传输快照。您可以使用下拉菜单从数据集中的任何日期字段中进行选择 (图 30)。选择属性后, 时间播放器会在屏幕底部显示。请注意, 时间播放器也将在其他视图中可见。这可能是地图视图中查看网络时空成长的一个有价值的特征。

图 30. 全球设置菜单上的时间轴选项卡。您可以从数据集中选择日期字段。



点击**播放(Play)**按钮将启动时间播放器, 您可以看到网络在数据集中的日期范围内发生变化(图 31)。您可以暂停、倒退或前进。

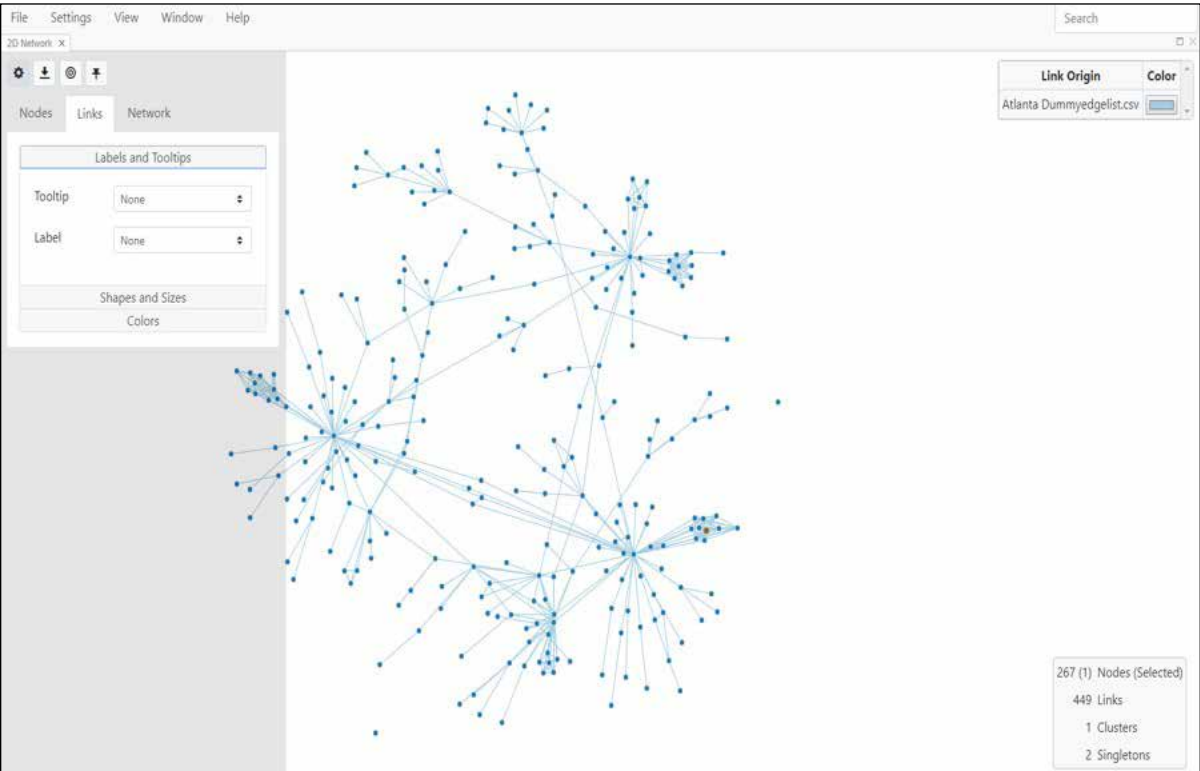
图 31. 网络时间播放器在日期范围内不同时间点的屏幕截图。随着网络的变化, 您可以看到链接在一段时间内是如何形成的。



# 节点和链接设置

选择前面提到的切换网络设置 (Toggle Network Setting) 图标, 以显示上下文菜单 (图 32)。您可以从三个选项卡中进行选择: “节点”、“链接”或“网络”来调整各种网络组件的设置。使用鼠标, 将鼠标悬停在每个属性名称上以查看它所代表的内容。

图 32. 在节点 (Nodes) 选项卡中更改网络配置的可用设置



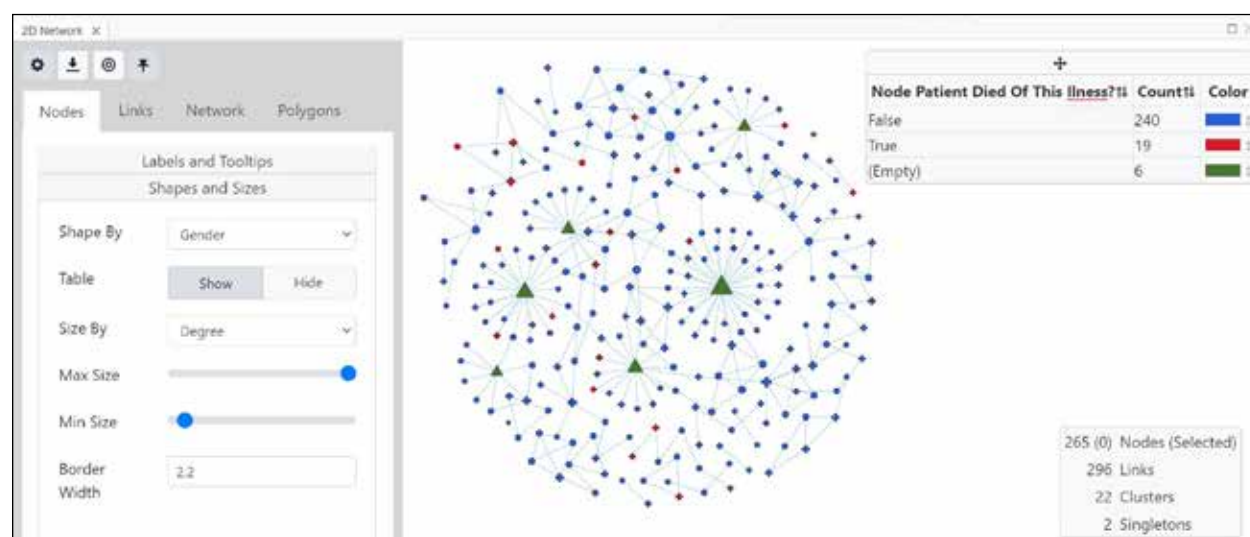
## 节点属性

**标签和工具提示:**您可以通过从每个参数的下拉菜单中进行选择来更改节点的标签和工具提示。您可以为工具提示选择多个变量。您可以使用滑块更改标签大小。您还可以选择标签相对于节点的方向。

**形状和大小:**您可以选择形状和大小(Shapes and Sizes), 将形状映射到节点, 并将节点的大小映射到从下拉菜单中选择的人口统计特征。节点的默认形状是圆形, 但可以通过使用下拉菜单进行更改。您可以设置节点的最小和最大大小, 以便即使按变量大小调整所有节点也是可见的。您还可以更改节点边框的粗细。

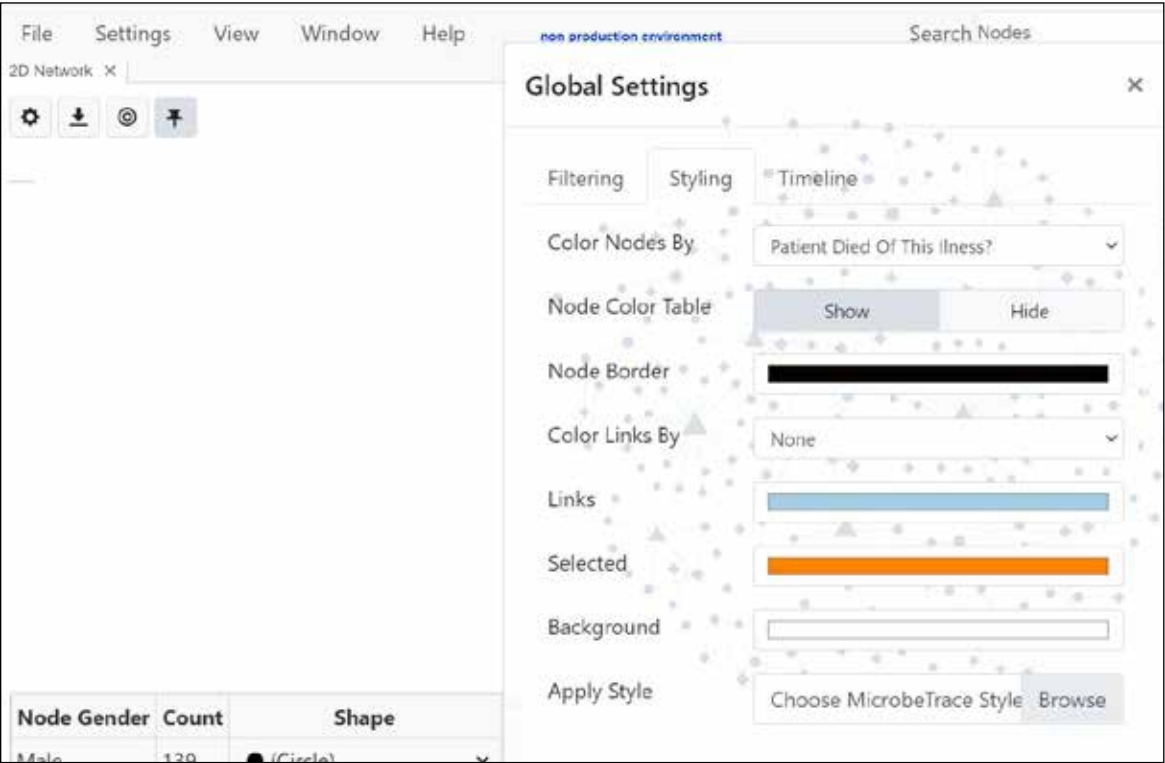
在图 33 中, 形状已映射到性别, 并且在窗口的右上角显示了一个键。节点按度(到节点的链接数)确定大小。

图 33. 形状和大小(Shapes and Sizes)菜单用于更改节点特征, 节点形状映射到性别, 按程度和边框宽度增加来调整大小。



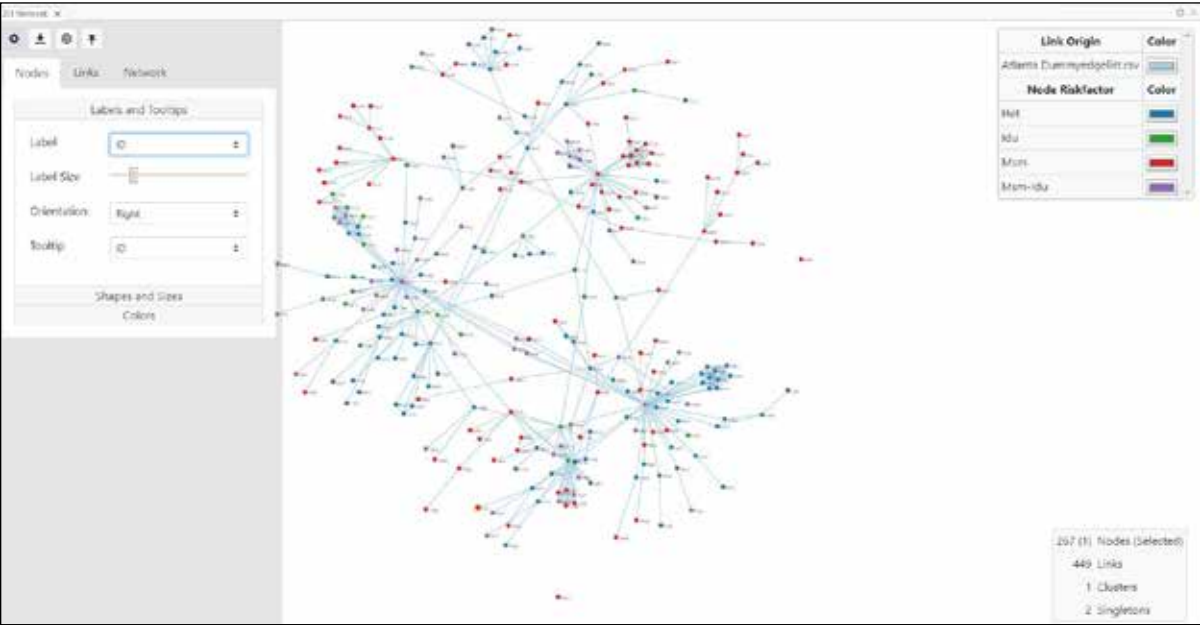
**颜色:**选择颜色(Colors)按钮会将您带到全球设置(Global Settings)菜单的样式选项卡, 在其中, 您可以使用下拉菜单将节点的颜色或链接映射到人口统计数据。您还可以更改节点边框颜色以及背景颜色(图 34)。还有一个名为应用样式(Apply style)的选项, 您可以在其中浏览并选择以前保存的 MicrobeTrace 样式文件。这将使您能够使用当前数据集重新创建该样式。重要的是, 当您在新会话中应用样式文件时, 您必须记住使用相同的文件类型(即节点与链接), 否则保存的样式将无法正确应用。保存样式文件的过程在[这里](#)描述。

图 34. 节点、边框、链接和背景的颜色和样式



在下面的示例中, 节点标签已从**标签 (Label)**下拉菜单中设置为 ID, 并已按风险因素着色, 方法是通过从**颜色节点依据 (Color Nodes By)**下拉菜单中选择**风险因素 (Risk Factor)**, 然后选择**完成 (Done)**。然后, 您将看到网络上节点的 ID 标签, 且在窗口右上角的文本框中将提供一个颜色映射键 (图 35)。

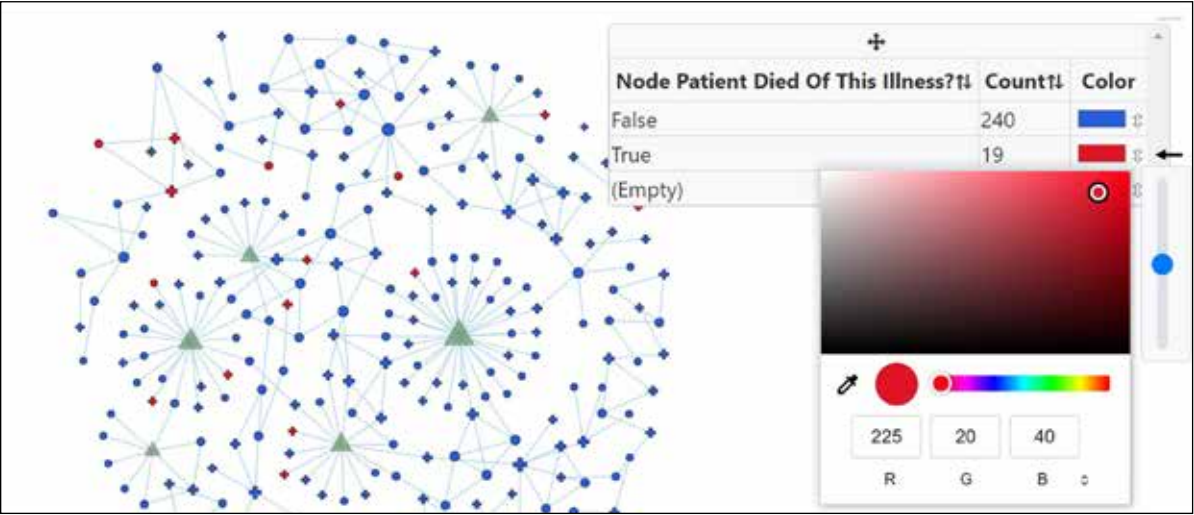
图 35. 节点标有 ID, 并按风险因素着色的节点设置示例





可以编辑右上角的键(图 36)。您可以通过点击颜色条来更改每个变量的颜色, 然后拉出一个颜色图表供您选择。点击颜色旁边的双向箭头将为您提供一个透明度滑块, 以更改该颜色的透明度。您还可以通过点击来编辑文本列。

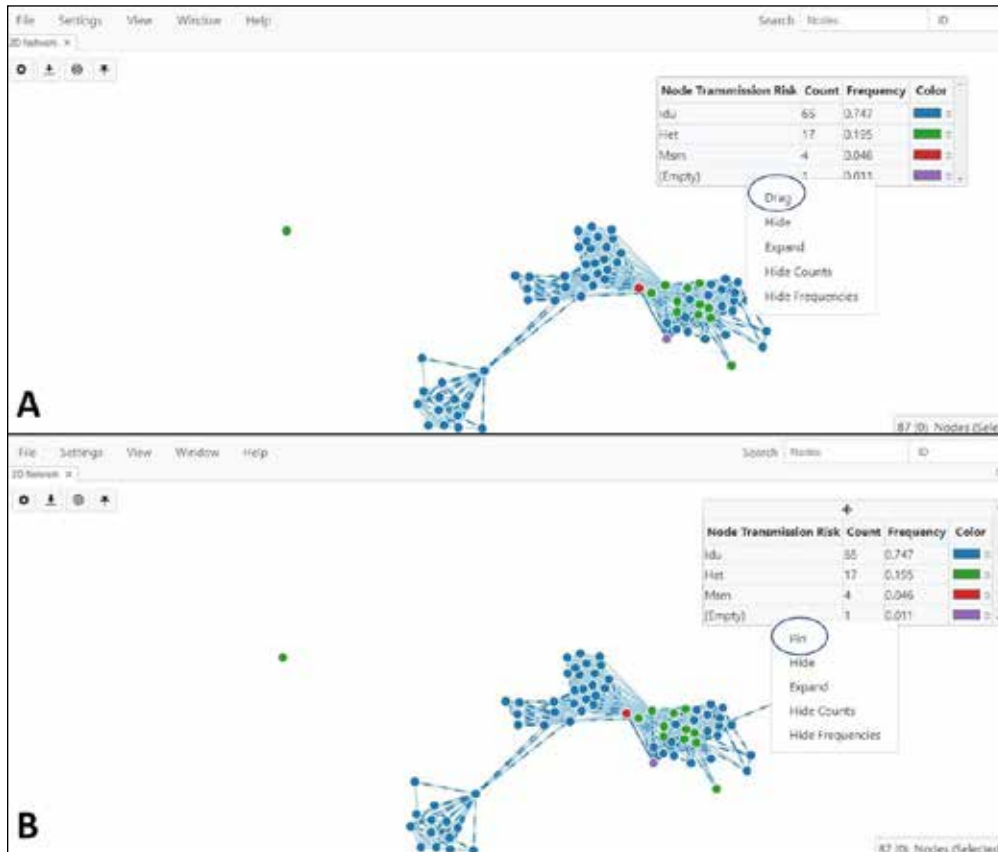
图 36. 编辑颜色键



此键将显示在您打开的每个新视图的右上角。您可以将它移动、隐藏和取消隐藏, 并通过右键点击框并选择拖动、固定、隐藏、展开、切换计数和切换频率来展开它以显示所有变量。一旦隐藏后, 您可以转到全球设置菜单 (Global Settings Menu), 并选择**显示所有内容 (Reveal Everything)**。切换允许您选择是否显示键中所有类别的节点计数和频率(图 37)显示两个选项。还可以使用键盘上的箭头键移动所有键框, 从而更好地控制和定位框。



图 37. 移动和扩展网络密钥箱/表。A. 选项包括拖动 (Drag)、隐藏 (Hide)、展开 (Expand)、切换计数 (Toggle Counts) 和切换频率 (Toggle Frequency)。B. 将框拖到您选择的位置后, 右键点击框会显示可用选项固定 (Pin)、隐藏 (Hide)、展开 (Expand)、隐藏计数 (Hide Counts) 和隐藏频率 (Hide Frequencies)。



## 链接属性

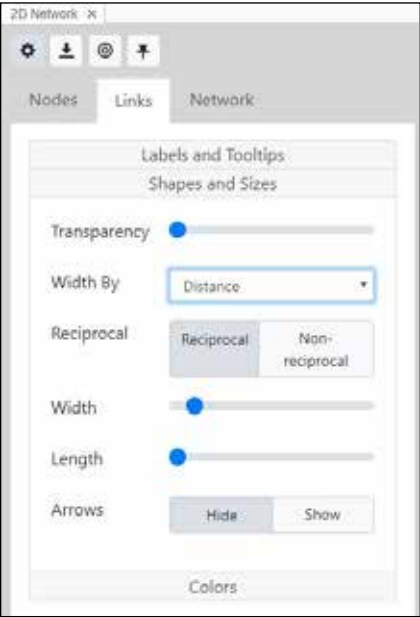
遗传链接或边通常使用定义的核苷酸距离截止值生成。切换网络设置菜单 (Toggle Network Settings Menu) 上的**链接 (Links)**选项卡允许您自定义链接颜色、宽度等, 并将这些属性映射到人口统计数据, 就像您可以使用节点一样 (图 38)。

**标签和工具提示:**这些是可定制的, 因为它们与节点一样。

**形状和大小:**您可以使用滑块来更改链接的透明度、宽度和长度。例如, 如果您的集群太紧 (非常密集), 您可以增加链接的长度 (拖动**长度 (Length)**选项上的滑块)。此选项允许集群结构变得更开放 (密度更低), 因此更容易查看集群节点和边。您还可以将宽度映射到链接列表中的任何变量。

**颜色:**选择**颜色**将带您进入全球设置菜单的样式选项卡, 如上面的节点部分所述。

图 38. 自定义链接属性和启用方向性的选项(箭头)

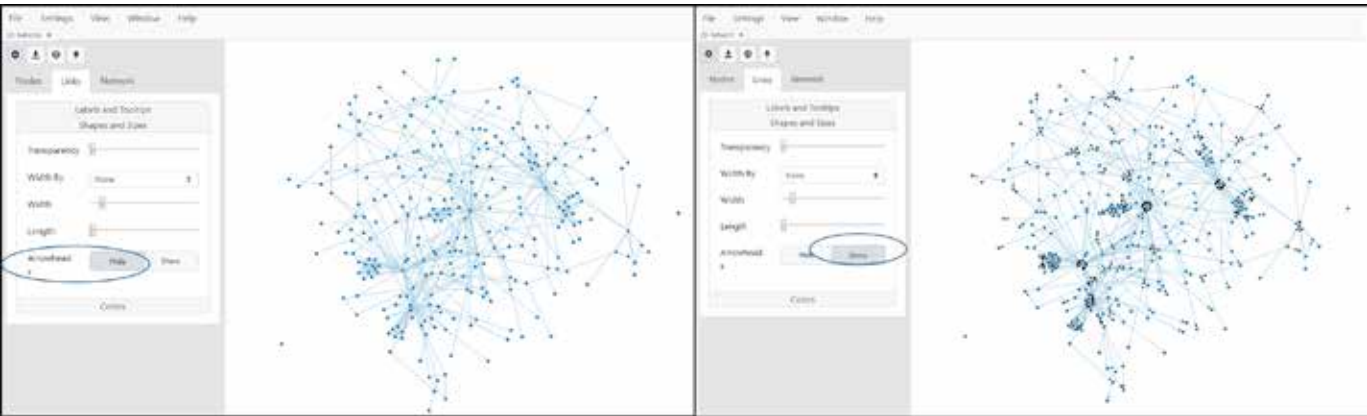


在网络中启用方向性(请阅读词汇表中关于推断传输方向性的重要注意事项)

方向性仅对包含接触者追踪数据的边列表有效, 而对仅包含序列数据的数据集无效。

HIV-1 分析的默认链接设置是链接是无方向性的。如果您上传包含接触者追踪信息的边列表, 那么您可以使用箭头(Arrowheads)按钮(图 39)来预测其方向性。在启用此功能之前, 请[参阅上面有关使用方向性](#), 探索 HIV 或其他病原体传播的限制的说明。

图 39. 包含遗传信息和接触者追踪数据的边列表的链接之间不带(左)和带(右)箭头的链接设置。

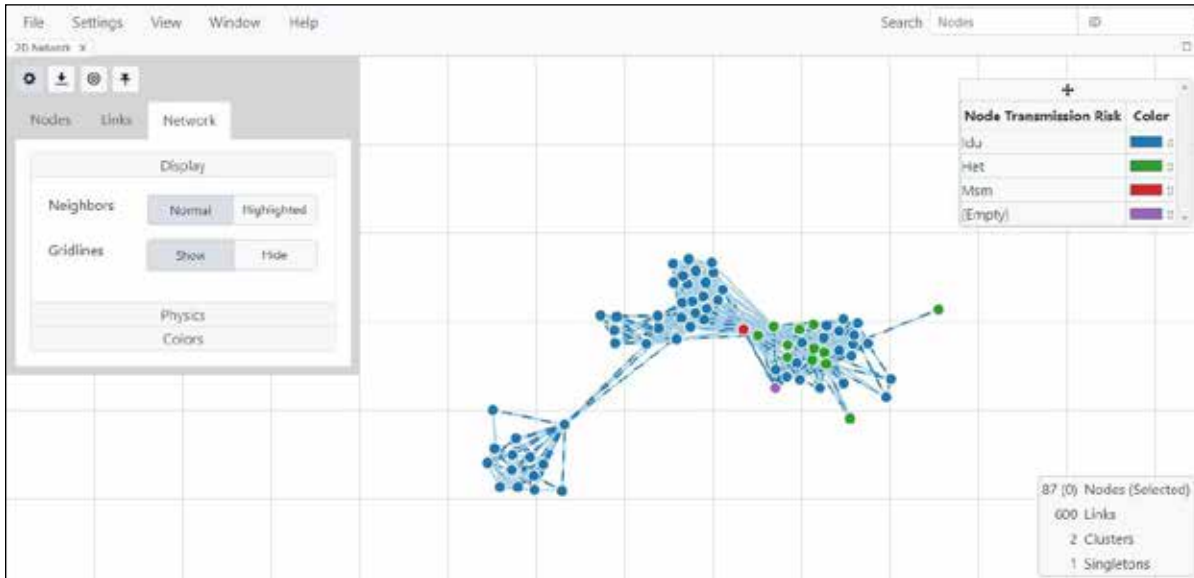


## 网络属性

**相邻:**使用此功能可以调整相邻节点的显示选项。当您鼠标悬停在特定节点上时, 选择**正常(Normal)**, 将会显示所有节点。然而, 如果您将此参数设置为**突出显示(Highlighted)**, 则将鼠标悬停在网络中特定的感兴趣节点上, 使其突出显示其所有相邻节点, 或链接到您关注节点的节点(图 40)。

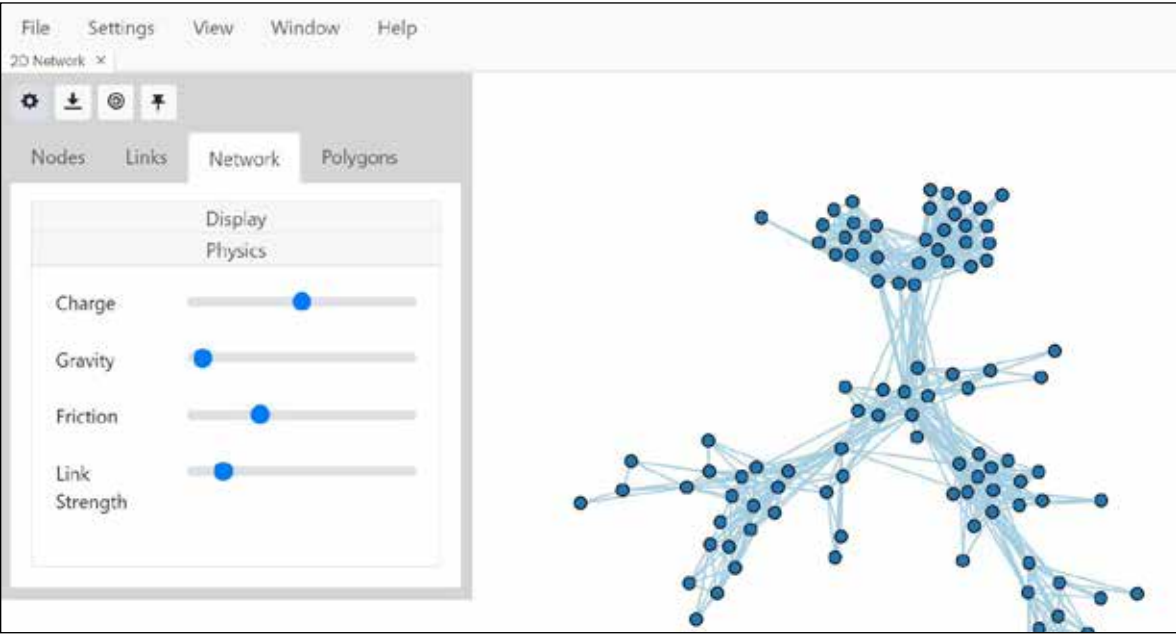
**网格线:**默认是网格线是隐藏的, 但您可以通过选择显示(Show)来选择显示它们(图 40)。网络实用程序使您能够以视觉和分析的定量方式比对节点和集群。它还可以帮助在后续分析中重现可视化。

图 40. 网络显示选项突出显示相邻节点;显示网格线。



**物理:**此按钮可让您更改网络的**摩擦力**、**电荷**和**重力**, 而这些参数决定了网络中节点的密集程度。通过点击上句中这些术语的突出显示链接, 了解在词汇表中有关这些属性的更多信息(图 41)。

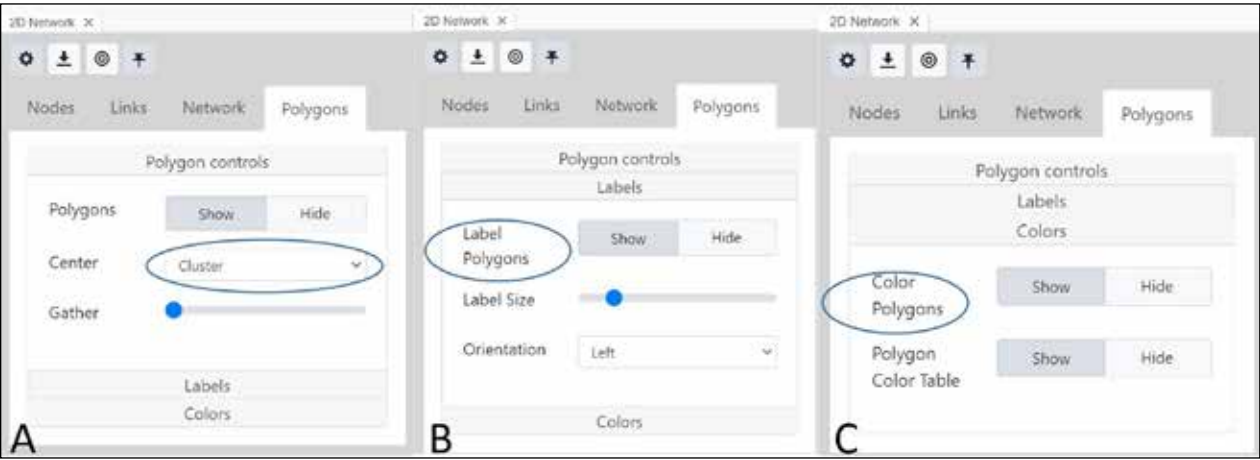
图 41. 用于更改网络物理属性的网络选项。



**颜色:**选择颜色将带您进入全球设置菜单的样式选项卡, 如上面的节点部分所述。

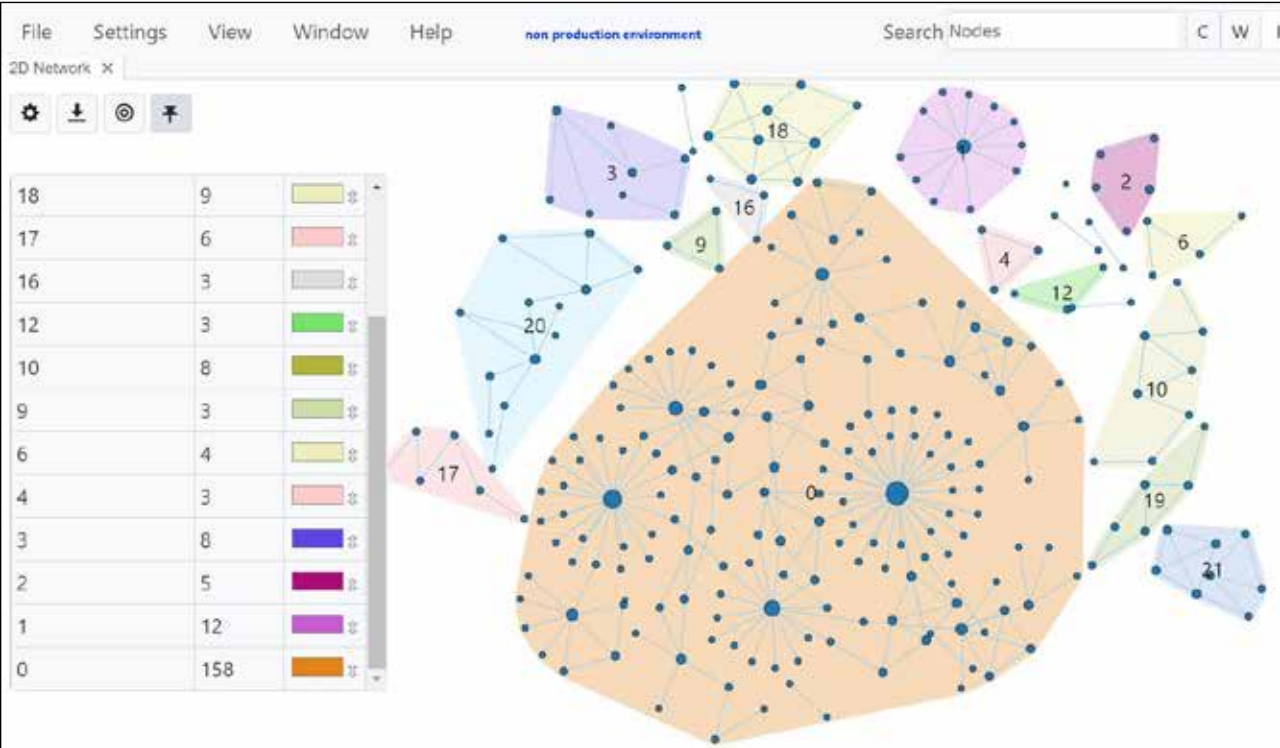
**多边形:**多边形选项卡允许您通过数据集中所选的任何节点属性对节点进行分组, 可通过下拉菜单进行选择(图 42A), MicrobeTrace 在组周围绘制彩色多边形。聚集滑块允许您调整多边形内节点的物理特性。增加聚集将每个多边形内的节点拉得更近。您还可以为多边形设置不同的颜色并进行标记(图 42B 和 C), 并能够通过键中的颜色图表更改多边形颜色和透明度。标签方向和大小也可以调整。

图 42. 多边形设置。多边形选项卡具有三个子选项卡 A. 多边形控件: 选择用于分组节点和调整聚集强度的属性。B. 标签: 选择标签大小和方向。C. 颜色: 修改多边形、颜色和透明度。



下图显示了一个多边形特征示例, 其中节点按集群分组, 在中心标记, 并调整透明度以清晰显示标签。

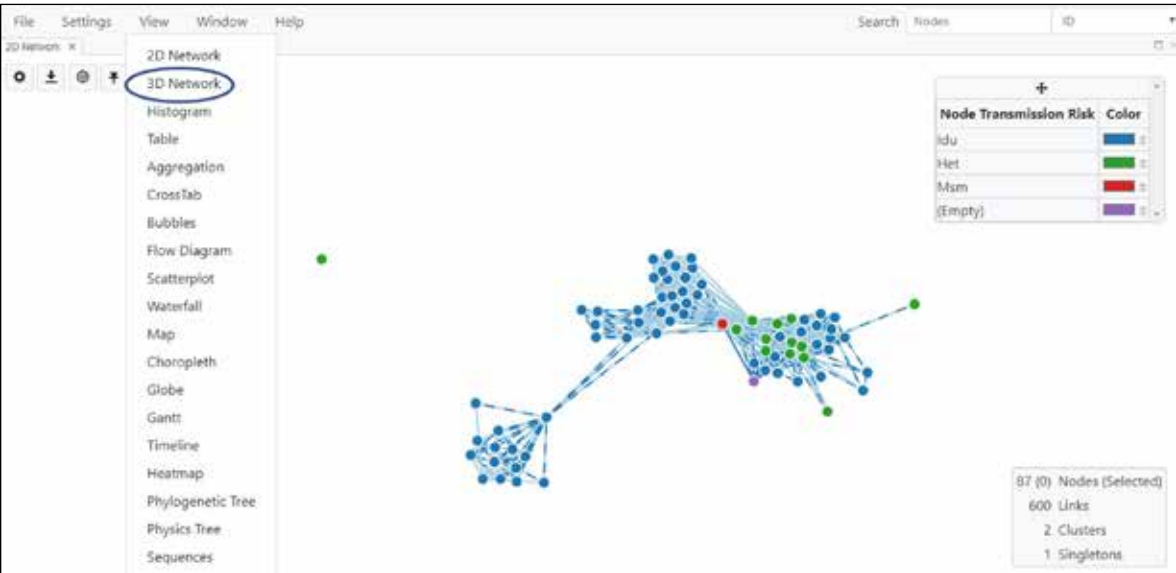
图 43. 节点按集群分组, 略微增强了聚集强度以分离多边形, 多边形被着色和标记。



### 3D 网络视图

您可以通过从下拉视图(View)菜单中选择 3D 网络来生成网络视图的 3D 版本(图 44)。

图 44. 选择 3D 网络视图







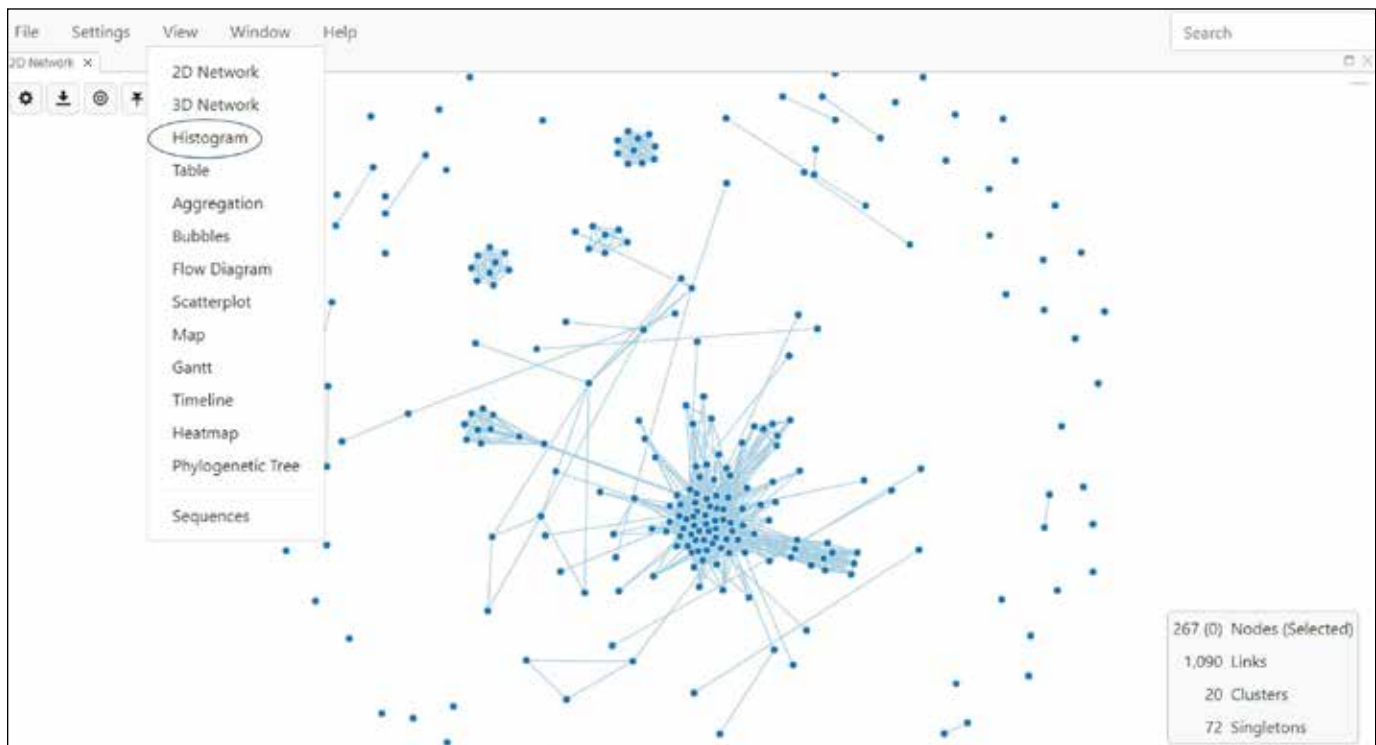
## 直方图视图


如果分析包括带有核苷酸序列的 FASTA 文件, 则使用 TN93 核苷酸替换模型计算的遗传距离结果可以看作是一个直方图。虽然查看序列中遗传距离的分布是直方图视图的主要用途, 它还可以与接触者追踪数据一起使用, 作为变量之间关系的替代可视化。

遗传距离直方图是一个条形图, 显示特定遗传距离在数据集中出现的频率。通常, 直方图中的频率分布呈现双峰(两个峰值)。一个峰值将包含非常密切相关的序列的遗传距离, 第二个将包含更远相关的序列。最能区分这两个峰的遗传距离可用于为您的特定分析细化遗传距离阈值选择。为方便起见, **全球设置菜单 (Global Settings Menu)** (通过选择 MicrobeTrace 主窗口上的设置访问) 也提供了此直方图的微型版本。您可以使用此功能更轻松确定数据的遗传距离截止值。

点击**视图 (View)**下的**直方图 (Histogram)**, 显示遗传距离直方图(图 47)。

图 47. 选择直方图视图



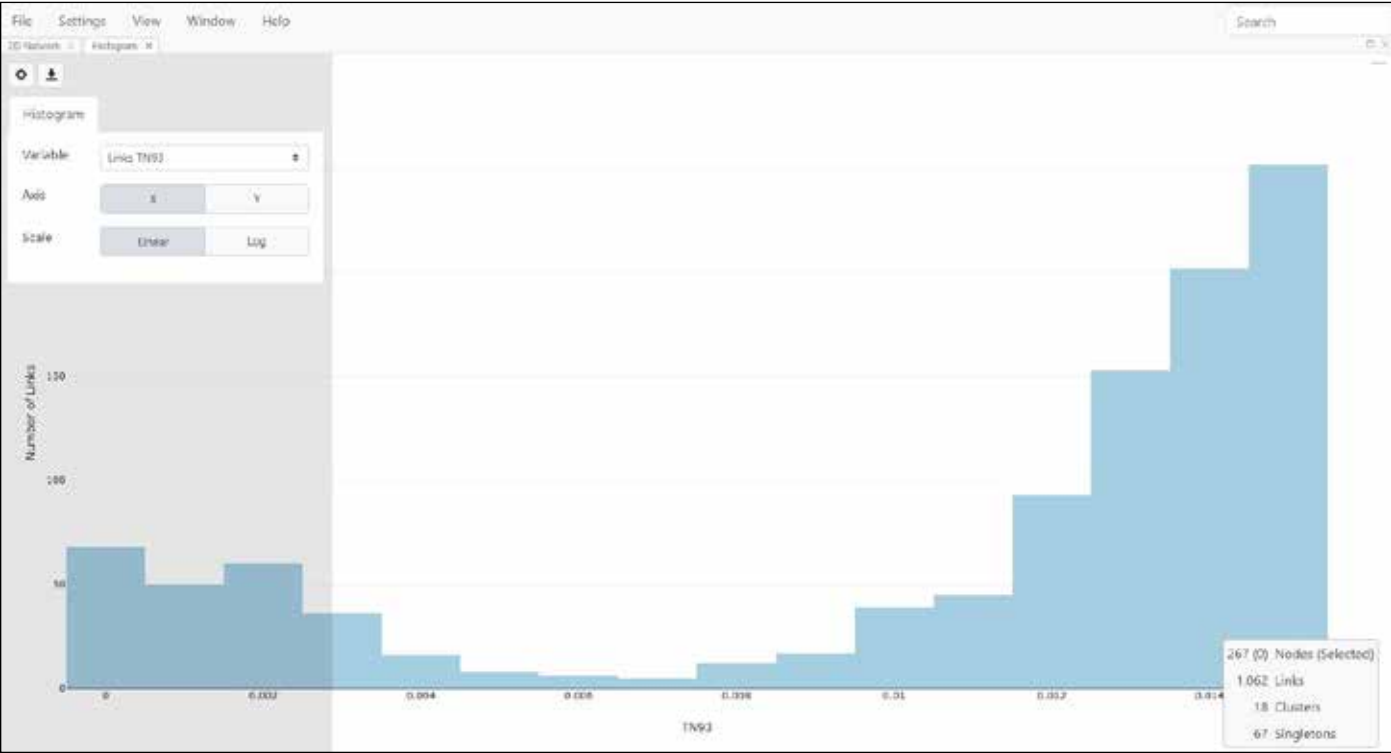
可以通过选择切换直方图设置 (Toggle Histogram Settings) 按钮来更改直方图设置 。从下拉菜单中选择相关变量, 从而选择用于配置直方图的链接类型(图 48)。

默认的直方图设置使用遗传距离作为变量, 显示所有链接之间距离的频率, 并使用线性比例来表示距离。当距离被选为绘图的距离时, 也可以通过在线性和对数尺度之间切换遗传距离或对数尺度绘制。如果您在 2D 视图中选择通过任何变量为节点或链接着色, 则这些相同的颜色将会显示在直方图视图中。



可以对双峰曲线上的距离进行检查, 以确定用于确定核苷酸序列连接的遗传距离截止值(详见词汇表中的遗传距离阈值)。如果您的特定分析需要使用不同的遗传距离截止值, 请转到网络视图(Network View)并在网络配置(Network Configuration)下的遗传距离阈值中输入此值。

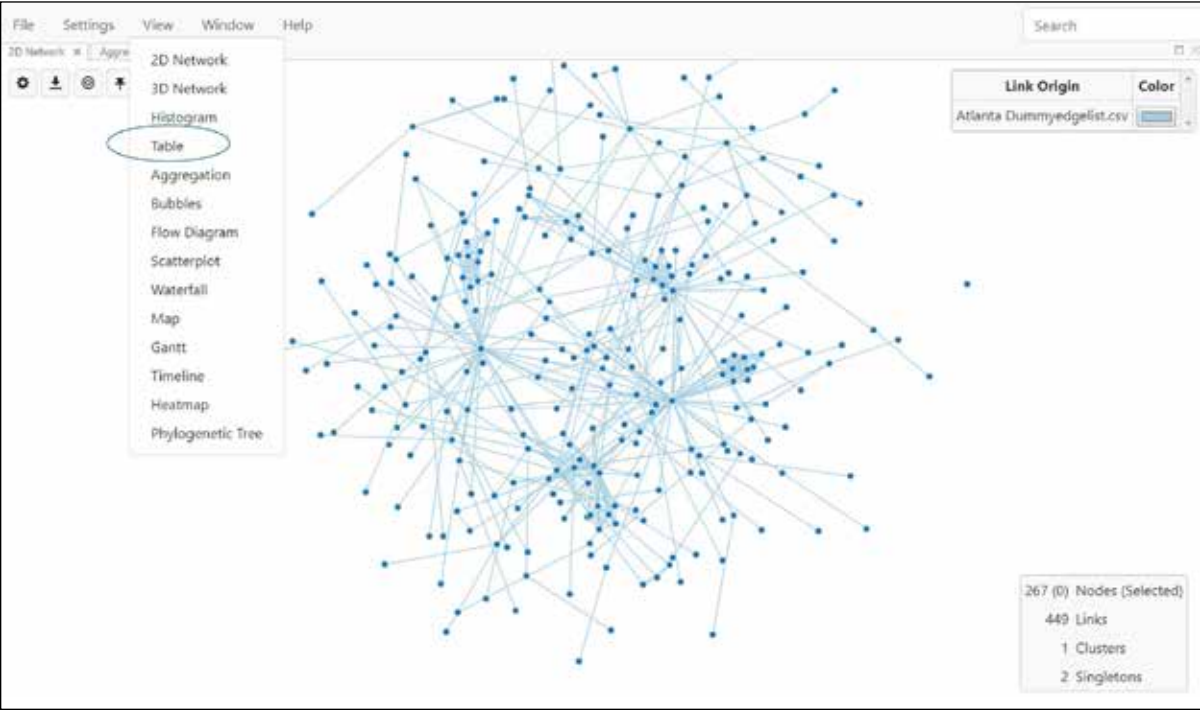
图 48. 使用 TN93 核苷酸替代模型计算的与遗传距离相关的链接数量的直方图视图。



# 表视图

表视图以表形式(如 Excel 工作表)显示与节点关联的数据。选择 **视图(View)**下的 **表(Table)** (图 49)。

图 49. 选择表视图



系统在新窗口中显示表(图 50)。

图 50. 表视图显示节点列表文件中的数据。

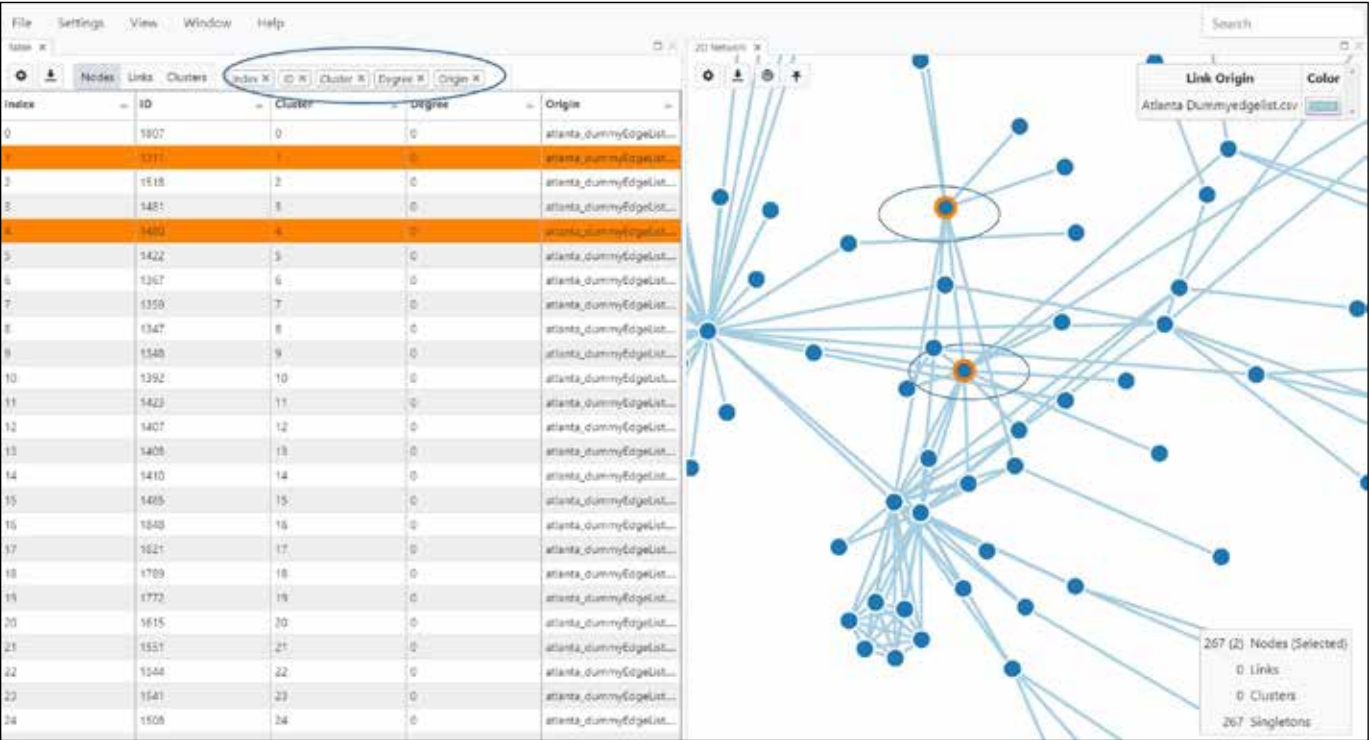
The screenshot shows the 'Table' view of the software. The 'Nodes' tab is selected, and the 'ID' column is highlighted. The table displays a list of nodes with their IDs, clusters, and subtypes.

Index	ID	Cluster	Subtype
11	10582_KF773581_H96c01	0	C
0	30582_KF773572_H96c05	0	C
1	30582_KF773570_H96c01	0	C
2	30582_KF773577_H96c10	0	C
3	30582_KF773573_H96c06	0	C
4	30582_KF773574_H96c07	0	C
5	30582_KF773575_H96c08	0	C
6	30582_KF773576_H96c09	0	C
7	30582_KF773578_H96c11	0	C
8	30582_KF773571_H96c02	0	C
9	30582_KF773579_H96c12	0	C
10	30582_KF773580_H96c13	0	C
12	30582_KF773583_H96c07	0	C
13	30582_KF773582_H96c04	0	C
14	30582_KF773584_H96c14	0	C
15	30582_KF773585_H96c12	0	C

搜索框允许您查找表数据中感兴趣的节点的信息。您可以通过 csv 文件中的任何属性来搜索节点, 这些属性作为搜索框右侧的下拉菜单提供(参见图 50 中圈出的部分)

如果您选择以表的格式突出显示特定节点的数据, 则可以通过在 2D 视图中按住 Ctrl 键来选择要包含在表视图中的任意数量的节点来选择节点。现在返回到表视图(Table View)窗口以查看所选节点的数据。为方便起见, MicrobeTrace 将选定的行置于表的顶部。您也可以点击并拖动网络视图(Network View)旁的表视图(Table View)选项卡, 以便两个窗口并排在同一个浏览器窗口中(图 51)。

图 51. 表视图(右)和 2D 网络视图(左)。选定的节点在网络中以橙色突出显示, 表视图中的相应数据行也是如此。选定的节点被带到表的顶部。




默认情况下, 表视图中仅显示四列。如果您想显示更多列, 请点击“空白”(在图 51 中圈出), 以从下拉菜单中选择更多列。或者, 您可以开始在“空白”中输入列名, MicrobeTrace 将自动尝试猜测您的选择, 并像拼写检查一样完成单词。您还可以通过在列名框中选择“x”从表视图中删除列。

点击表中的列标题将根据该列中的数据对所有行进行排序。选择表中的行将突出显示网络中的相应节点以及地图视图中链接的地图数据(Map View) (但仅当地理数据包含在您的节点列表数据文件中时)。可以对表中的所有文本条目执行搜索。您可以使用 ID 或其他标识符搜索特定节点, 在表中选择它, 然后在网络视图(Network View)中查看相应突出显示的节点(节点以橙色突出显示), 以便查看链接和集群或网络位置。请注意, 如果您选择不显示单例(使用全球设置过滤(Global Settings Filtering)选项卡), 并且您在表视图中选择了一个恰好是单例的节点, 则单例将不会显示在网络视图(Network View)中。

默认表视图 (Table View) 设置在表列中显示节点数据。您可以使用表上方、设置按钮右侧的按钮在查看表中的链接或集群数据之间进行切换。链接 (Link) 和集群视图 (Cluster View) 具有相同的功能, 包括选择 ID、排序、选择要显示的列等。

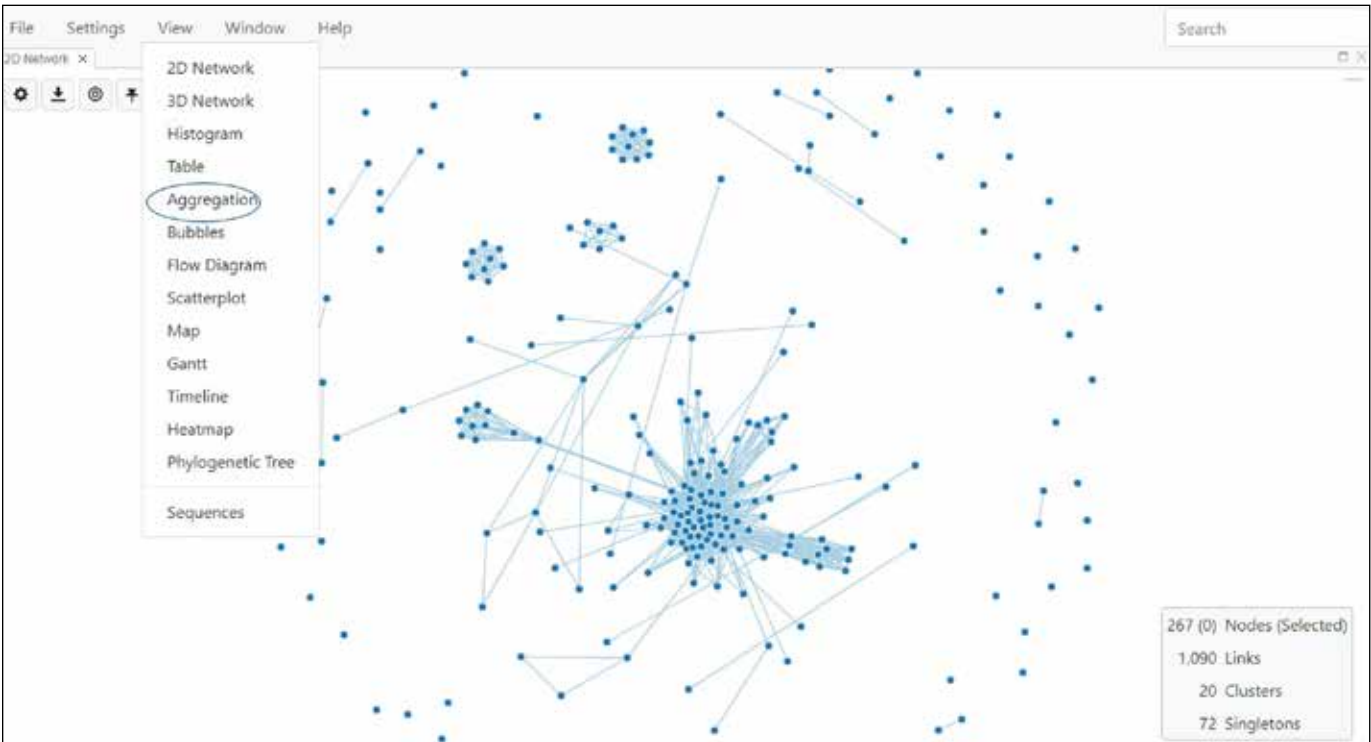
## 表设置

您可以使用切换表设置 (Toggle Table Settings) 按钮上提供的滑动比例选项来调整表视图中的文本大小 

## 聚合视图

聚合视图用于汇总节点数据中的变量。聚合视图提供了一个包含两列的表: 您要合计变量的值, 以及具有这些值的实体总数。此功能对于生成表或报告特别有用。要使用此功能, 请从视图 (View) 菜单中选择聚合 (Aggregation) (图 52)。

图 52. 选择聚合视图




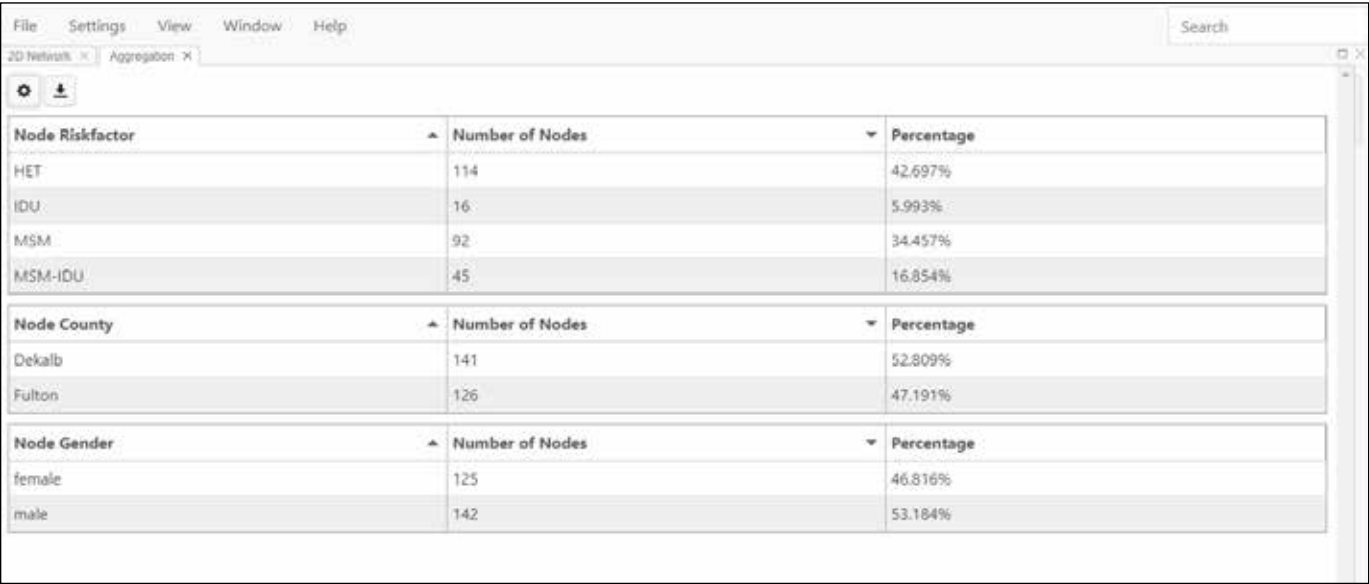
在下面的示例中, 当您第一次加载聚合表时, 变量的默认设置是**集群 (Cluster)**。  
在第二列提供了每个集群中的节点数, 在第三列列出了每个集群中节点的百分比。选择切换聚合设置 (Toggle Aggregation Settings)  按钮 (该按钮将打开设置菜单), 以自定义您想要比较节点、链接或集群数据的变量。



图 55. 三个变量的聚合表, 显示每个类别中节点的数量和百分比。



The screenshot shows the 'Aggregation' window with three tables. Each table has columns for the variable name, 'Number of Nodes', and 'Percentage'.

Node Riskfactor	Number of Nodes	Percentage
HET	114	42.697%
IDU	16	5.993%
MSM	92	34.457%
MSM-IDU	45	16.854%

Node County	Number of Nodes	Percentage
Dekalb	141	52.809%
Fulton	126	47.191%

Node Gender	Number of Nodes	Percentage
female	125	46.816%
male	142	53.184%


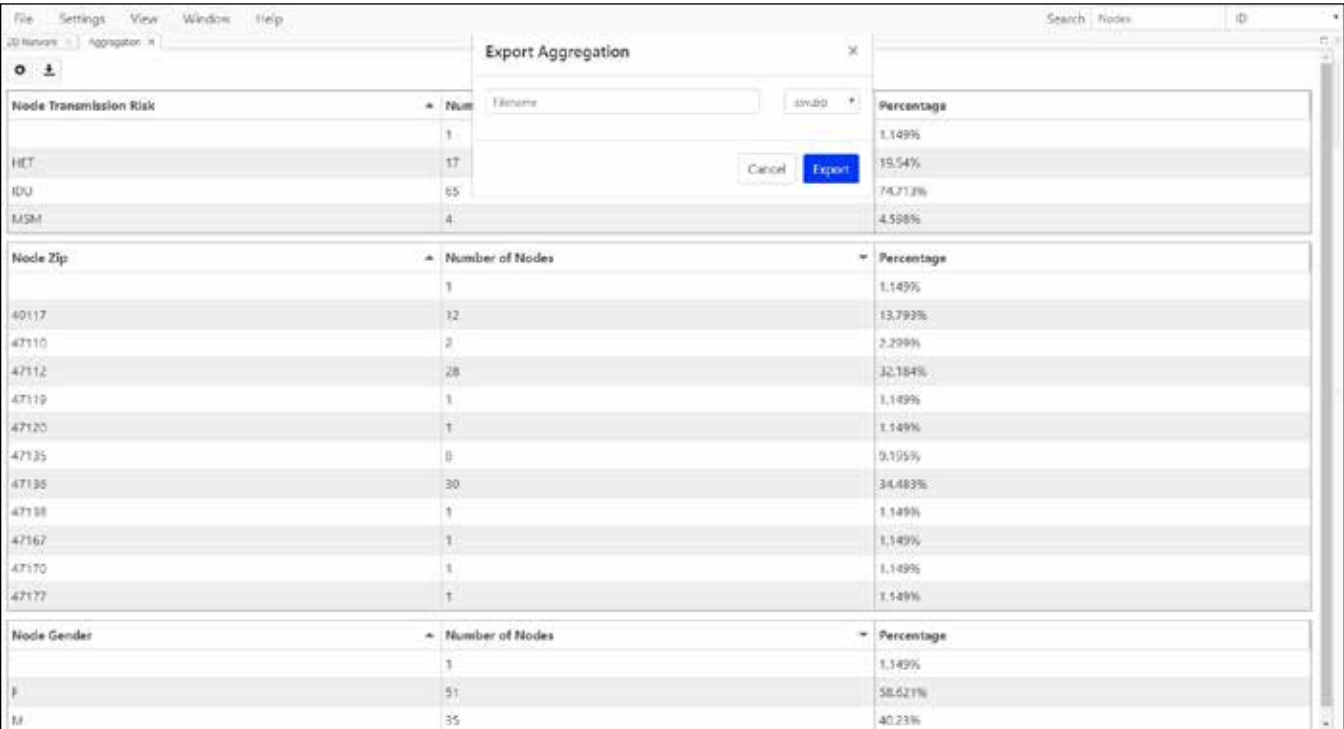
然后可以以您选择的格式导出此视图。选择下载按钮  , 从下拉菜单中选择文件格式。  
在下图中(图 56), 选择了 .xlsx 格式。这会将您的表以 Excel 文件导出, 每个表位于单独的工作表中。

图 56. 导出聚合表



The screenshot shows the 'Aggregation' window with the 'Export Aggregation' dialog box open. The dialog box has a 'File name' field, a 'Format' dropdown menu set to 'xlsx', and 'Cancel' and 'Export' buttons. The background table shows data for 'Node Transmission Risk'.

Node Transmission Risk	Number of Nodes	Percentage
	1	1.149%
HET	17	19.54%
IDU	65	74.712%
MSM	4	4.598%

Node Zip	Number of Nodes	Percentage
	1	1.149%
40117	12	13.793%
47110	2	2.299%
47112	28	32.184%
47119	1	1.149%
47120	1	1.149%
47135	0	0.195%
47136	30	34.483%
47138	1	1.149%
47167	1	1.149%
47170	1	1.149%
47177	1	1.149%

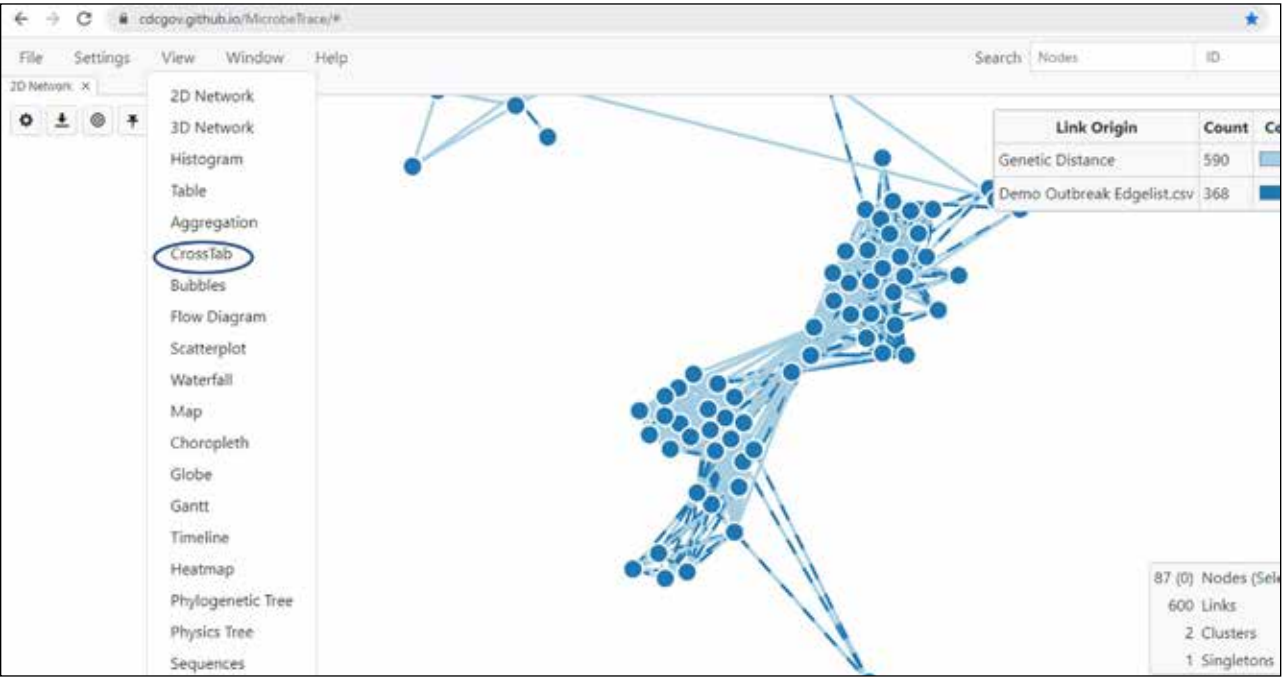
Node Gender	Number of Nodes	Percentage
	1	1.149%
F	51	58.621%
M	35	40.23%



# 交叉表视图

此视图允许对数据进行双变量交叉制表, 可用于评估可能不明显的分类变量之间的关系。它通常用于评估对调查问题的多重回答, 通常是在合作伙伴服务访谈期间收集的。顶部打开, 从下拉视图 (View) 菜单中选择交叉表视图 (Cross Tab)。

图 57. 选择交叉表视图



然后, 使用设置按钮选择您想要交叉制表的实体 (节点、链接或集群)。接下来, 从 X 和 Y 变量下拉菜单中选择您想要评估的两个分类属性 (图 58)。两个选定变量的横截面的结果计数将显示为表。此表可以导出为 CSV、XLSX、PDF 或 JSON 文件。

图 58. 跨使用频率的主要药物交叉制表

	Methamphetamine	Oxycontin	Marijuana	Cocaine	Heroin	(Empty)
frequently	5	16	7	7	7	0
occasionally	5	12	1	4	8	0
rarely	1	2	3	7	1	0
	0	0	0	0	0	1

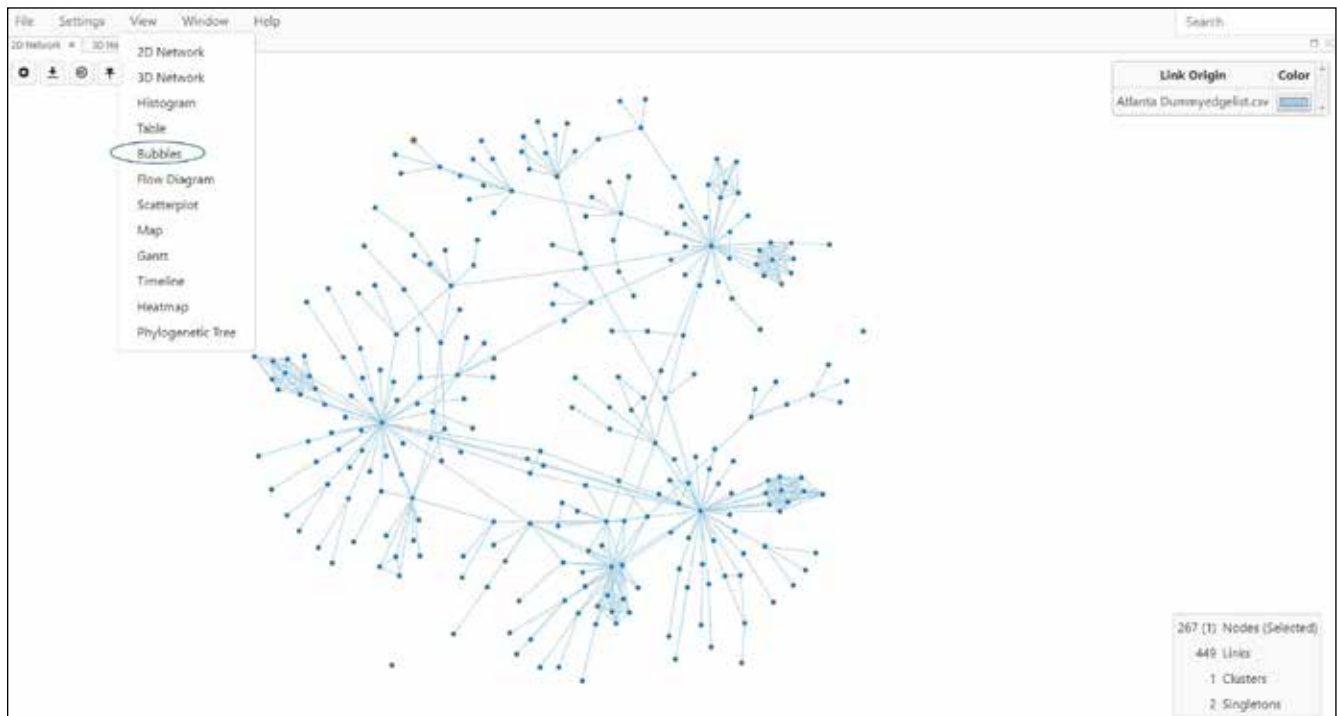


## 气泡视图

当您不想显示网络中节点之间的链接时, 气泡视图(Bubbles View)用于可视化网络内的人口统计分布。例如, 当信息未知或不可用时, 显示节点或人之间的联系可能会被误解为传输方向性(Oster 等人, Barré-Sinoussi 等人)。在将网络数据作为图像共享时, 删除显示的链接有助于消除任何不必要的混淆。

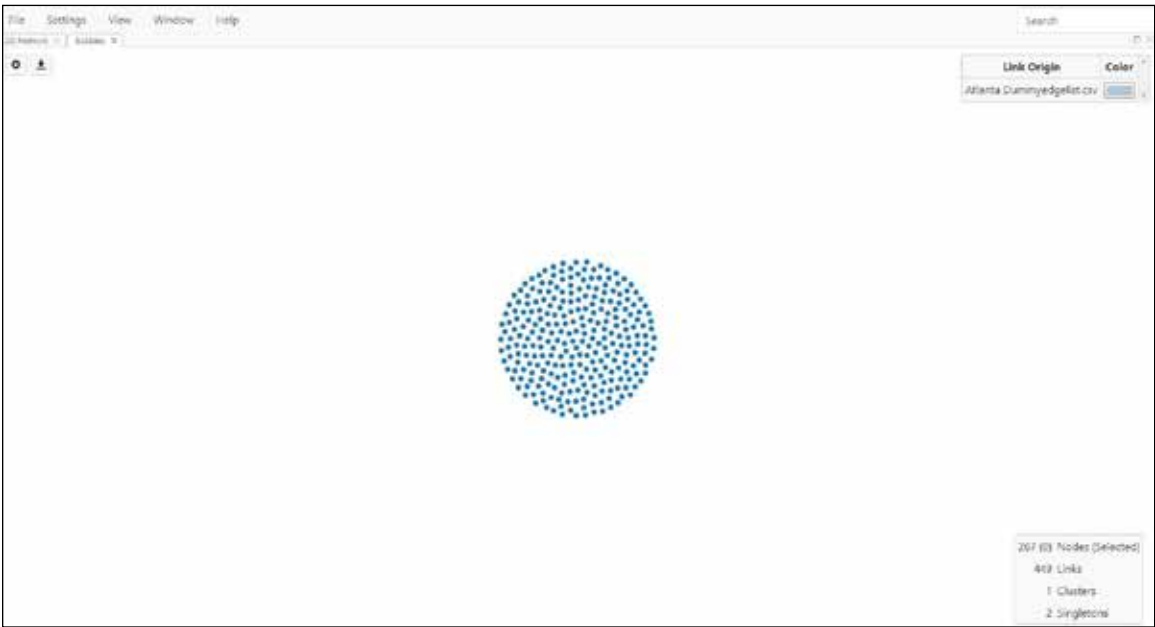
气泡视图(Bubbles View)是 MicrobeTrace 会话中所有节点的强制导向图, 并无节点之间的链接。不同于网络视图(Network View), 您不能用鼠标移动单个节点。但是, 根据数据集中的变量, 您可以在屏幕上移动它们, 以使用相似的节点进行聚类。要选择此视图, 请从下拉菜单中选择气泡(Bubble)(图 59)。

图 59. 选择气泡视图



这将打开仅显示数据中节点的默认视图(图 60)。

图 60. 默认的气泡视图

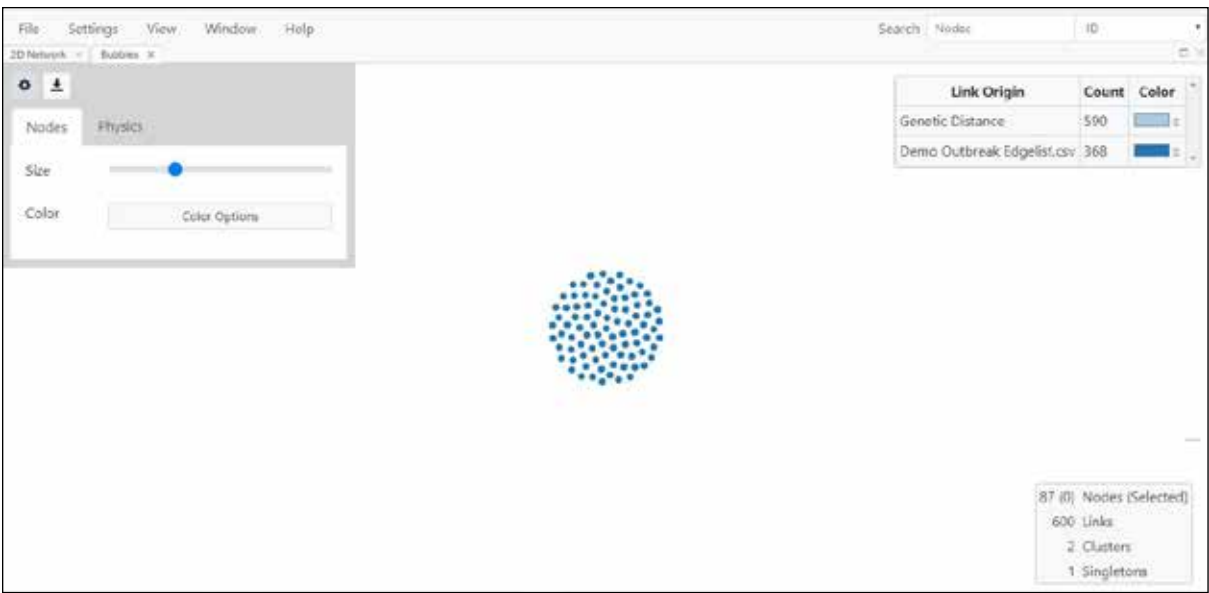


## 气泡视图设置

为了通过数据中的变量探索聚类, 使用**切换气泡设置(Toggle Bubble Settings)**按钮打开设置菜单。这将打开两个选项卡 — **节点和物理(Nodes and Physics)**(图 61)。

**节点**选项卡允许您使用滑块更改节点的大小。与其他视图一样, 您可以使用**颜色选项(Color Options)**按钮将节点颜色映射到您选择的变量。如果您已经在 2D 网络中映射了节点颜色, 则该颜色映射将转移到其他视图, 包括气泡视图(Bubble View)。

图 61. 气泡视图的切换设置



物理(Physics)选项卡(图 62)允许您将节点映射到 X 轴和 Y 轴上的变量。在下面的示例中, 我们正在研究传播风险因素与性别之间的关系。

图 62. 显示风险因素与传输集群中节点县的气泡视图。  
A. 从下拉菜单中选择变量(已圈出); B. 风险因素与县。HET - 异性恋者, IDU - 静脉吸毒者, MSM - 男男性行为者。

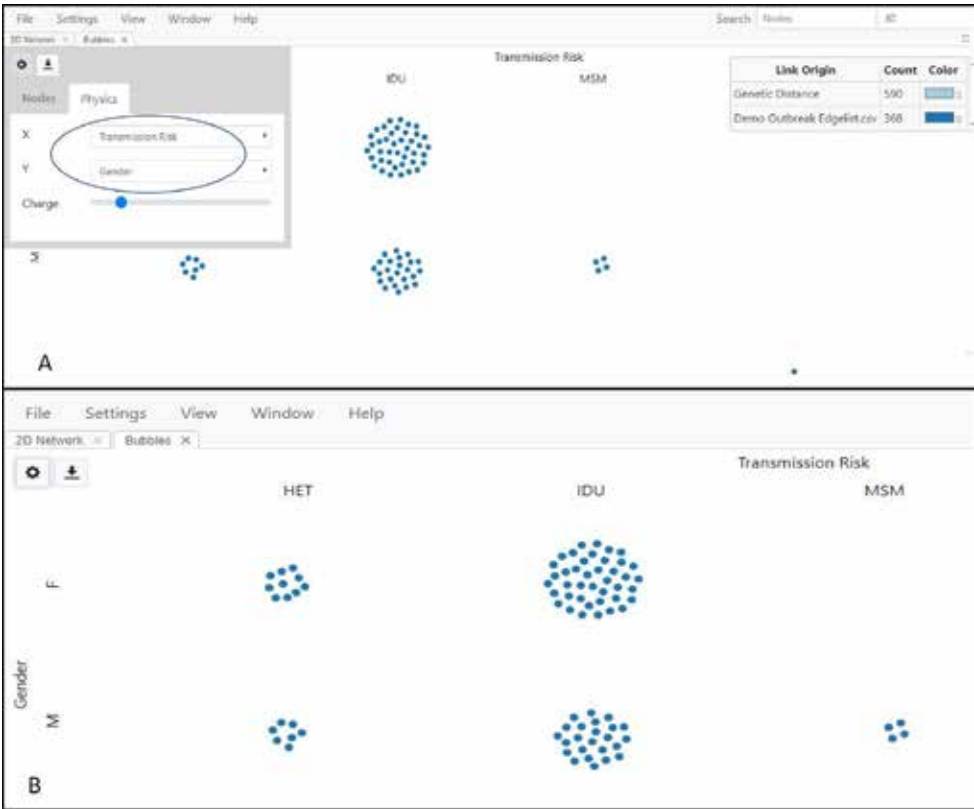
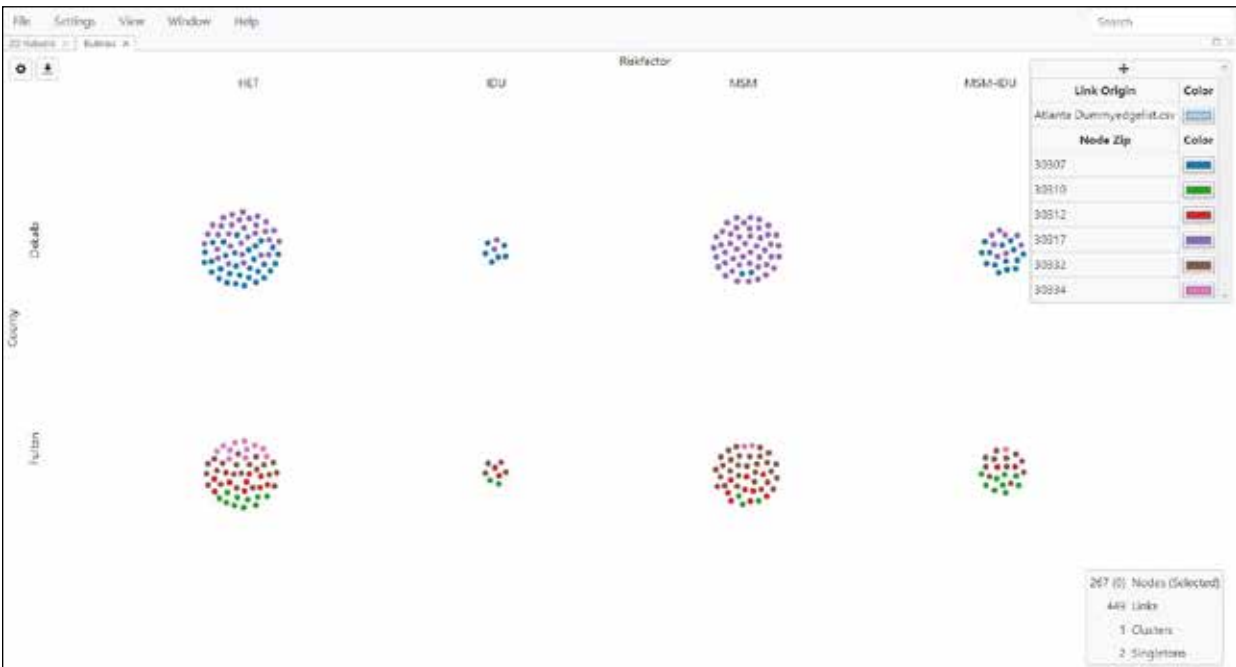


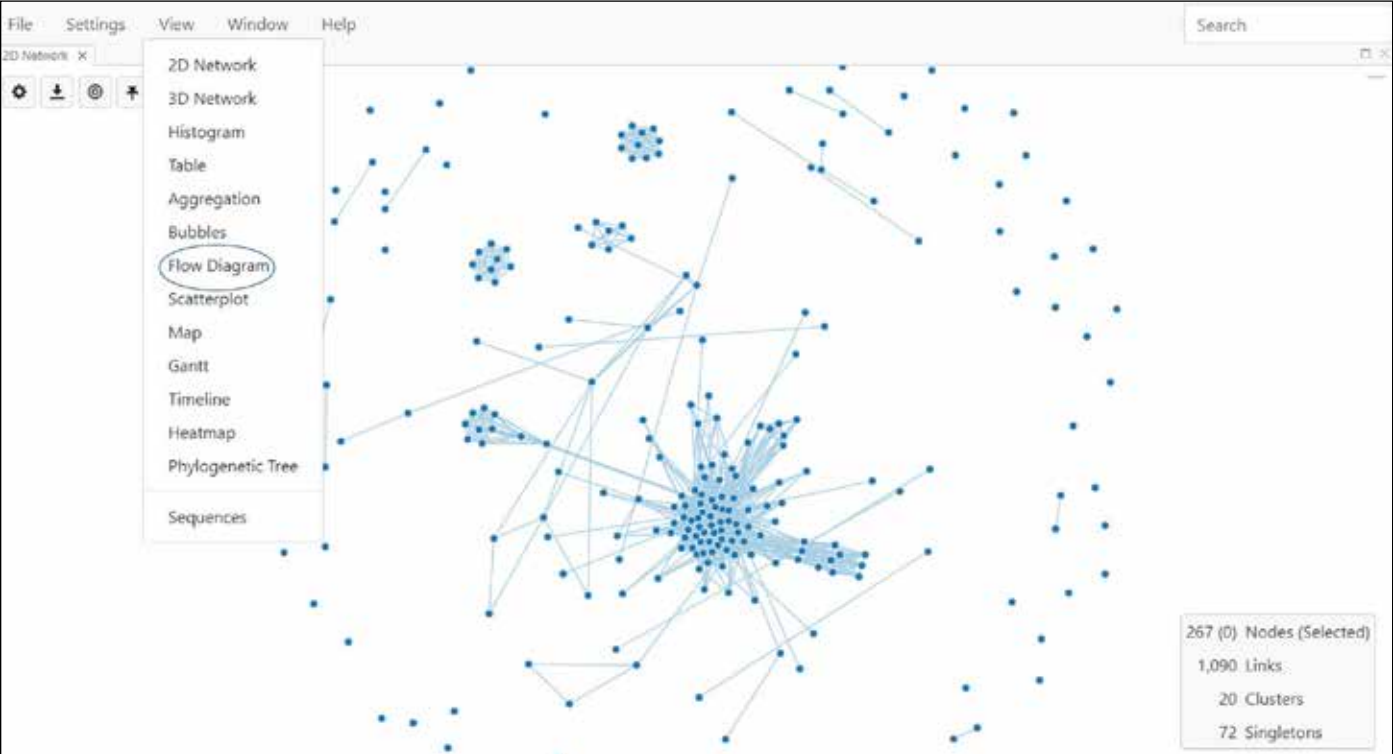
图 63. 气泡视图显示风险因素与县(乔治亚州亚特兰大的 Dekalb 和 Fulton), 节点按邮政编码着色。



# 流程图视图

流程图视图 (Flow Diagram View) 以流程图的形式 (特别是冲积图 (alluvial) 或桑基图 (Sankey diagram)) 将数据可视化, 并允许比较数据集中的变量。流程图可以从任何数据变量生成, 因为它们完全依赖于节点特征。从视图 (View) 菜单中选择流程图 (Flow Diagram) (图 64)。

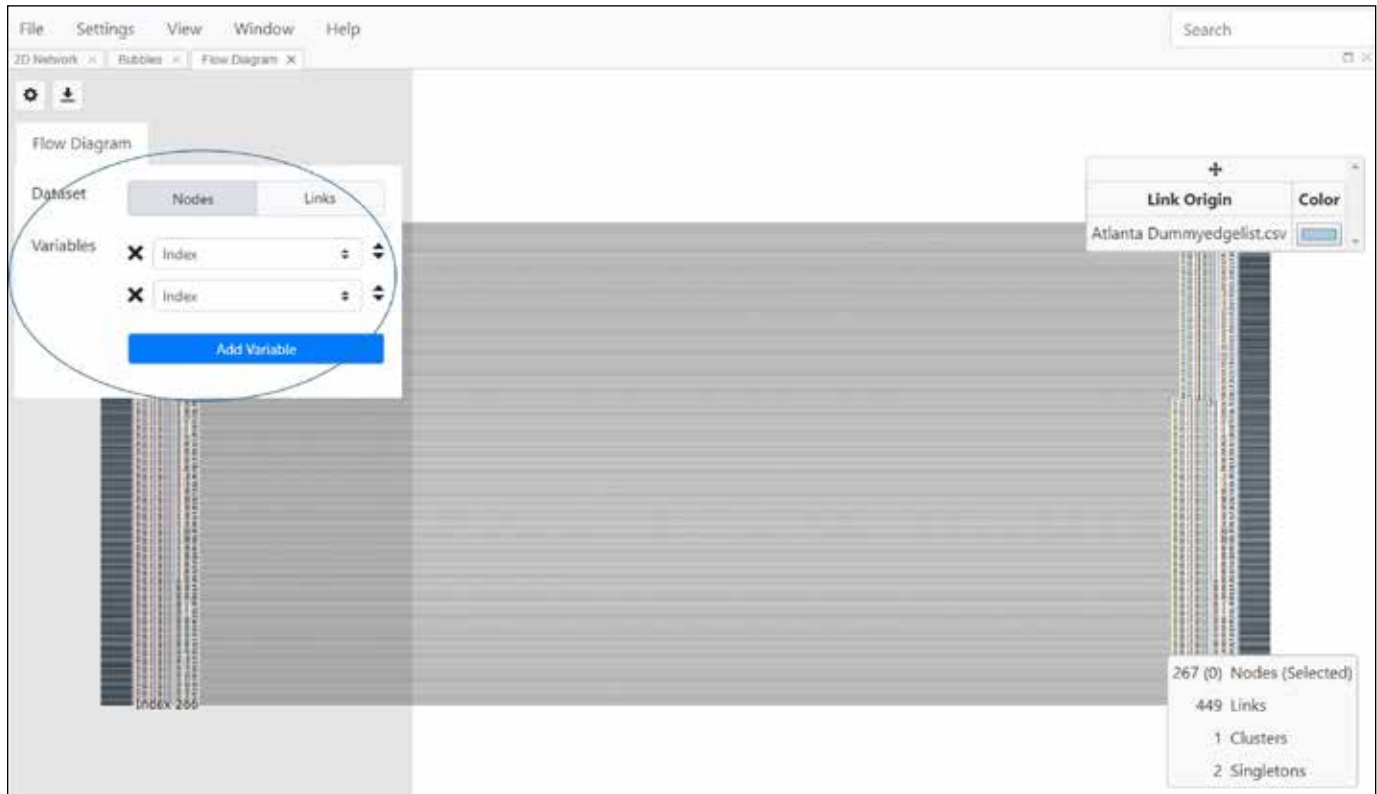
图 64. 选择流程图视图



## 流程图视图设置

选择后, MicrobeTrace 在新窗口中显示流程图视图(Flow Diagram View)(图 65)。

图 65. 映射到数据集中特定变量之前的流程图视图。



流程图视图(Flow Diagram View)默认设置为索引,所以在使用**切换流程图(Toggle Flow Diagram)**打开的流程图(Flow Diagram)对话框中选择变量之前,不会显示流程图。使用流程图视图(Flow Diagram View)浏览数据至少需要两个变量。使用下拉菜单选择要在流程图中可视化的两个变量。您还可以使用添加变量(Add Variable)按钮添加更多变量。在此示例中,检查了感染风险因素、县和邮政编码。图 66 显示了风险因素与县和邮政编码的映射。

图 66. 流程图视图 (Flow Diagram View) 显示了跨邮政编码和县 (乔治亚州亚特兰大的 Fulton 和 DeKalb) 的风险因素比较。

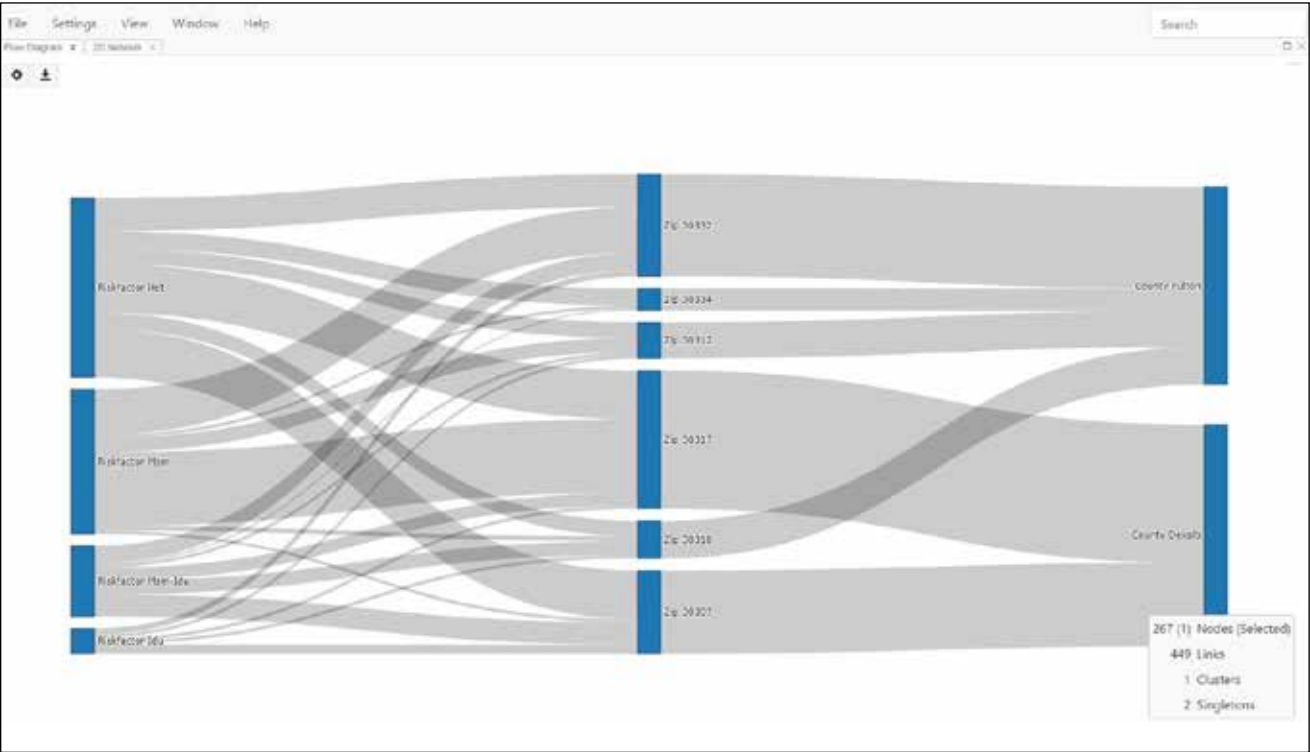
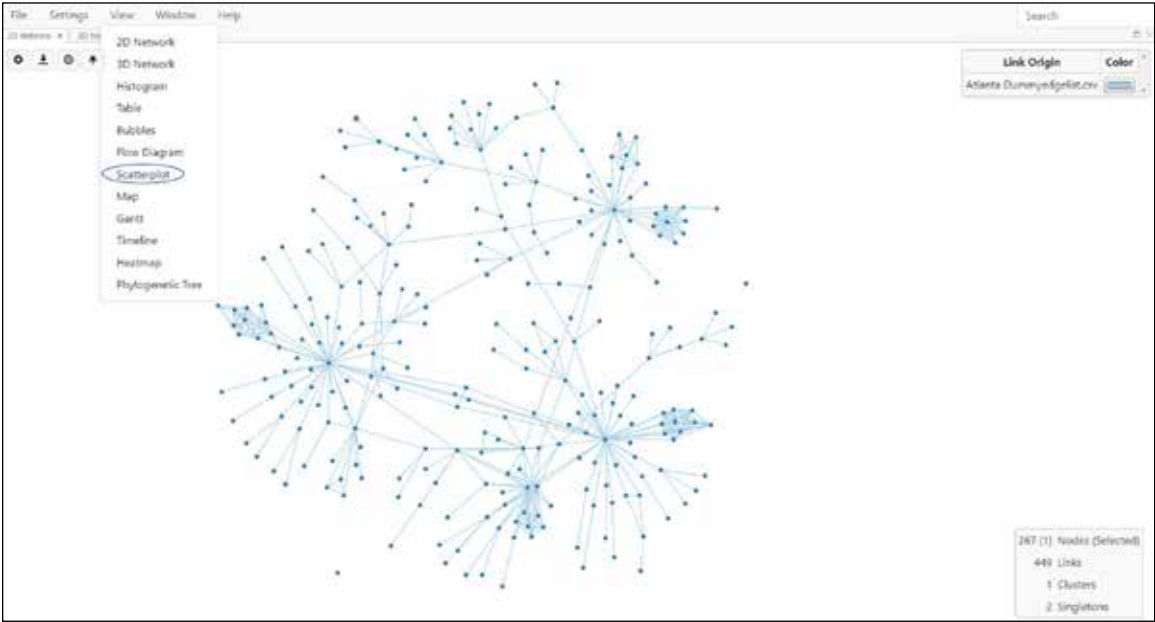


图 66 流程图中的蓝色块表示所选变量的相对流行度, 块之间的灰色流场(可变厚度的曲线)表示选定变量之间的关联。

# 散点图视图

散点图 (Scatterplot View) 视图提供数据中数值变量相关性的可视化。从视图 (View) 菜单中选择散点图 (Scatterplot) (图 67)。

图 67. 选择散点图视图




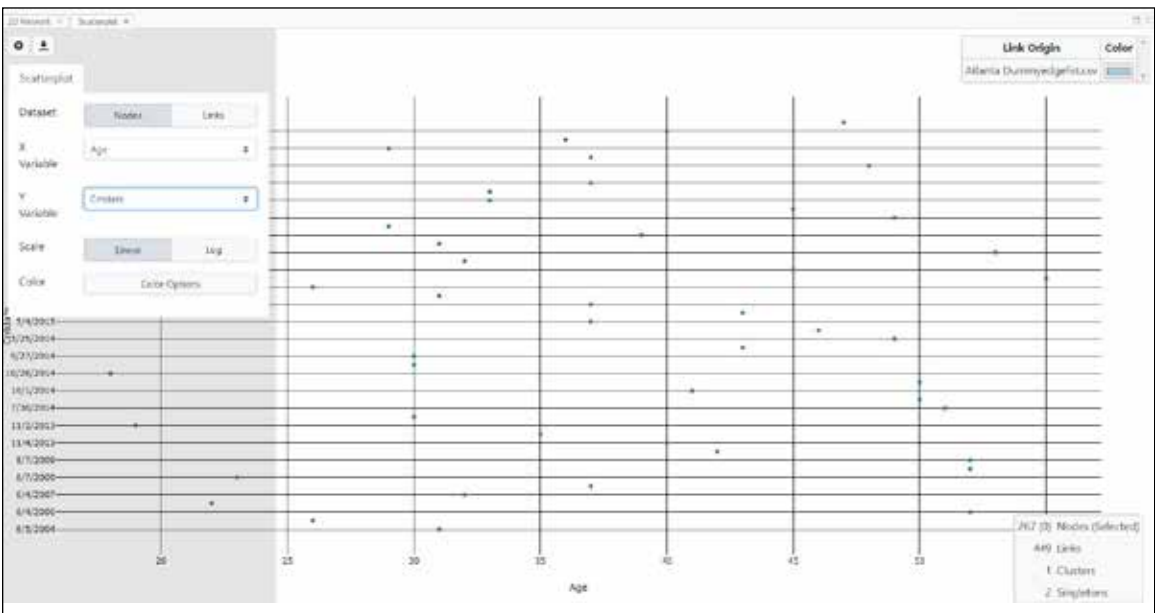
MicrobeTrace 在新选项卡中显示散点图。默认情况下, 它会按照 SNP 绘制 TN93 距离。但是, 您可以使用切换散点图 (Toggle Scatterplot) 按钮  打开一个对话框, 使用下拉菜单选择要在 X 轴和 Y 轴上进行比较的变量 (图 68)。在数据集 (Dataset) 选项中选择节点或链接以选择要比较的变量。您还可以将数字刻度调整为对数或线性。选择颜色选项 (Color Options) 会将您带到全球设置 (Global Settings) 菜单, 您可以在该菜单和其他视图中自定义颜色。

图 68. 散点图 (Scatterplot View) 查看和调整显示选项

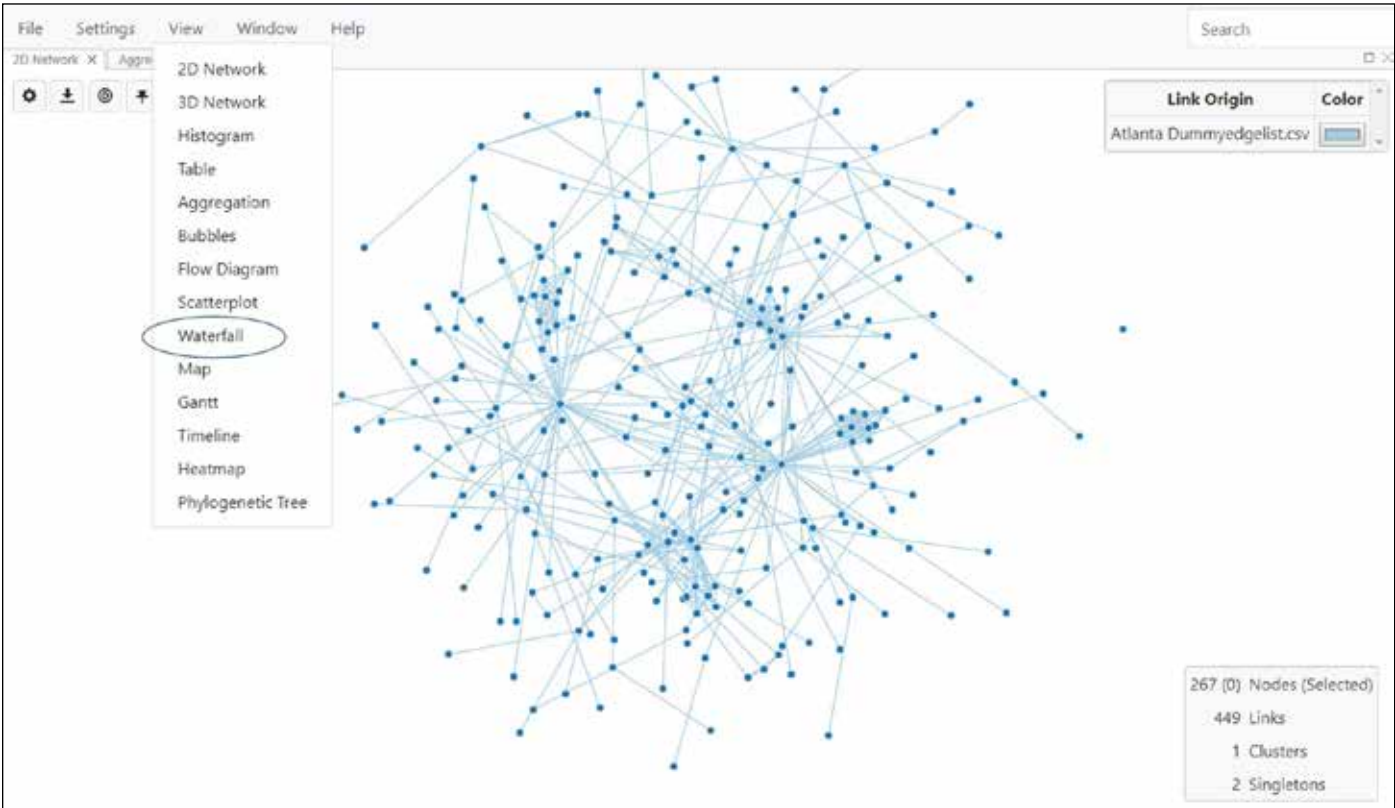




# 瀑布视图

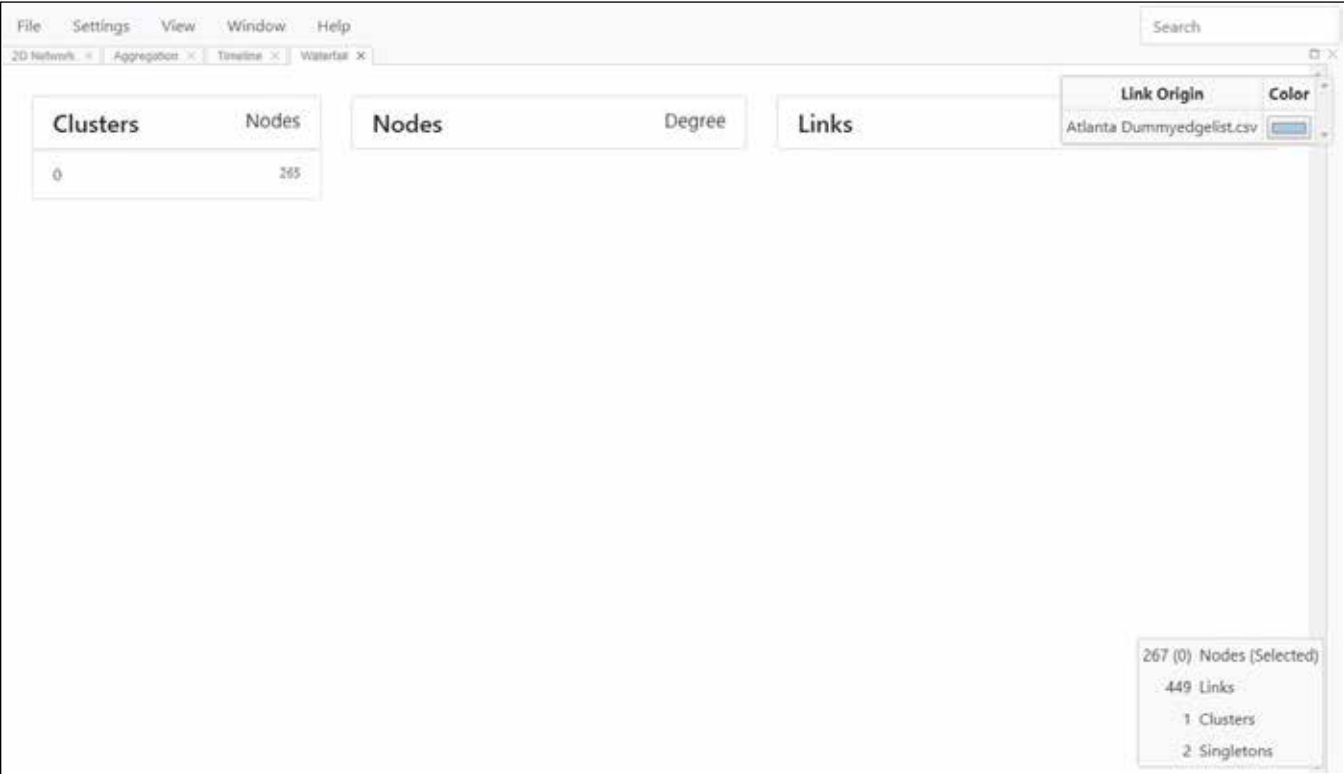
此视图以表式而非网络格式显示集群、节点和链接数据。从列表选择一个集群会显示该集群的节点和链接信息, 然后您可以依次从列表中选择节点以查看链接信息。此视图以文本格式(而非以网络为基础的可视性视图)提供有关这三个变量之间关系的详细信息。要选择此视图, 请从下拉视图菜单中选择瀑布视图(Waterfall View)(图 69)。

图 69. 选择瀑布视图



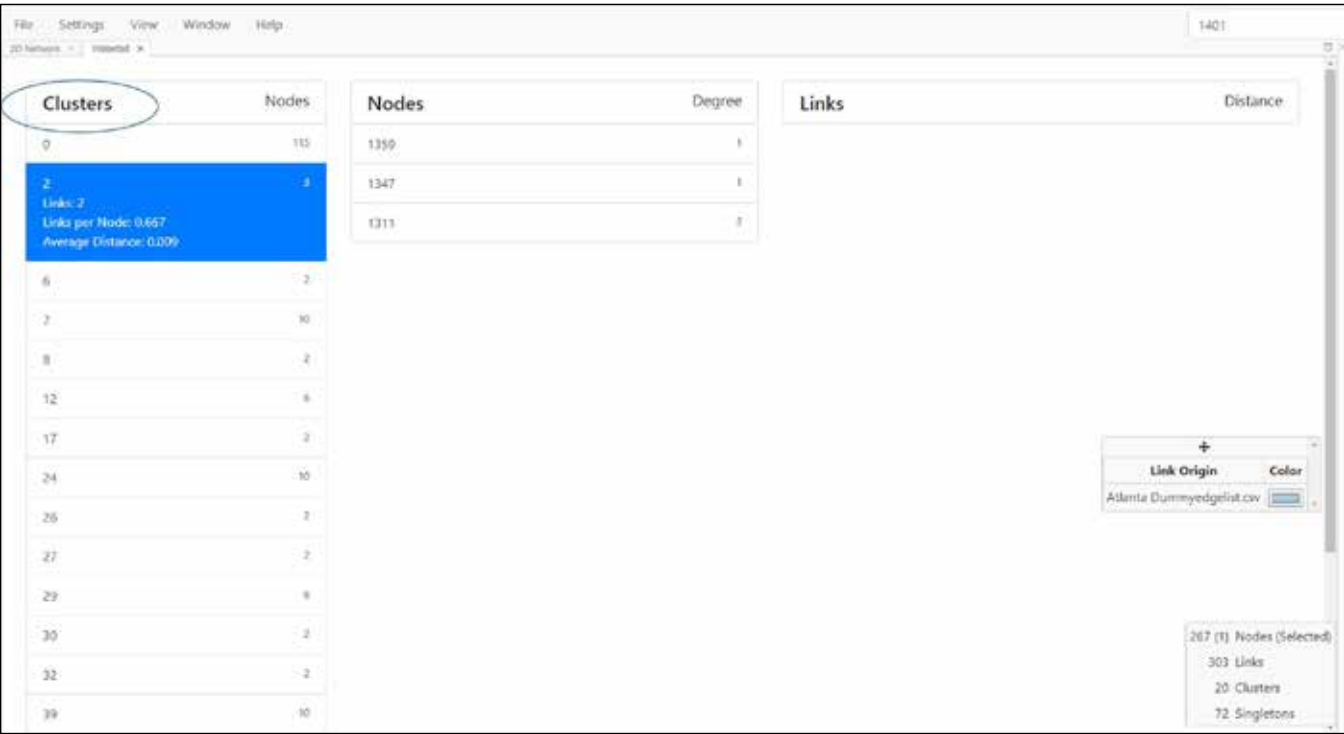
瀑布视图显示三列:集群、节点和链接各一列(图 70)。

图 70. 瀑布视图, 打开屏幕



点击特定集群, 然后列出该集群中的节点以及节点度(图 71)。

图 71. 通过点击感兴趣的集群来显示节点列表的瀑布视图。



**图 72. 以下拉/文本格式查看节点详细信息。**



图 73. 显示链接详细信息的瀑布视图



# 地图视图

如果地理坐标包含在节点文件中, 则数据中的节点可以使用地图视图(Map View)显示在全球性地理地图中。该地图允许您缩放到节点列表中指示的地理位置。尽管纬度和经度为您提供了最精确的位置, 但如果纬度经度数据不可用, 邮政编码和其他地缘政治分界线(县和州)也会显示在地图上。

选择**视图(View)**, 并选择**地图(Map)**(图 74), 以显示地图视图(Map View), 默认情况下将只显示一个空地图(图 75)。

图 74. 选择地图视图

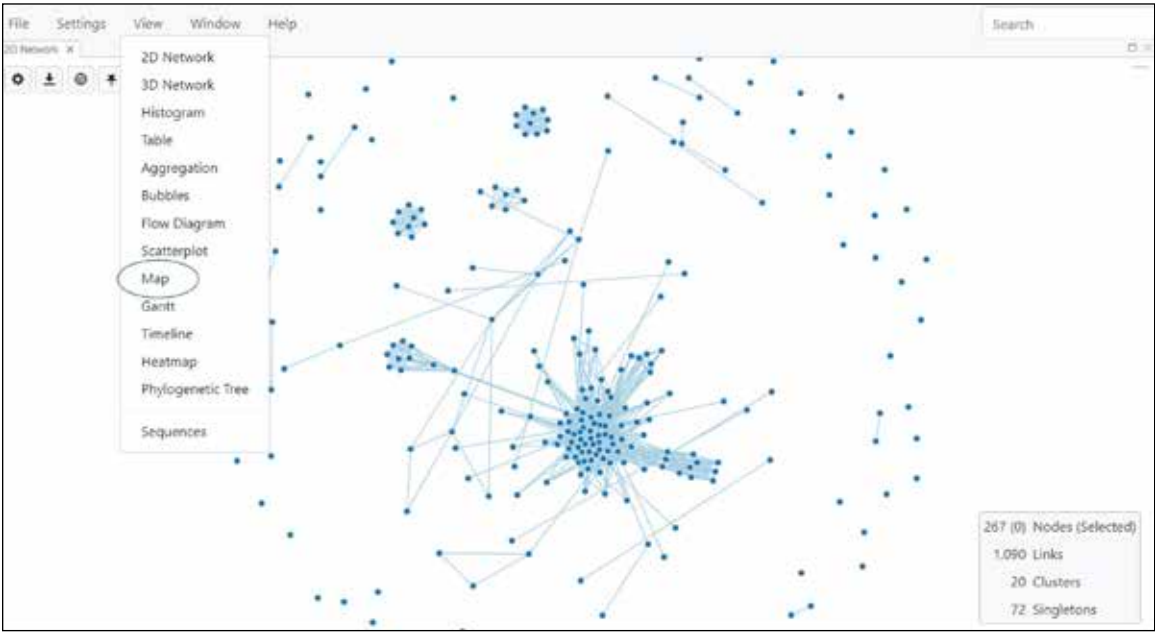



图 75. 地图视图的默认设置是不显示节点。



## 地图视图设置

使用以下部分中描述的各种设置, 根据节点文件中包含的地理信息自定义数据的可视化。

选择**切换地图设置**按钮  以打开设置对话框, 该对话框有四个选项卡: 数据、组成、节点和链接。

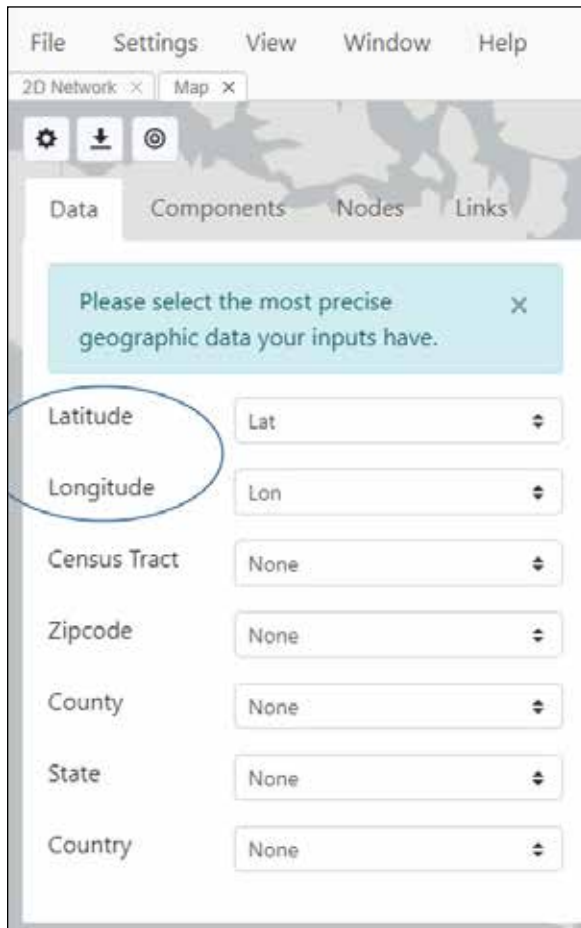
**数据选项卡:**选择**数据**会显示此功能的下拉菜单选项。

图 76. 使用数据选项卡更改地图选项



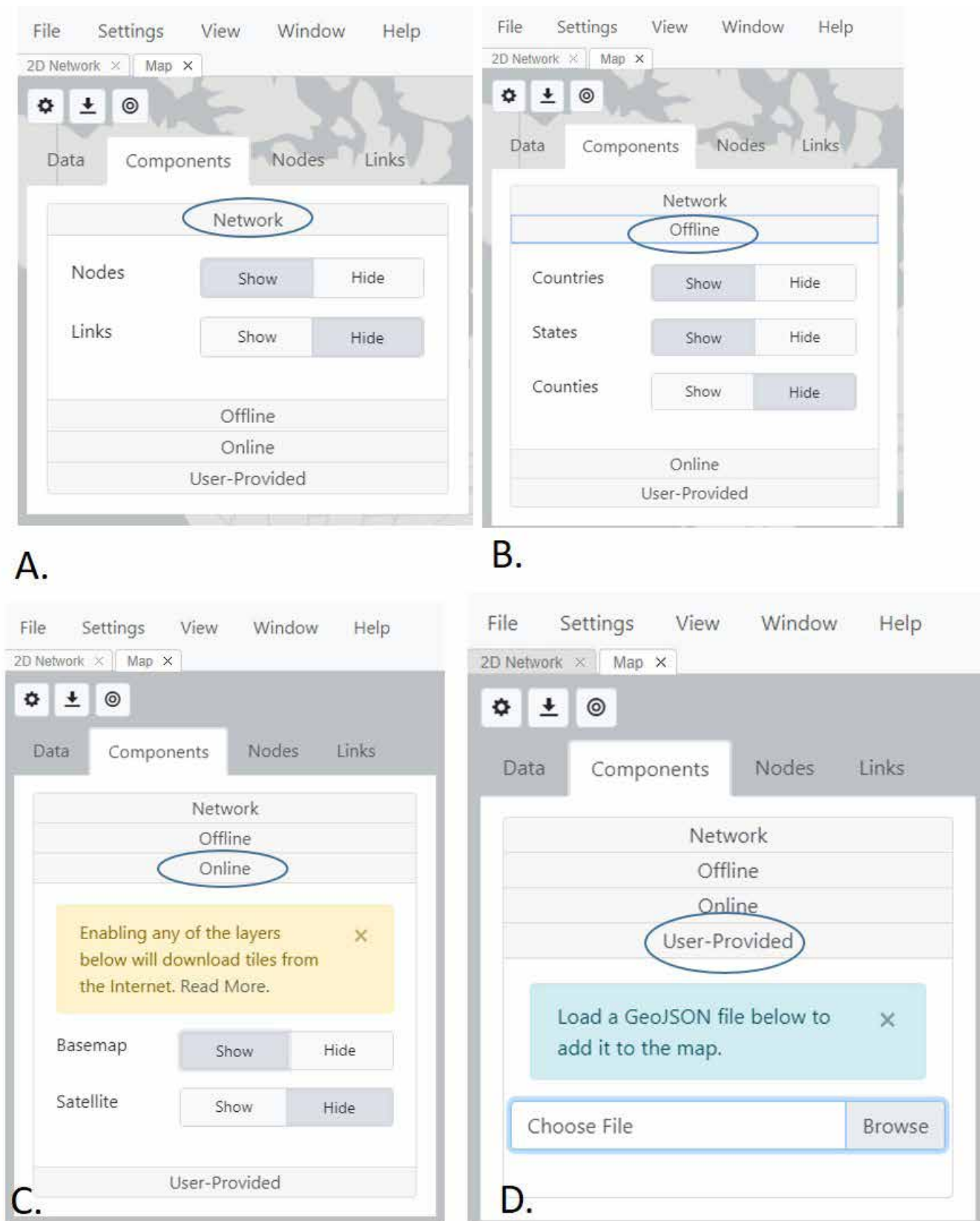
在这个例子中, 我们想通过纬度和经度将数据进行可视化。**(注意: 地图显示是分层的, 因此如果您的数据集包含下面列出的所有数据列, 并且您选择了多个属性, 则显示的地图将默认为最高可用的地理精度级别。)**请确保您只选择最适合您的数据集的变量, 而将其他变量保留为无。在这个亚特兰大地区 HIV-1 *pol* 序列数据集的示例中, 我们在节点列表中有纬度和经度 (lat lon) 信息, 并选择了这些作为需要使用的地理参数 (图 77)。

图 77. 地图选项和选择纬度和经度, 以在地图上显示节点的地理位置。



**组件选项卡:**选择 **组件 (Components)** 将允许对将显示的实际网络进行更改 (图 78A), 包括节点和链接。如果您在线或离线, 您还可以选择是否使用地图视图 (Map View) (图 78B-C)。此选项将确定显示哪些功能。离线时, 您可以选择显示或隐藏国家/地区、州和县。在线时, 您可以选择显示在 MicrobeTrace 中离线不可用的基本图或卫星视图图层。如果在线且您选择了这些选项, MicrobeTrace 将从互联网上下载基本地图和卫星地理数据 (称为图块), 这些数据类似于 Google 地图的功能。如果您生成了具有该文件格式的特定位置信息的数据, MicrobeTrace 还能够加载 GeoJSON 文件。要将 GeoJSON 文件添加到 MicrobeTrace 中, 请选择**用户提供的 (User-Provided)**, 浏览到计算机上存储文件的位置, 然后加载文件 (图 78D)。

图 78. 地图设置, 组件选项卡.A. 网络、B. 离线、C. 在线和 D. 用户提供的选项卡。

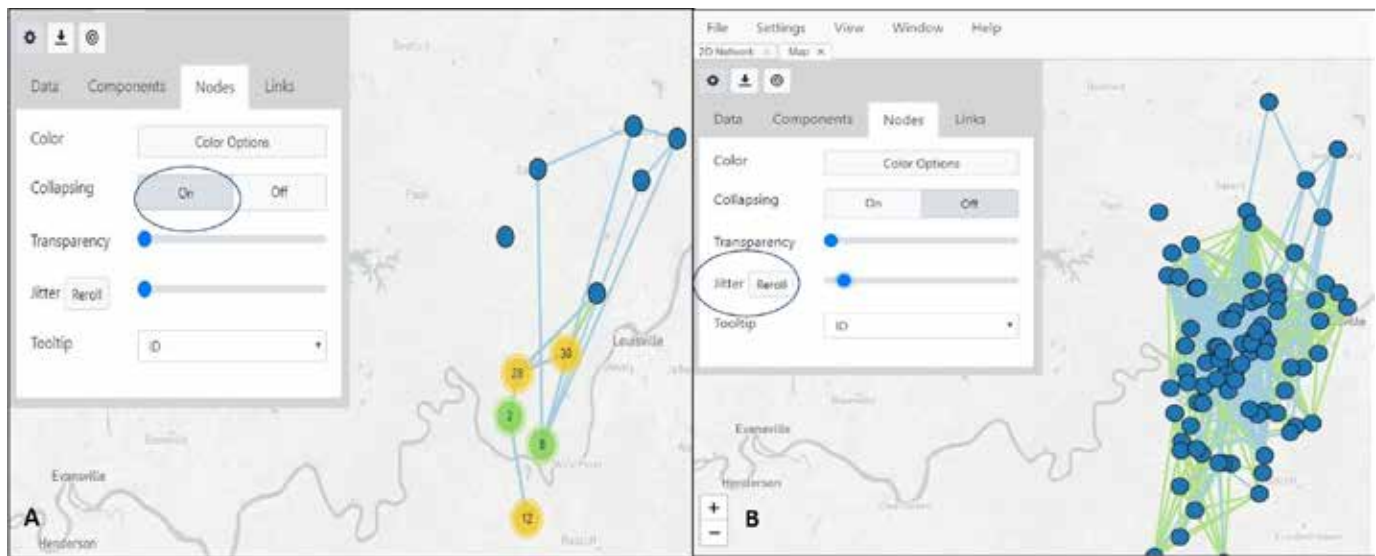




**节点选项卡:**选择**节点(Nodes)**, 以更改地图上节点的外观(图 79)。节点可以由节点文件中的任何变量着色。选择颜色选项会打开全球设置(Global Settings)菜单, 您可以像在其他视图中一样自定义样式设置、背景等。

使用相应的滑块更改节点的透明度和抖动速度。**在许多数据集中, 可能有许多节点共享相同的地理坐标, 导致地图中的节点密度非常高, 因此这些节点显示为单个大点。为了分隔具有相同地理坐标的节点, 使用抖动滑动条(图 79, 圈出)来增加抖动水平以分离或抖动节点, 使它们清晰可见。**使用**工具提示(Tooltip)** 更改当鼠标指针放置(“悬停”)在节点上时显示哪些变量。例如, 如果您从工具提示下拉菜单中选择 **ID**, 则当鼠标指针悬停在该节点上时, 该节点 ID 将会显示。

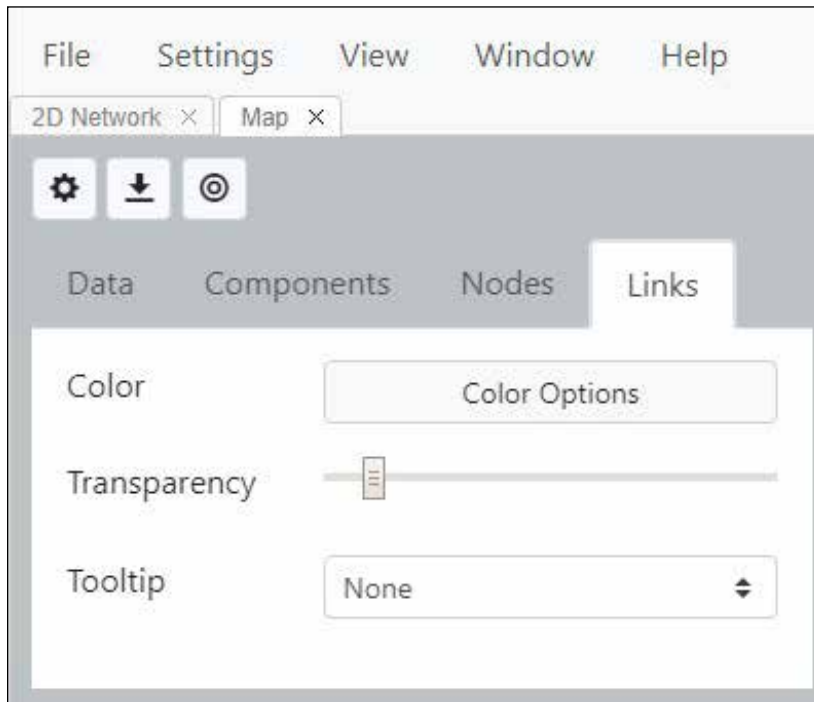
**图 79. 地图视图节点设置以更改节点颜色和工具提示功能的显示方式。**



节点折叠可以打开(面板 A)或关闭(面板 B)。增加抖动会随机分散节点;随机点击重新滚动会打乱周围的节点, 这是敏感数据的一个有用功能

**链接选项卡:**选择**链接(Links)**, 以更改链接设置(图 80)。这些功能与节点选项卡中的功能相同。您可以调整颜色、透明度和工具提示设置。

图 80. 地图视图链接设置



MicrobeTrace 根据您选择的各项选项在地图中显示节点数据。图 81 显示了带有按风险因素着色的节点和隐藏链接的地图。查看地图时，鼠标上的滚动条可用于平移或放大和缩小。默认情况下，地图会缩小，您会看到一个带有数字的圆圈，该数字代表节点的数量（图 81，顶部面板）。您将其放大时，节点会弹出，以形成更小、更离散的组。可以使用鼠标指针选择或取消选择单个或多个节点。这些选择将传播到网络 and 表视图。这使得能够在多个可视化窗口之间追踪特定个体。与其他视图一样，地图图像可以导出并保存为 .png、.svg 或 .jpg 图像文件。

图 81. 地图视图显示映射到纬度和经度坐标的节点, 按风险因素着色, 并显示底图。节点折叠, 面板 A 中的抖动关闭。面板 B 显示当您悬停在折叠节点上时的节点分布; 组件节点显示为圆形或螺旋形。所有数据均为假设, 仅用于演示目的

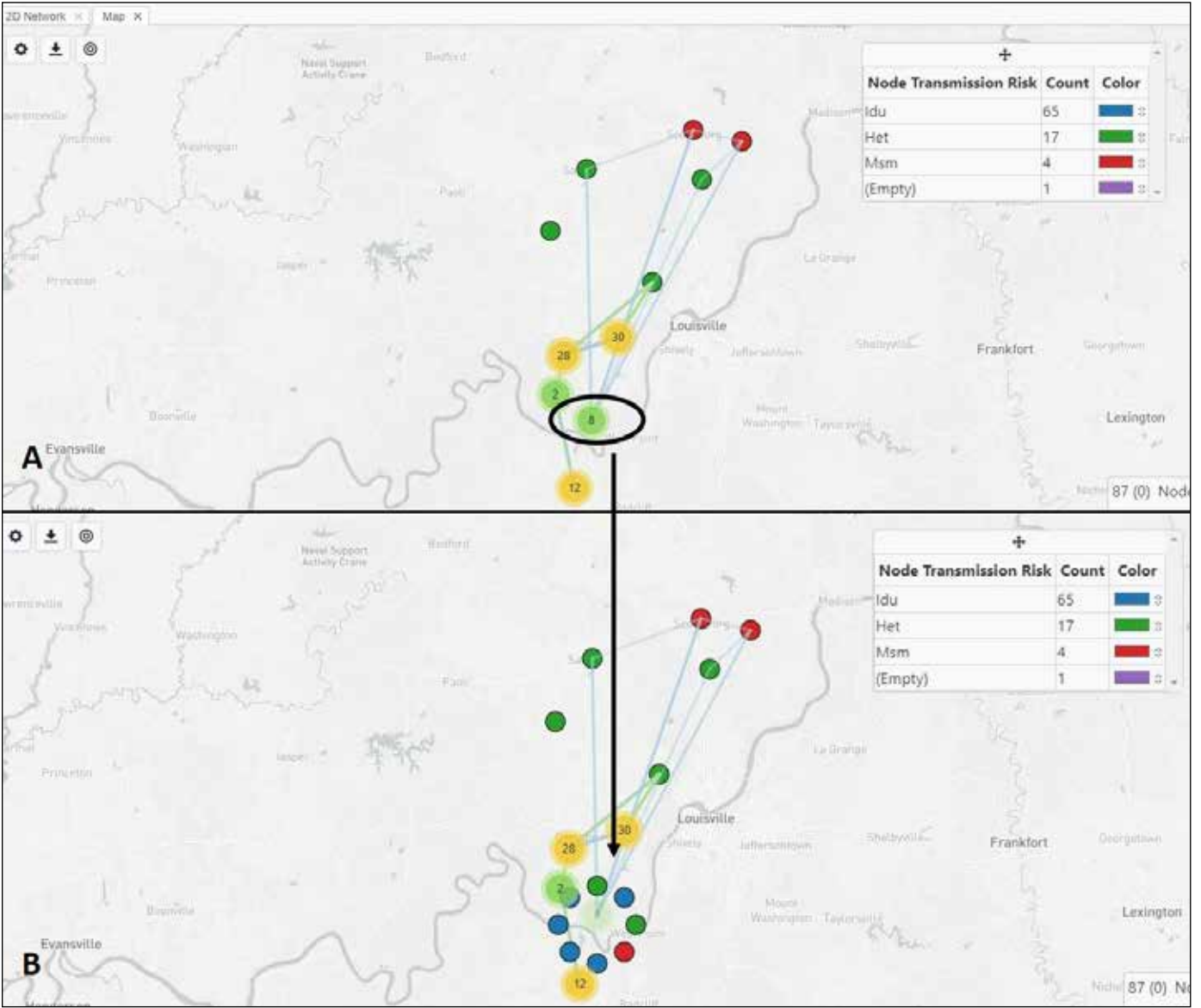
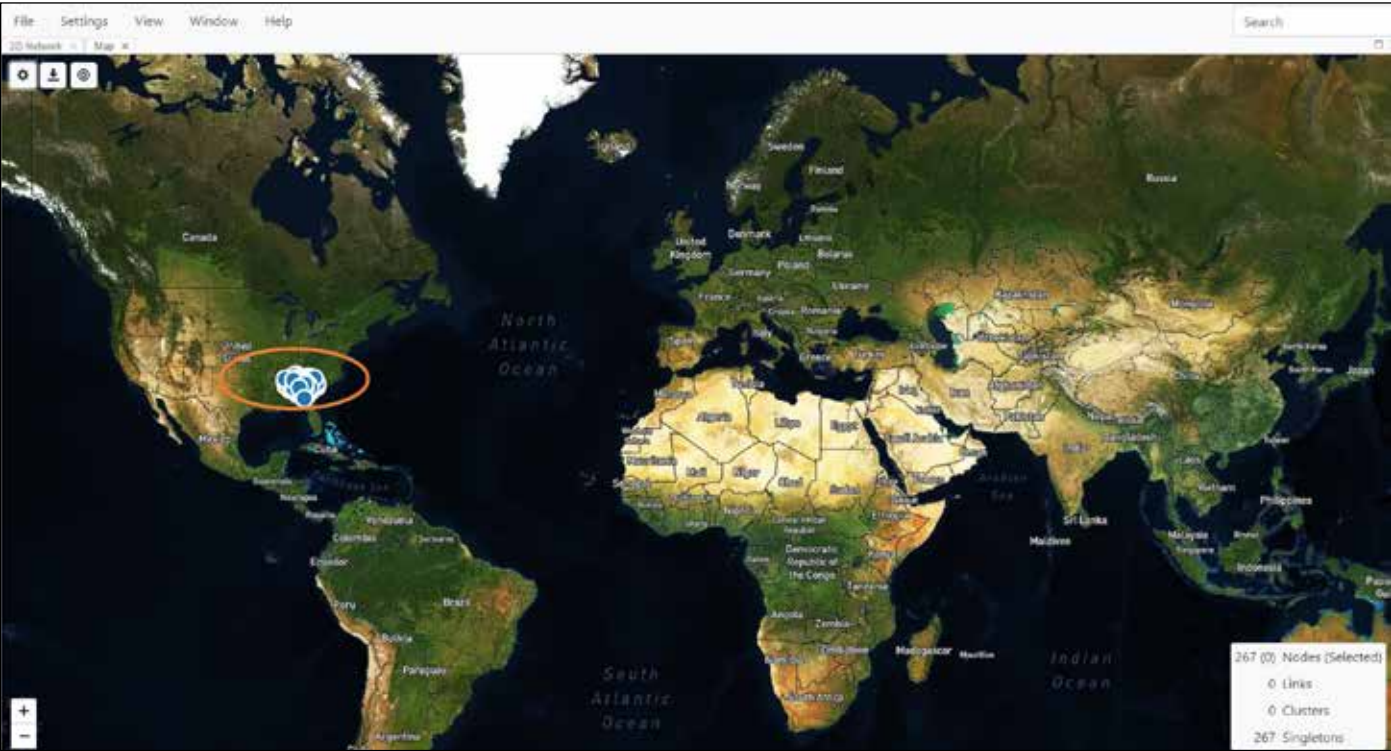


图 82. 地图视图显示映射到纬度和经度坐标 (圆圈) 的节点, 按风险因素着色, 并显示卫星特征。

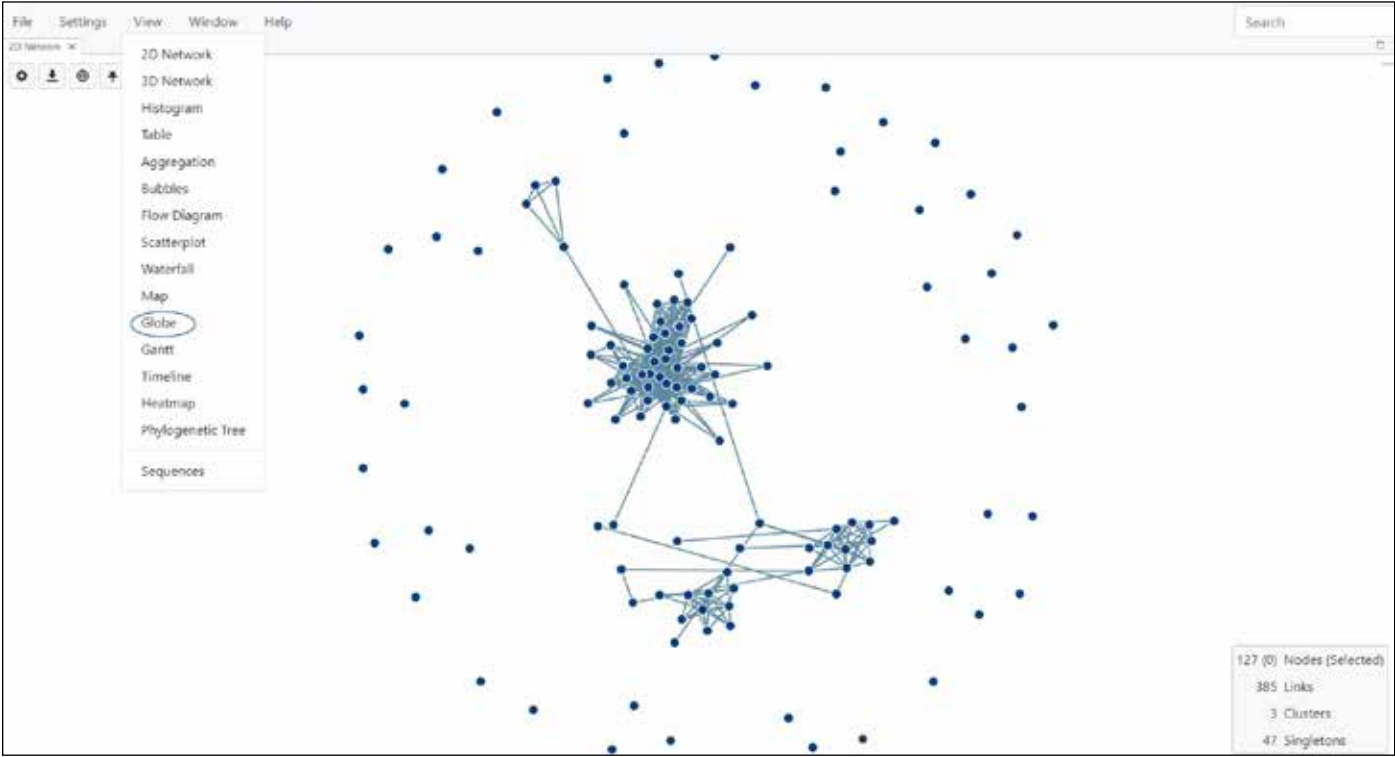


# 地球视图

该视图类似于地图视图, 如果节点文件中包含地理坐标, 则数据中的节点可以以全球格式显示。不同之处在于, 在这个视觉中, 节点和链接被绘制在一个可以旋转的三维地球上。这种观点对于国际性爆发或查看各国家/地区之间的潜在联系特别有用。布局和选项与地图视图中的非常相似。尽管纬度和经度为您提供最精确的位置, 但如果纬度经度数据不可用, 邮政编码和其他地缘政治分界线(县和州)也会显示在地图上。

选择**视图(View)**, 并选择**地球(Globe)**(图 83), 以显示地球视图(Globe View)。

图 83. 选择地球视图



打开视图是一个标准的 3-D 地球(图 84)。

图 84. 地球视图 - 打开窗口



要显示节点, 需要将它们映射到节点列表中提供的地理坐标。

## 变更地球视图选项

使用以下部分中描述的各种设置, 根据节点文件中包含的地理信息自定义数据的可视化。


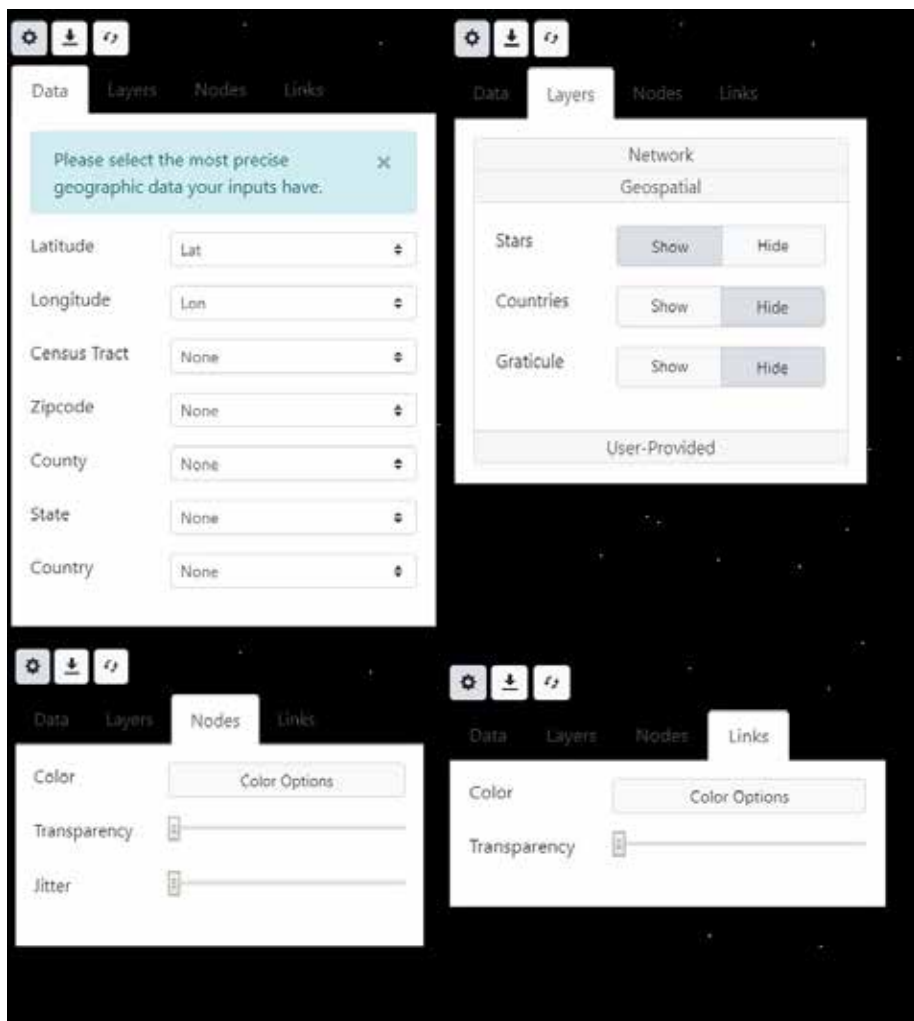
选择**切换地图设置**按钮  以打开设置对话框, 该对话框有四个选项卡: 数据、层、节点和链接(图 85)。

图 85. 使用数据、层、节点和链接选项卡显示可用设置的地球视图

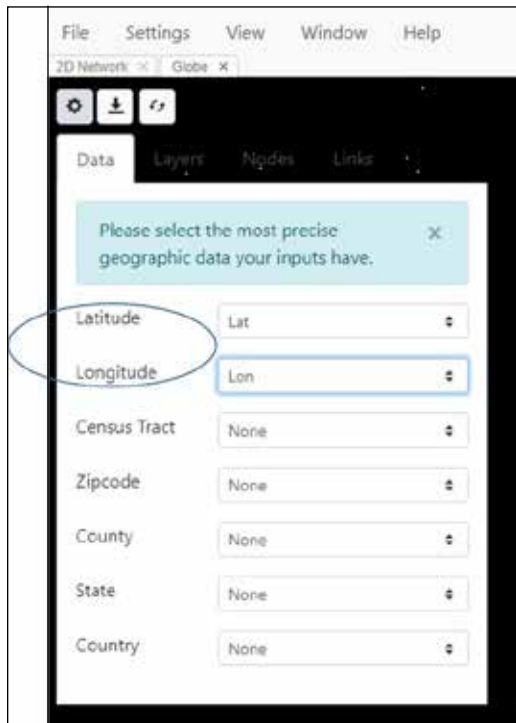


## 数据选项卡

选择**数据**会显示此功能的下拉菜单选项。在这个例子中, 我们想通过纬度和经度来将数据进行可视化。**(注意: 地图显示是分层的, 因此如果您的数据集包含下面列出的所有数据列, 并且您选择了多个属性, 则显示的地图将默认为最高可用的地理精度级别。请确保您只选择最适合您的数据集的变量, 并将其他变量保留为无。在这个欧洲 HIV-1 pol 序列的示例数据集中, 纬度和经度(lat-lon)信息位于节点列表中, 并被选定为要使用的地理参数(图 86)。**

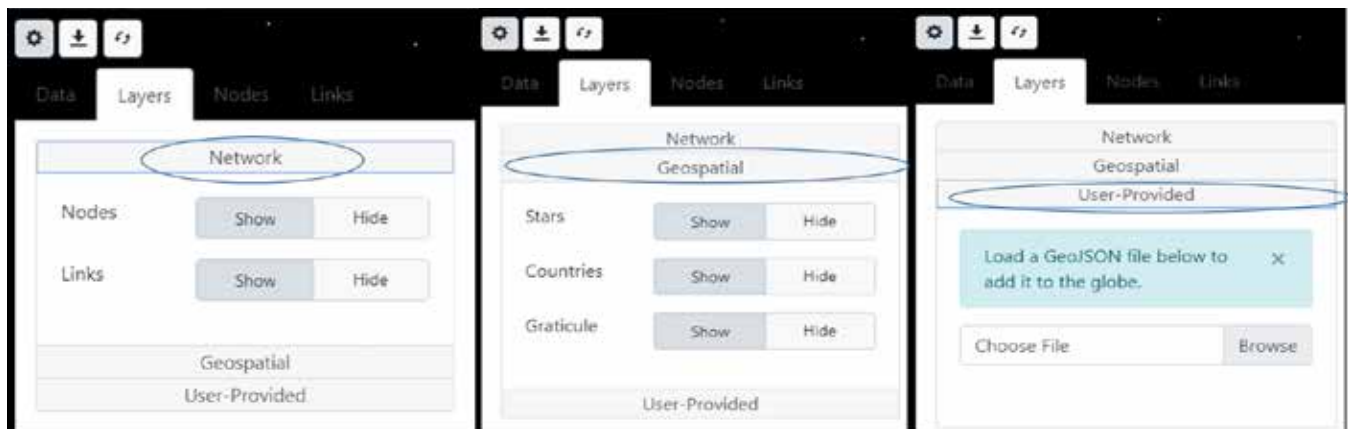


图 86. 用于选择纬度和经度的数据选项卡以显示地球上节点的地理位置。



层选项卡:选择层(Layers)将允许您使用网络、地理空间和用户提供的数据进一步自定义显示(图 87)。

图 87. 使用数据层选项卡自定义地球视图设置



**网络:**此选项允许在地球上显示网络节点和链接。

**地理空间:**允许显示某些地理空间特征, 包括星星、国家和经纬网(地球上表示经度和纬度的线)。背景中显示的夜空是为了反映真实的星座而开发的。

**用户提供:**MicrobeTrace 还能够加载您生成的具有该文件格式的特定位置信息的 GeoJSON 文件。要将 GeoJSON 文件添加到 MicrobeTrace, 请选择用户提供的(User-Provided), 浏览到计算机上存储文件的位置, 然后加载文件。


**节点选项卡:**选择**节点(Nodes)**, 以更改地图上节点的外观(图 88)。节点可以由节点文件中的任何变量着色。选择颜色选项(Color Options)会打开全球设置(Global Settings)菜单, 您可以像在其他视图中一样自定义样式设置、背景等。

使用相应的滑块更改节点的透明度和抖动速度。**\*请注意:**在许多数据集中, 可能有许多节点共享相同的地理坐标, 导致地图中的节点密度非常高, 因此这些节点显示为单个大点。为了分隔具有相同地理坐标的节点, 请使用抖动滑块增加抖动级别, 以分隔或抖动节点, 使它们清晰可见。

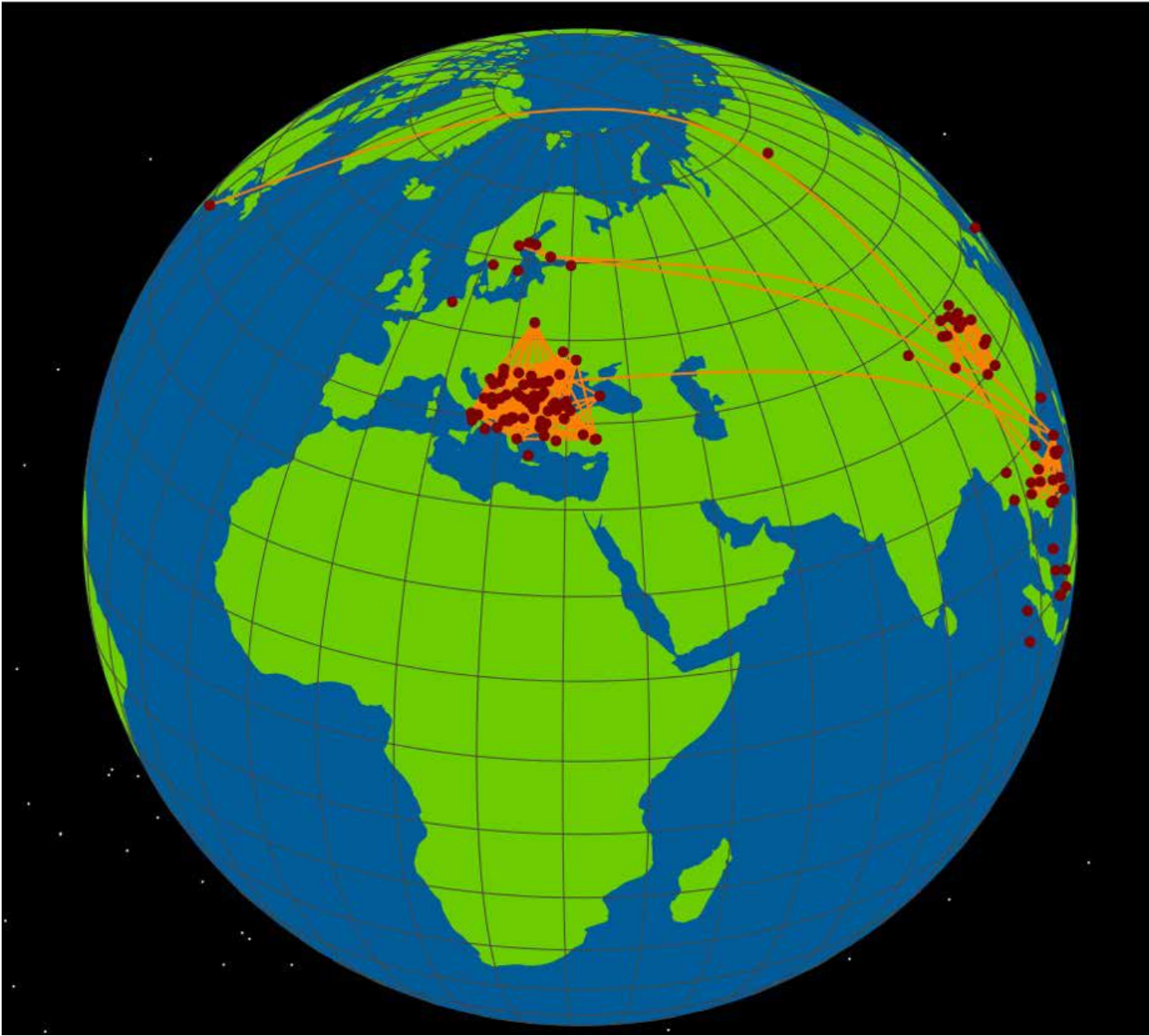
图 88. 在地球视图(Globe View)中更改节点设置



**链接选项卡:**选择**链接(Links)**以更改链接颜色和透明度设置。这些功能与节点选项卡中的功能相同。

MicrobeTrace 根据您选择的各项选项在地球上显示节点数据。图 89 显示了地图, 其中节点颜色随着显示的链接而改变。使用鼠标上的滚动条平移或放大和缩小。左键点击时移动鼠标可让您旋转地球, 并轻松查看洲际网络连接。除了左上角的常用按钮外, 地球视图还有一个额外的旋转按钮 。点击此图标可缓慢旋转地球。与其他视图一样, 地图图像可以导出并保存为 .png、.svg 或 .jpg 图像文件。

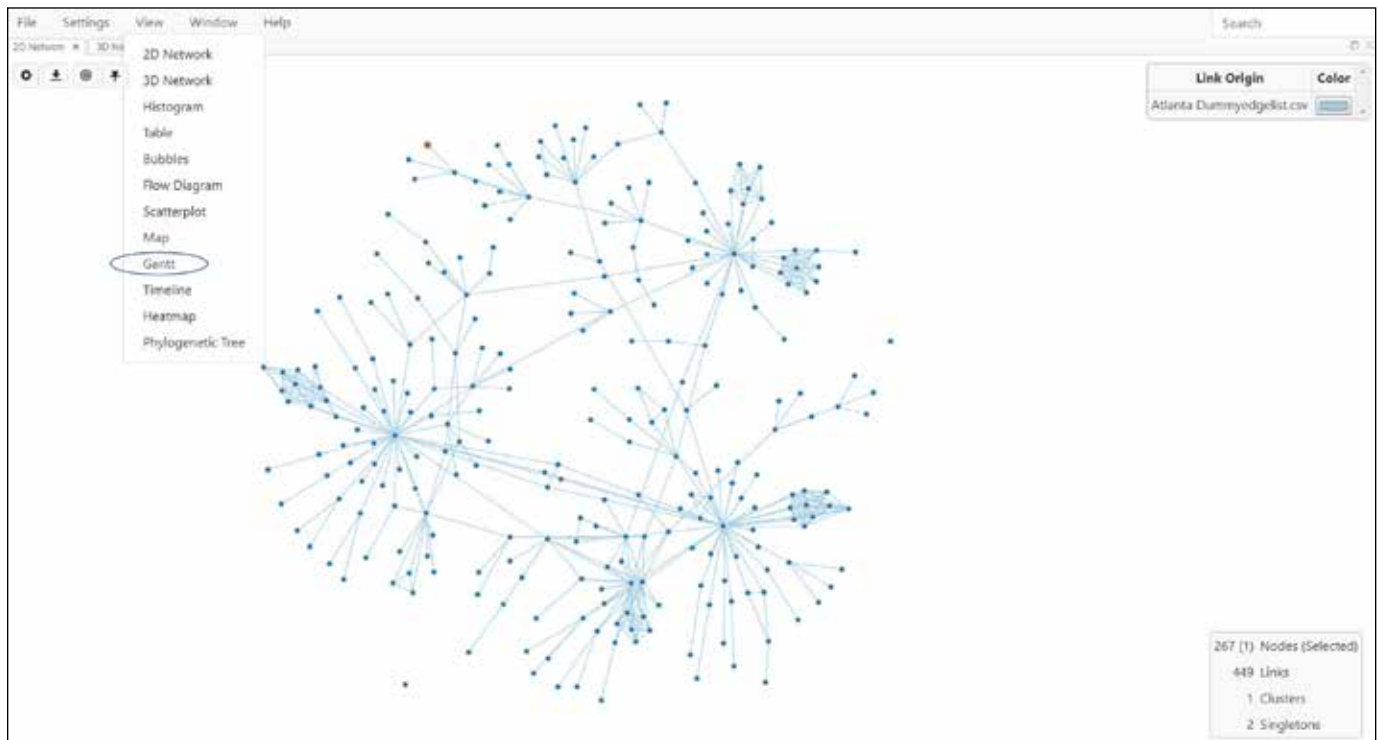
**图 89. 具有自定义设置的地球视图, 节点映射到纬度经度坐标; 节点和链接颜色改变, 抖动增加。**



## Gantt 视图

此视图用于以开始日期和结束日期列的形式包含时间跨度或时间增量的数据(例如感染期、在医院或其他机构的时间长度,或用于资源管理)。Gantt 图在 X 轴上绘制时间跨度,在 Y 轴上绘制节点(单个)ID。该图表使您能够查看个人的一个或多个时间跨度,并在需要时查找重叠。从**视图(View)**菜单(图 90)中选择 Gantt 图。

图 90. 选择 Gantt 图视图




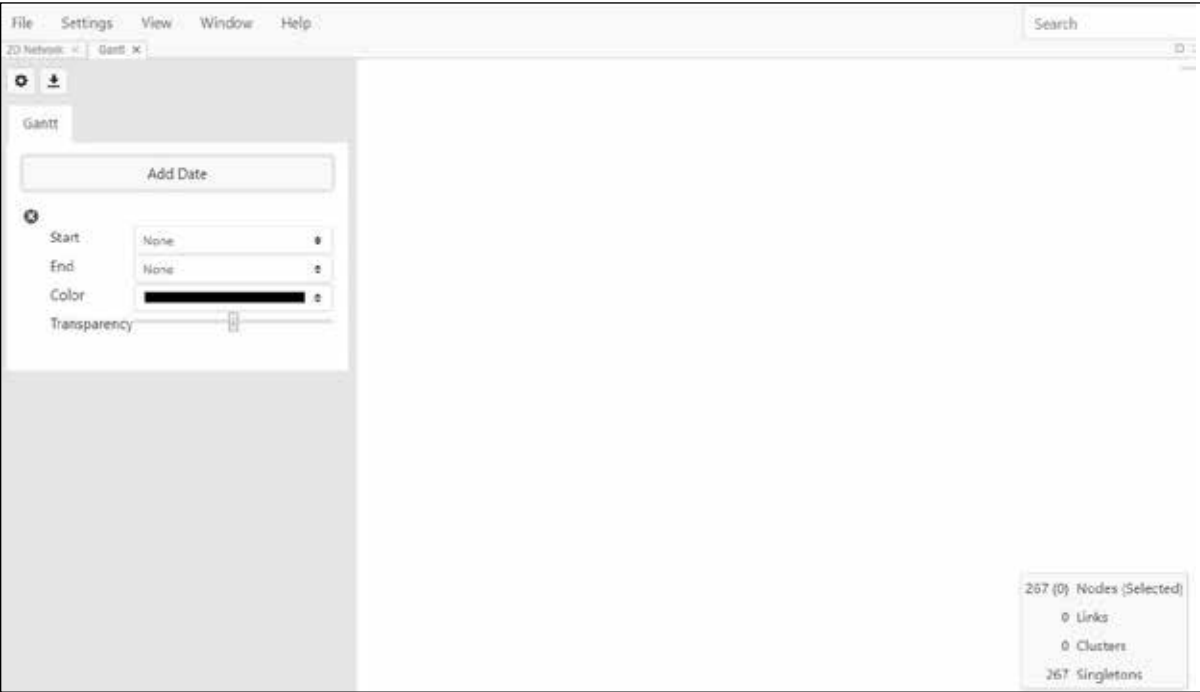
Gantt 图最初是空白的,但选择**切换 Gantt 设置(Toggle Gantt Settings)**按钮会  打开一个菜单,以添加节点列表中包含的日期(图 91)。选择**添加日期(Add Date)**,并从下拉菜单中选择开始和结束日期。如果数据有多个开始日期和结束日期列,您可以再次选择**添加日期(Add Date)**以添加更多日期。您还可以为每个时间跨度设置颜色。

图 91. 在 Gantt 图视图中添加一个或多个日期范围的菜单, 并为每个范围选择配色方案。



更改将立即反映在显示的 Gantt 图中(图 92)。

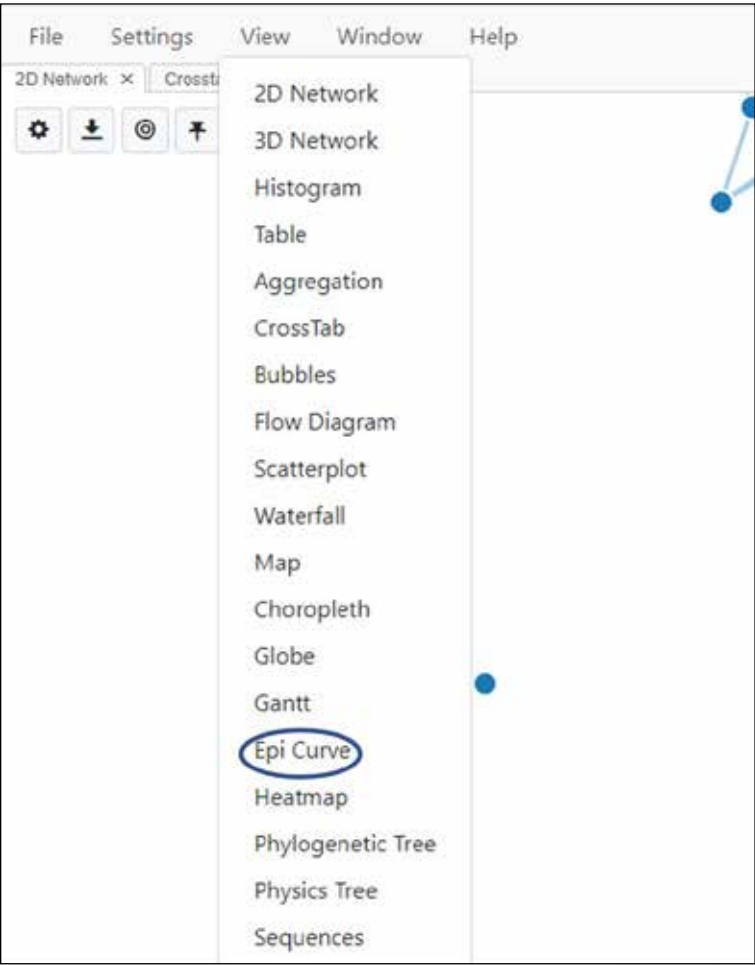
图 92. 选择了两个时间跨度的 Gantt 视图, 并针对 Y 轴上的节点 ID 绘制。每个时间段都用不同的颜色表示。



# Epi 曲线

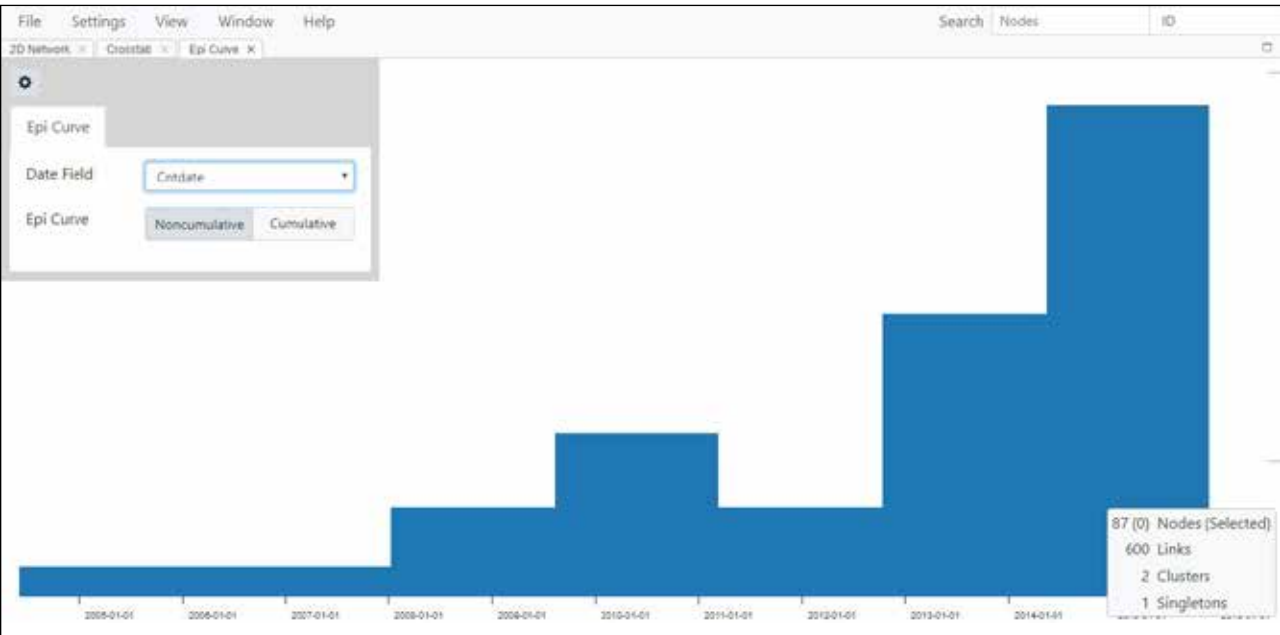
X 轴上绘制的日期的图形表示可以使用 Epi 曲线视图进行可视化, 该视图将您的数据绘制在时间轴上, 以创建累积或非累积的 Epi 曲线。从视图 (View) 菜单 (图 93) 中选择 **Epi 曲线 (Epi Curve)**。这将打开一个没有任何日期字段选择的窗口 (图 94)。

图 93. 选择 Epi 曲线视图



从下拉菜单中选择相关的日期字段(图 94)。您可以选择累积型或非累积型 Epi 曲线。

图 94. Epi 曲线(Epi Curve)视图显示默认视图(未选择日期)。



此功能通常与网络视图结合使用。有关如何并排或以平铺模式排列一个或多个视图以进行同时分析的说  
明, 请参阅[平铺视图\(Tiling Views\)](#)。

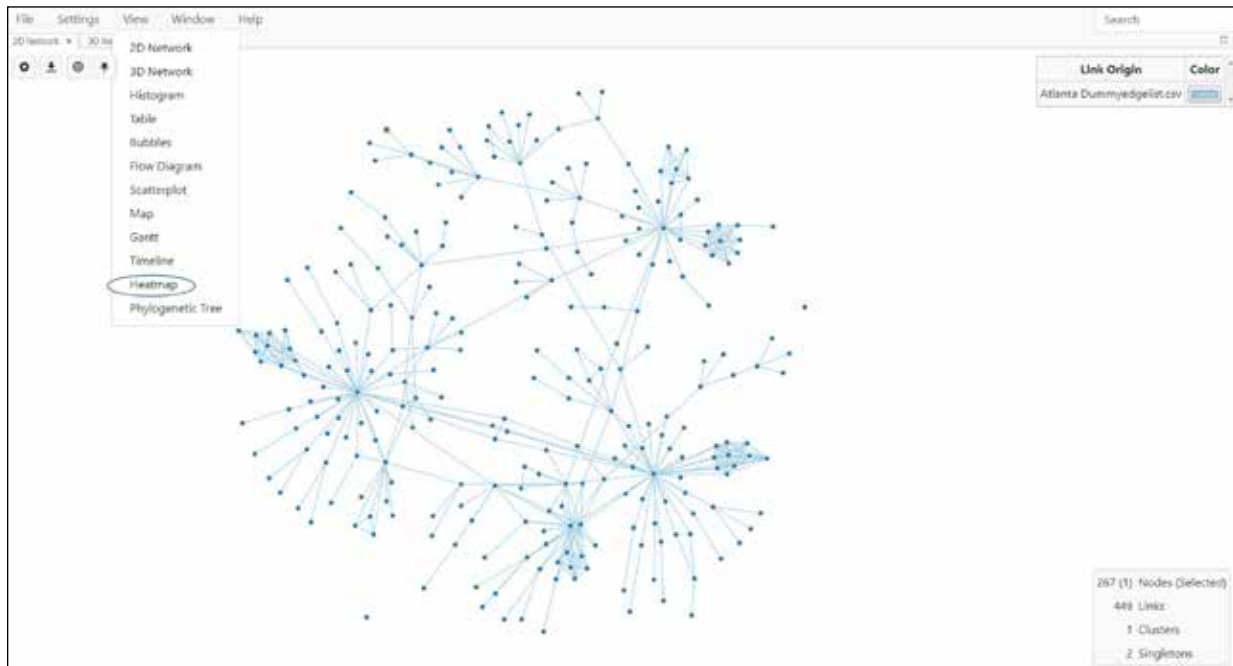
## 热图视图

如果分析包括带有核苷酸序列的 FASTA 文件, 则 MicrobeTrace 提供计算的核苷酸遗传距离矩阵的热图会  
为可视化。

要显示热图, 请选择菜单(Menu)栏上的查看(View), 然后选择热图(Heatmap)(图 95), 以开启一个新窗  
口, 显示热图矩阵中的遗传距离(图 96)。

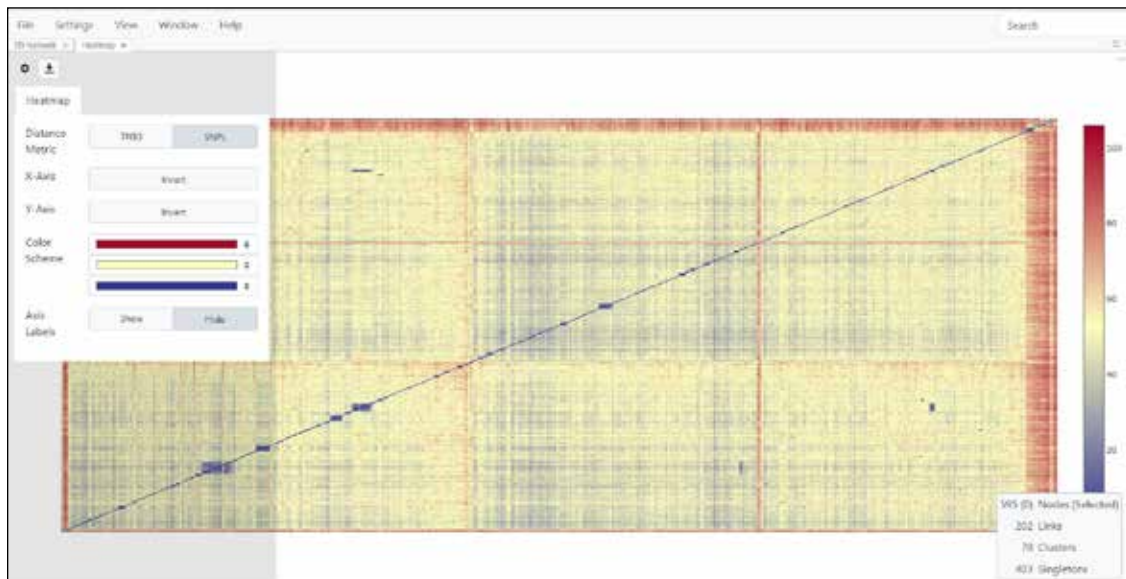



图 95. 选择热图视图



请注意, 渲染遗传距离矩阵需要一些时间, 因此窗口将显示为空白, 直到计算完成并加载遗传距离热图。

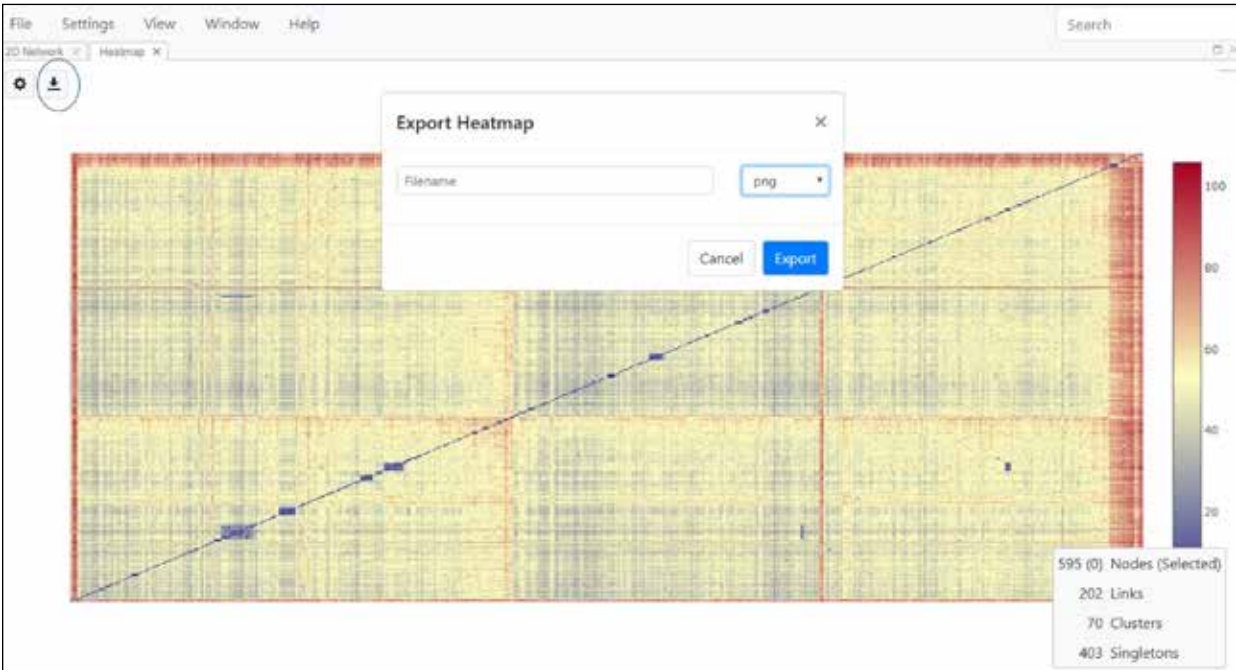
图 96. FASTA 文件中序列的遗传距离矩阵的热图视图 (Heatmap View)



矩阵中的每个单元格代表数据集中两个序列之间的遗传距离(替换/位点)。对角线上的单元格(从左下到右上)表示序列与其自身的比较(即, 遗传距离为零, 导致明显不同的对角线)。热图右侧的比例尺表示用于分析数据集中遗传距离范围的配色方案。例如, 深蓝色单元格表示与黄色或红色单元格相比在遗传上更密切相关的序列。通过将鼠标指针悬停在矩阵中所需的单元格上, 可以查看两个序列的实际遗传距离值。弹出的气泡将显示两个序列的 ID 及其计算的遗传距离。可以通过选择切换热图设置 (Toggle Heatmap Settings) 按钮来更改设置 。这使您可以更改距离度量、反转轴、更改配色方案或显示或隐藏轴标签。

通过选择左上角的导出热图(Export Heatmap)按钮,可以将热图图形保存为 .png 或 .jpg 图像文件(图 97)。使用导出(Export)对话框导航到计算机上的所需目的地,然后输入要保存的图像的文件名。选择保存以保存图像文件。您还可以将实际距离导出为 .csv 文件,然后可以在 Excel 中查看。

图 97. 将热图导出为图像文件或包含计算距离的 .csv 文件。



## 序列视图

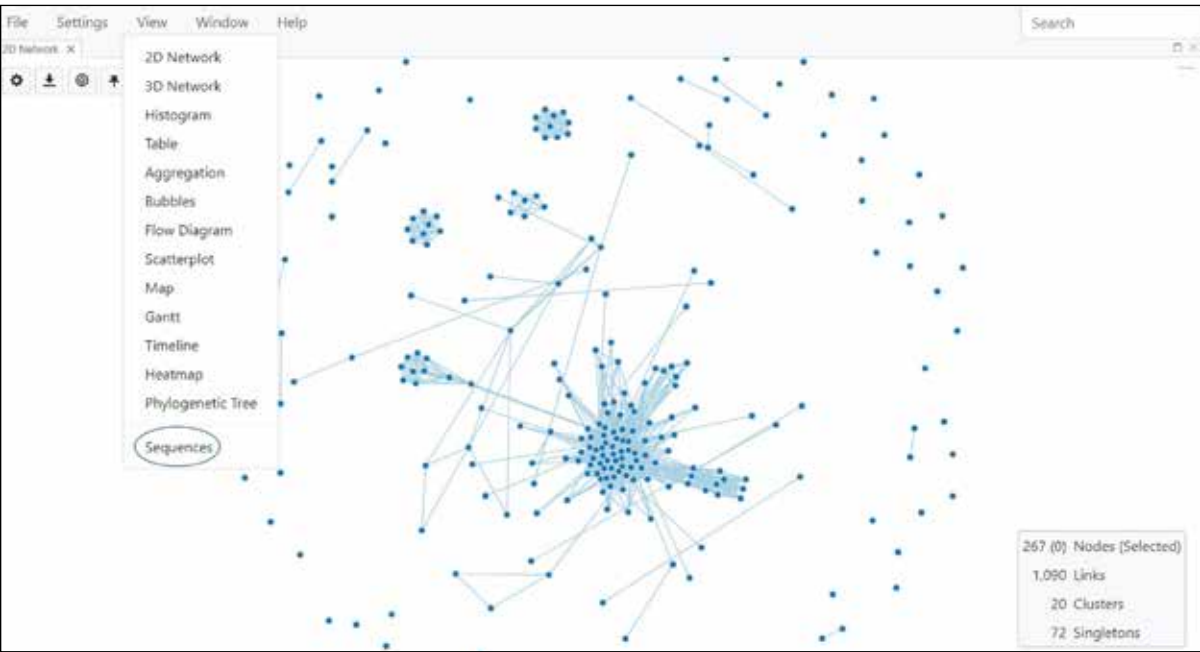
序列视图(Sequence View)的主要功能是允许对最终比对的质量进行目视检查,以确保比对中没有意外的间隙、插入等。不正确的比对将极大地影响遗传距离的确定,从而影响推断的网络或系统发育树。这对于非 HIV 序列可能尤其如此,因为 MicrobeTrace 中的比对器未配置为处理来自 HIV-1 以外的病原体的序列。使用通常用于 HIV-1 的 TN93 核苷酸替换模型,非 HIV-1 序列可能并不总是正确比对。如果在分析中使用预比对的 FASTA 文件,则序列视图的价值可能有限,除非使用序列视图重新检查该比对。

**\*重要提示\*** 我们强烈建议在 MicrobeTrace 中使用所有预比对之前检查其质量。另外,请不要使用 MicrobeTrace 来比对预先比对的序列,这可能会改变比对。请注意,在序列查看器中对比对所做的任何编辑都不会自动呈现在比量的网络中。编辑后的序列比对文件必须先保存,然后可以在 [文件加载步骤](#) 中从 MicrobeTrace 分析开始时使用。

选择视图(View)菜单下的序列(Sequence)(图 98),以在新窗口中将序列显示为比对(图 99)。

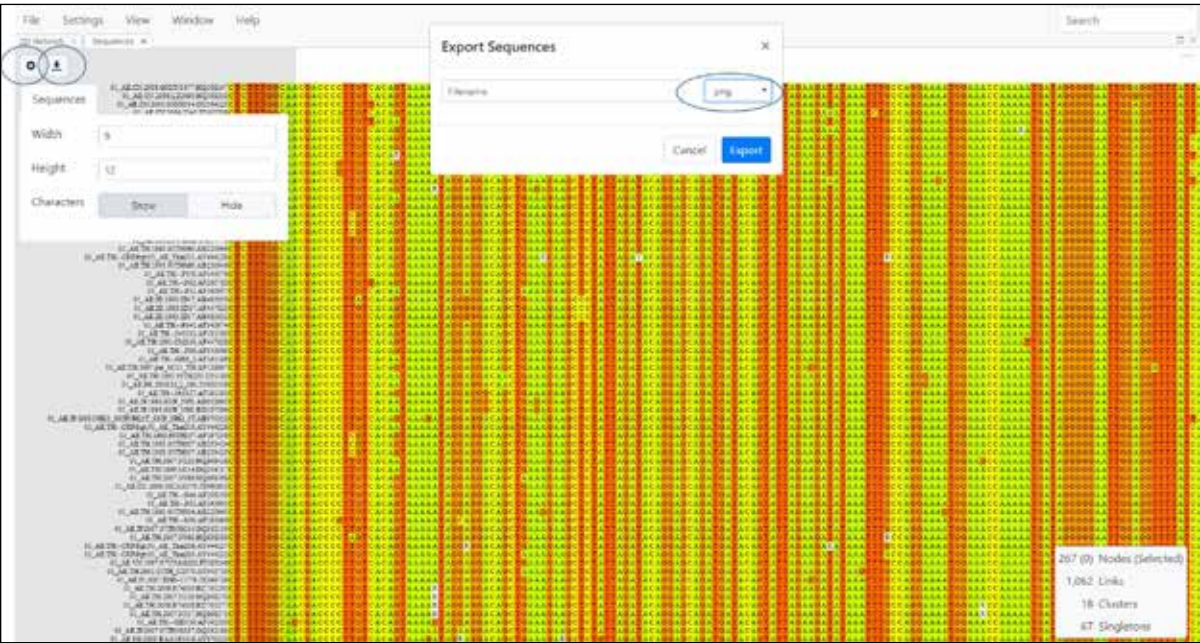
使用滚动条来操纵显示的序列视图。

图 98. 选择序列视图



您可以使用鼠标上的滚轮水平和垂直地滚动, 以检查您的比对情况。您可以使用设置按钮更改核苷酸字符的外观 (图 99, 圈出)。您还可以使用选择下载按钮时出现的对话框, 以多种格式(png - 便携式网络图形、FASTA 或 MEGA - 是 [分子遗传学进化分析 \(Molecular Genetics Evolutionary Analysis, MEGA\)](#) 程序中使用的文件格式) 导出序列 (图 99, 圈出)。序列 ID 将以您选择的文件格式附加到序列中。

图 99. 显示比对的核苷酸序列的序列视图。对话框允许您更改比对外观并导出比对。

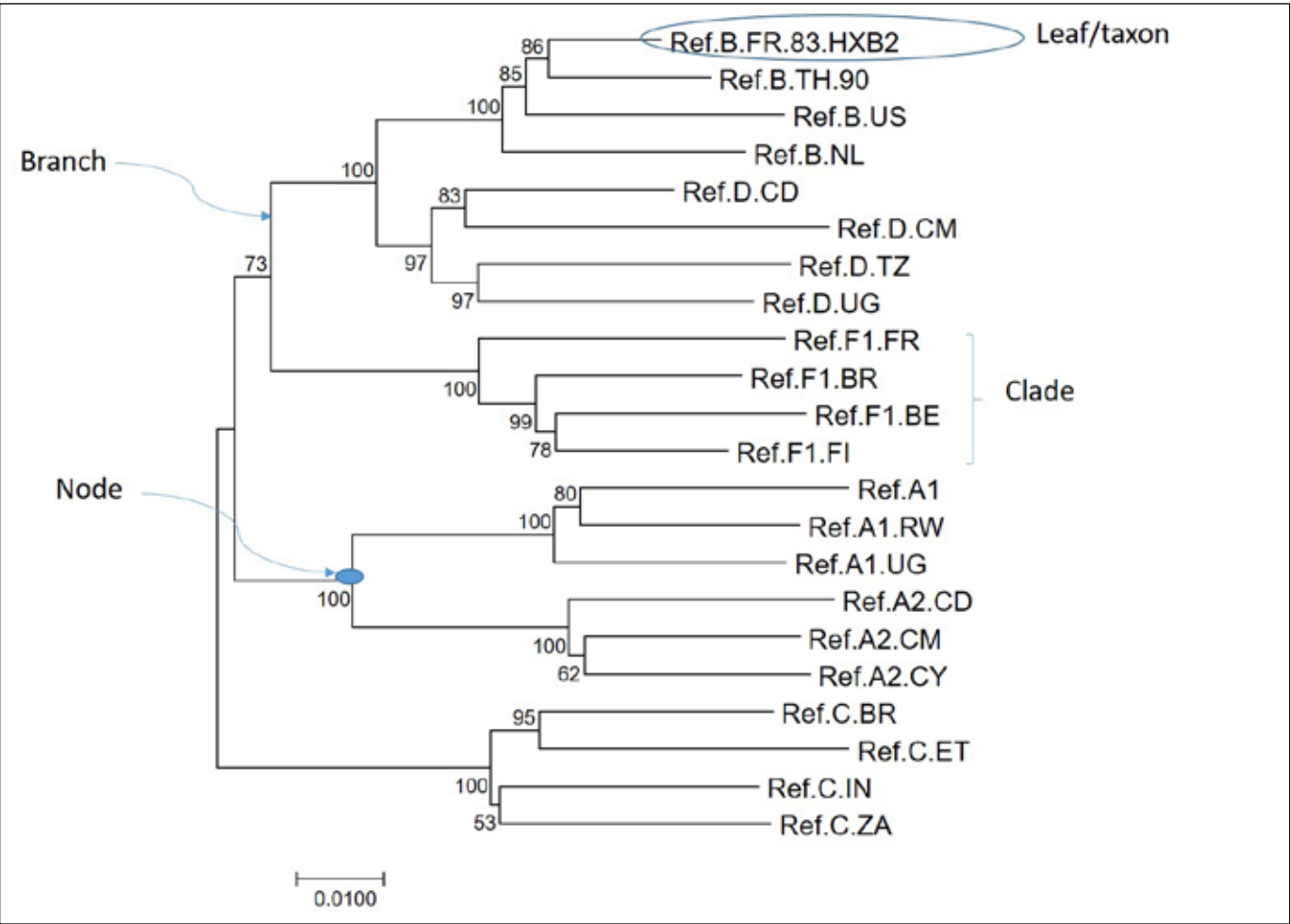




# 系统发育视图

系统发育视图(Phylogeny View)用于使用 FASTA 文件中的序列和基于遗传距离的邻接(NJ)方法(neighbor-joining(NJ) method)生成系统发育树。与遗传距离网络一样,系统发育树代表了来自一组生物的一组序列之间的进化关系。例如,系统发育分析表明,HIV-1 由 M、N、O 和 P 四个主要组群组成。M 组是在全球范围内传播的最常见的组。M 组进一步分为多个亚型,代表密切相关但不同的病毒基因型。B 亚型是美国最常见的亚型。在矩形系统发育树(图 100)中,水平线称为树的分支或尖端,代表每个分类单元或后代,例如“家谱”中的后代。树的节点代表一个或多个分类单元(分类单元的复数)的推断共同祖先。序列或分类群的簇称为进化枝,它们代表了密切相关的遗传序列。进化枝是一组分类群,包括祖先和该祖先的所有后代或分类群。例如,HIV-1 B 亚型被认为是由所有 B 亚型序列组成的进化枝。

图 100. 带有标记的组件的矩形邻接树示例

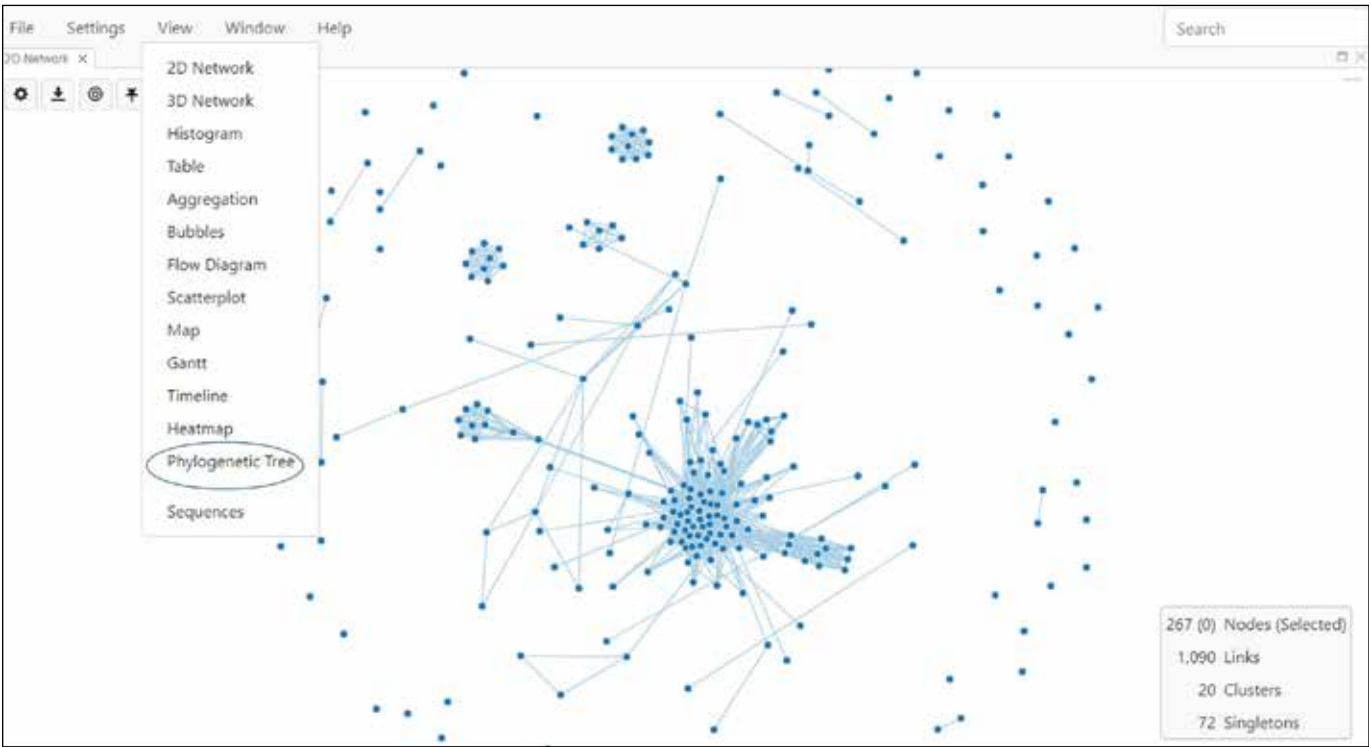


系统发育树可以有根或无根。有根树提供有关树中节点顺序的信息。树的根是所检查数据集的最古老的祖先谱系。无根树显示了分类群的关系,而不对祖先做出假设。用于系统发育视图的 NJ 方法推断有根树。

树中水平分支的长度与数据集中的遗传变化量成正比。系统发育视图(Phylogeny View)中的比例尺提供了数据集中核苷酸替换/位点的数量,用于推断树中的分支长度。垂直线没有任何意义,但用于在树中均匀显示分类群。

要显示系统发育树,请从视图(View)菜单中选择系统发育树(Phylogenetic Tree)(图 101)。

图 101. 选择系统发育树视图




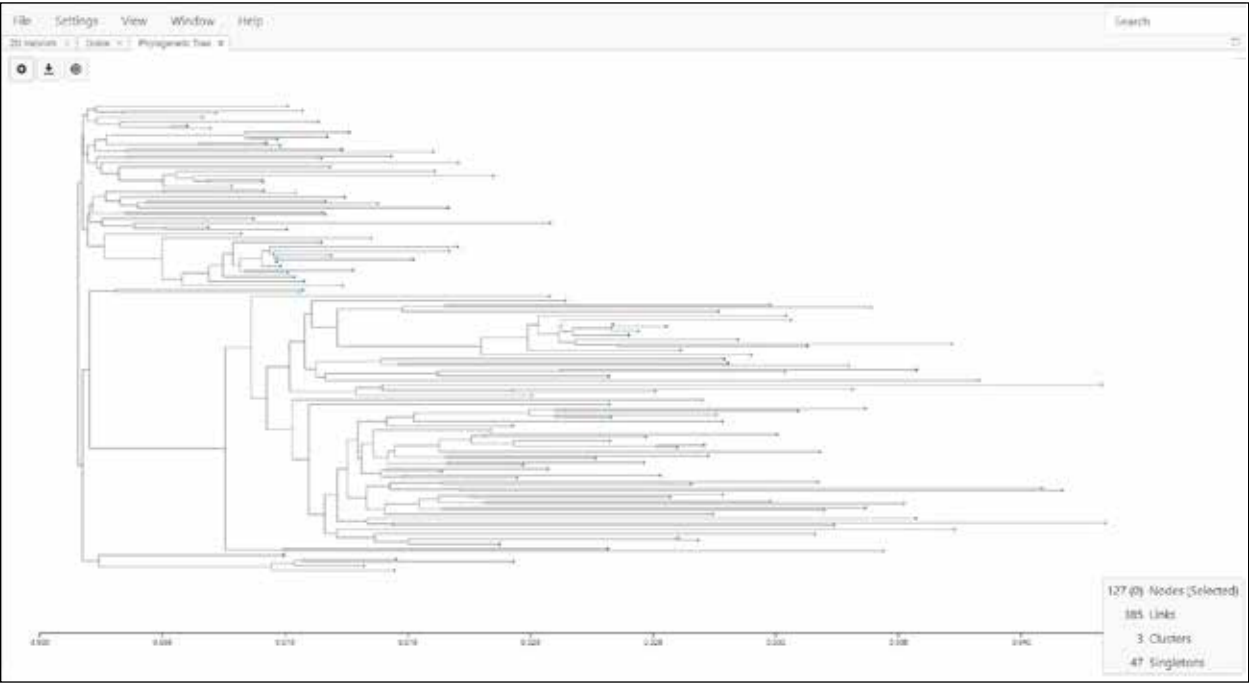
然后 NJ 树将显示在一个新窗口中(图 102)。您可以使用鼠标放大和缩小,并使用中心和缩放图标  将树重新定向到其默认大小和位置。

图 102. 默认加入树视图的比邻



系统发育设置


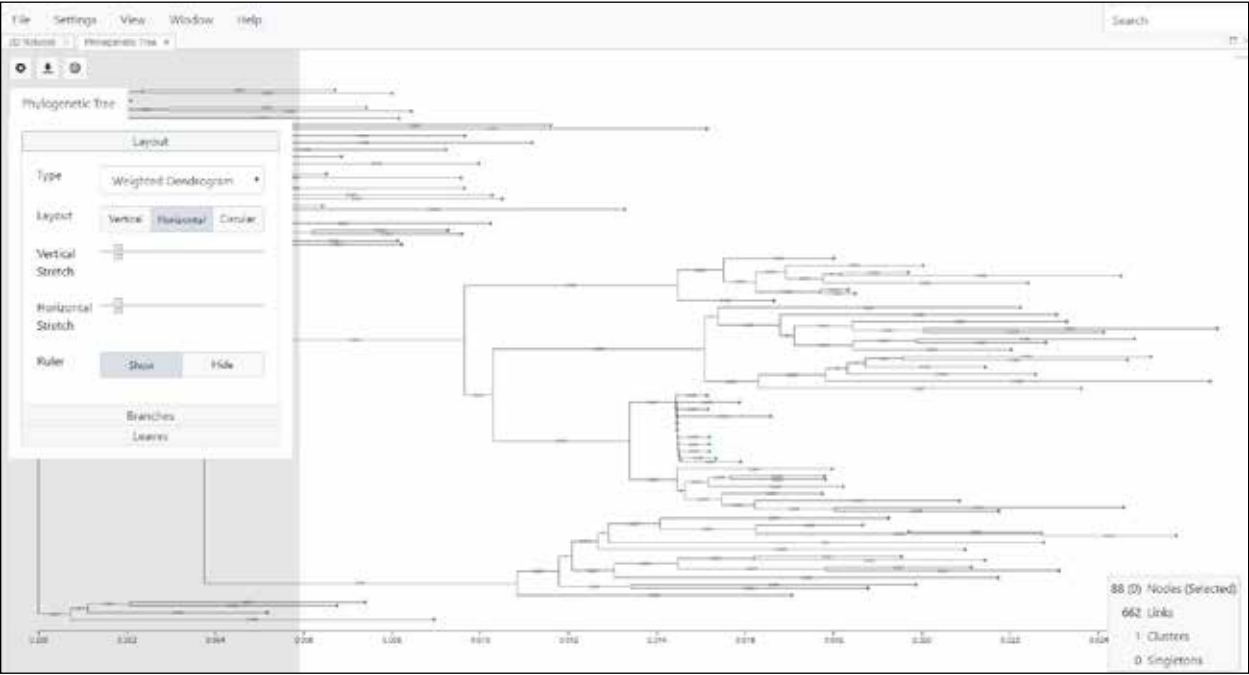
可以使用切换系统发育设置 (Toggle Phylogeny Settings) 按钮  来自定义树设置以访问菜单。此菜单中有三个选项卡: 树、树枝和树叶 (图 103)。

图 103. 显示设置菜单选项卡的系统发育树视图 (Phylogenetic Tree View)



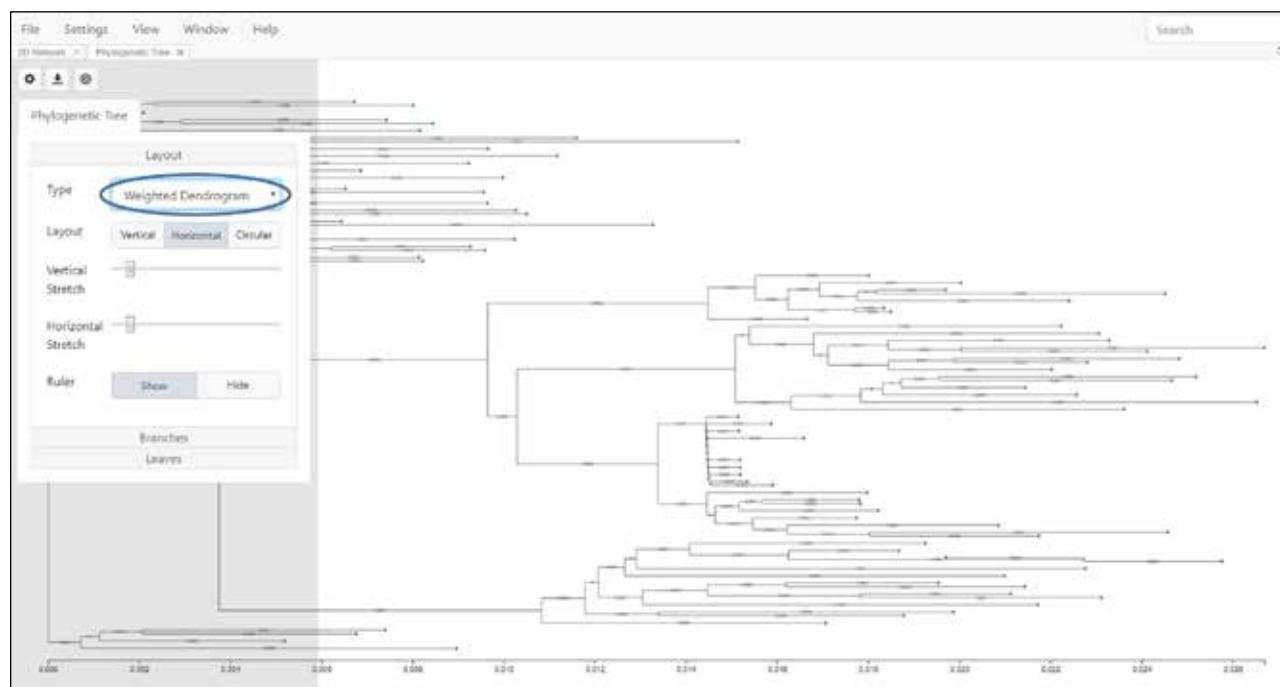
## 树选项卡

使用三个选项卡修改树设置:布局、动作和元

**布局:**选择布局(Layout)选项卡以更改树的类型、布局以及是否要显示标尺(比例尺)。您还可以垂直或水平拉伸树。当您有一棵包含许多分类群的大树时,此功能很有用。拉伸功能将分支展开,并允许更大的清晰度。

默认是垂直加权树状图,该树是一棵树枝长度按比例缩放的树。分支的长度与核苷酸取代的数量成正比(图 104)。树状图是表示系统发育树的图表,它显示了祖先与后代之间的关系。加权树状图推断出反映遗传距离矩阵中存在的结构的有根树。

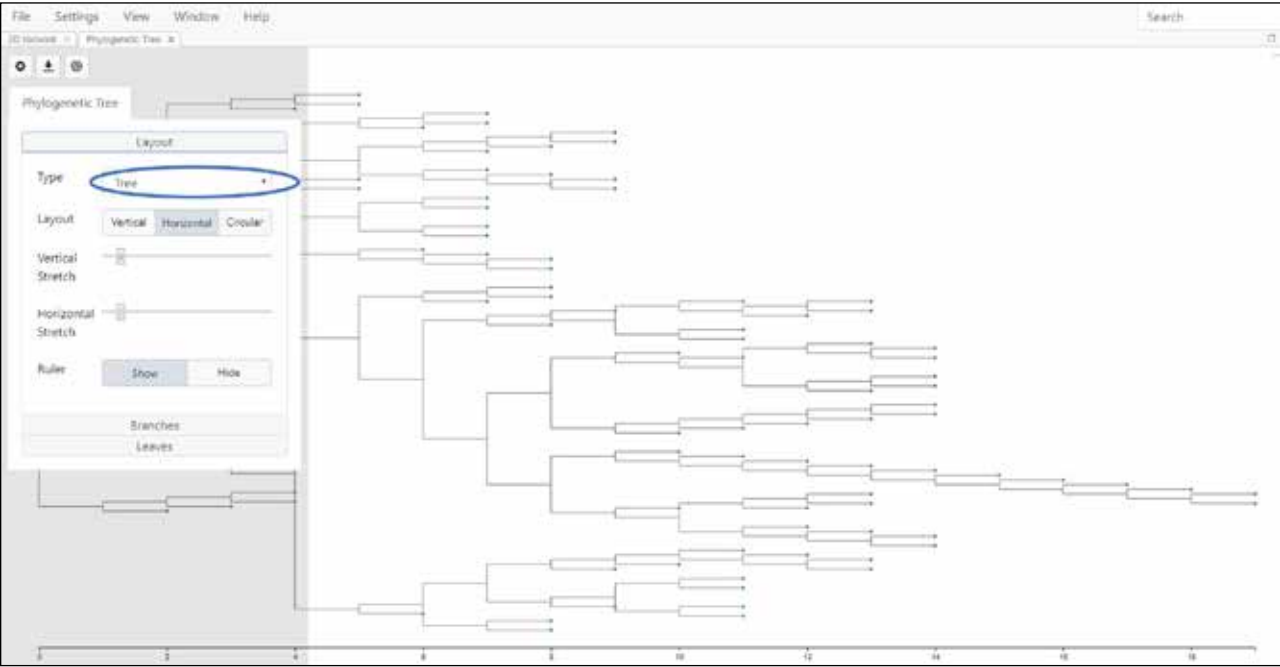
图 104. 选择显示比例尺的水平加权树状图



如果从下拉菜单中选择“树”,您将看到分支没有缩放,即分支长度是偶数,并且与核苷酸取代不成比例(图 105)。



图 105. 显示比例尺的水平树



如果您从下拉菜单中选择“未加权树状图(Unweighted Dendrogram)”，分支长度未按比例缩放，而距离有助于计算的每个分支长度平均值，并且分支的长度被调整，因此分类群(或有时称为叶子)在右侧比对(图 106)。

图 106. 未加权的树状图



您可以通过选择相应的树布局(Layout)选项将树布局更改为垂直(图 107)或圆形(图 108)。

图 107. 垂直加权树状图

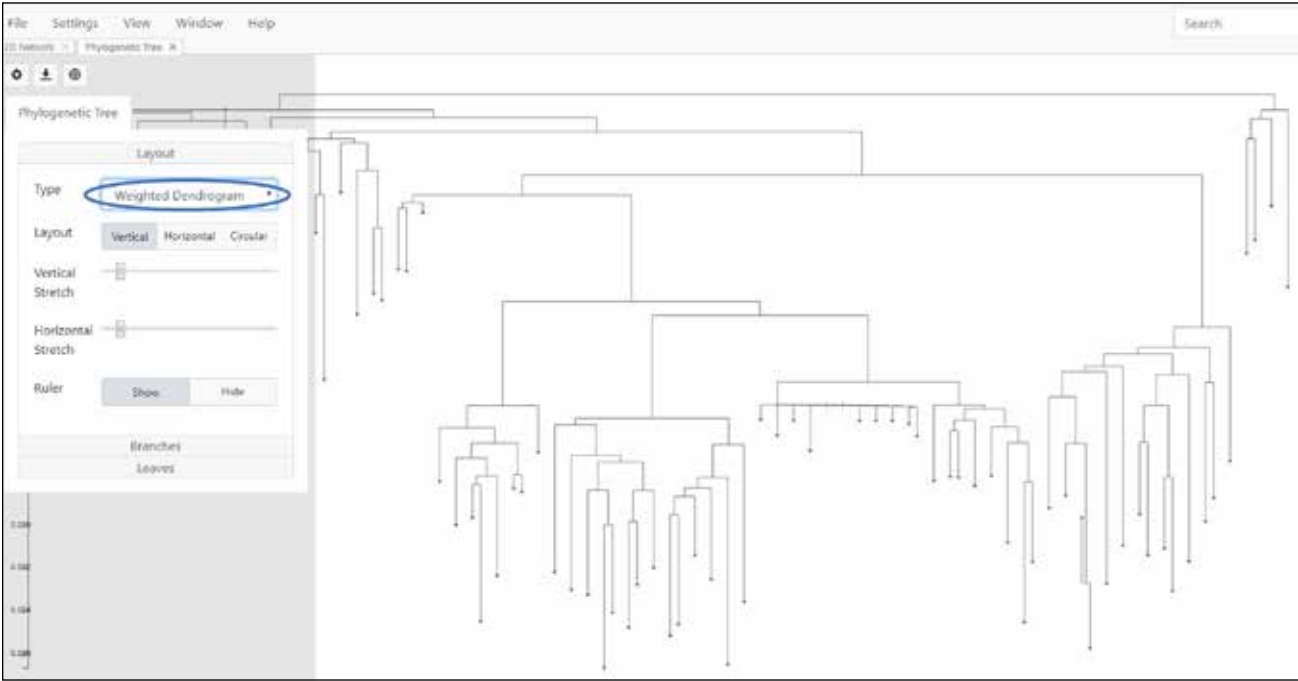
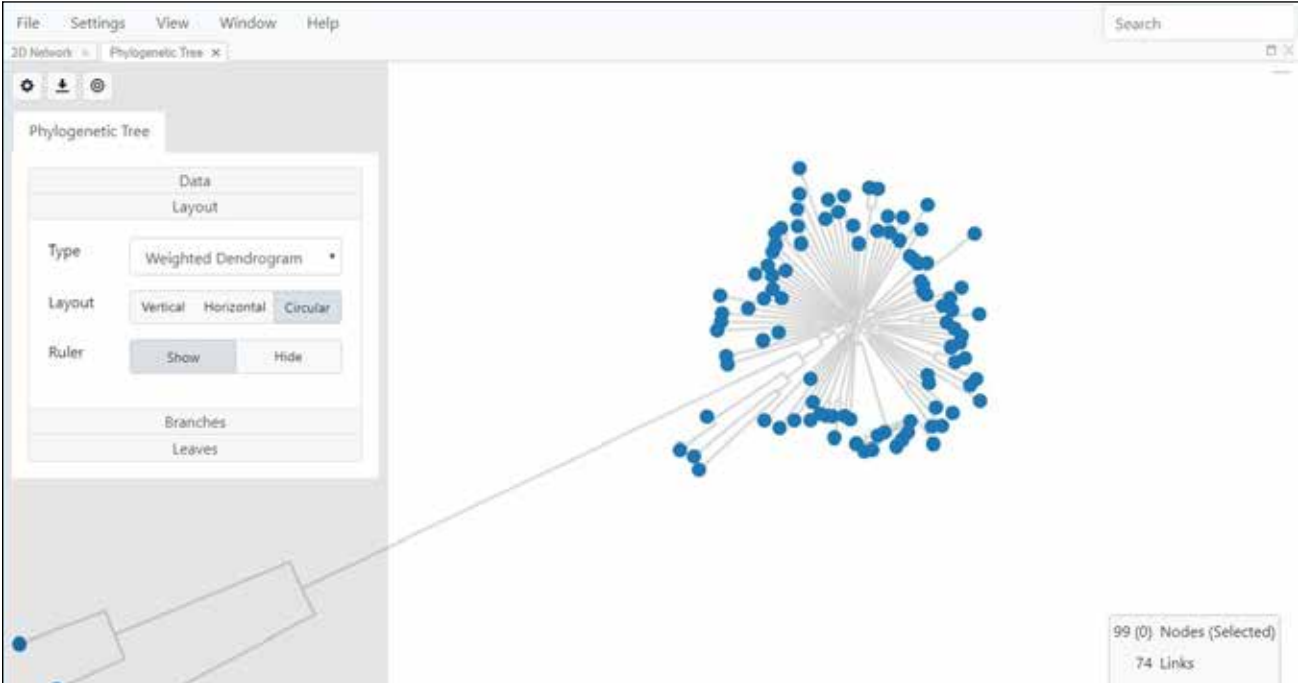


图 108. 圆形加权树状图



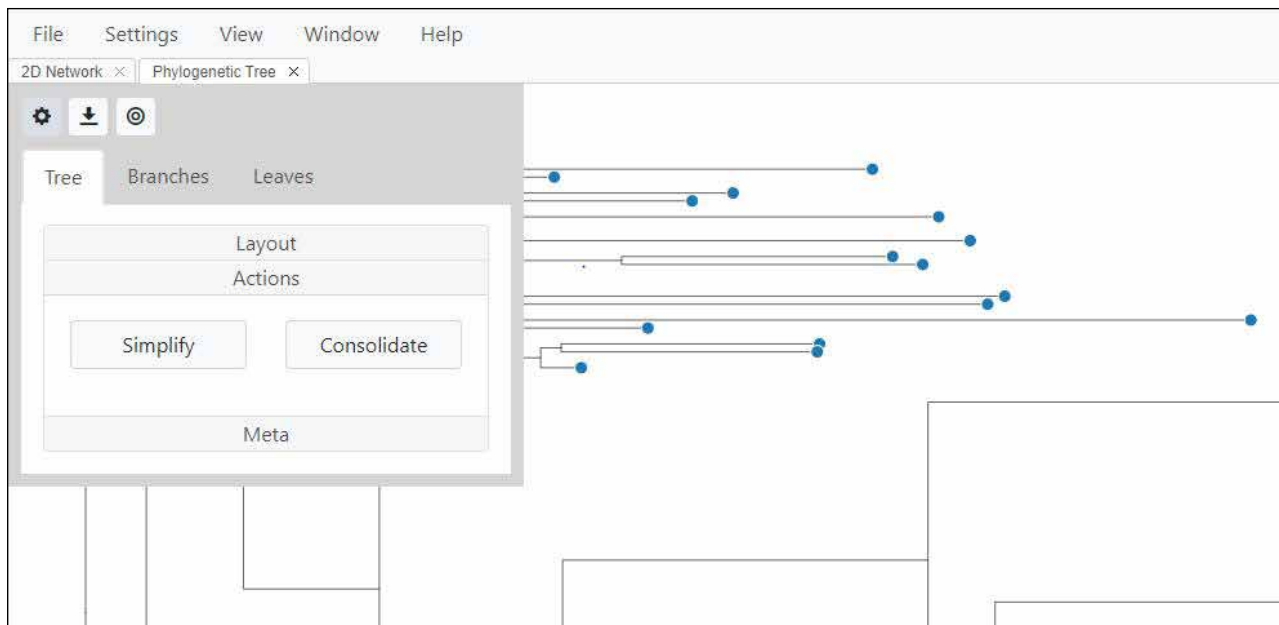
## 操作子选项卡

此选项卡允许您修改树枝, 以改进可视化(图 109)。

**简化:**将具有单个后代的分支折叠成一个长而连续的分支。

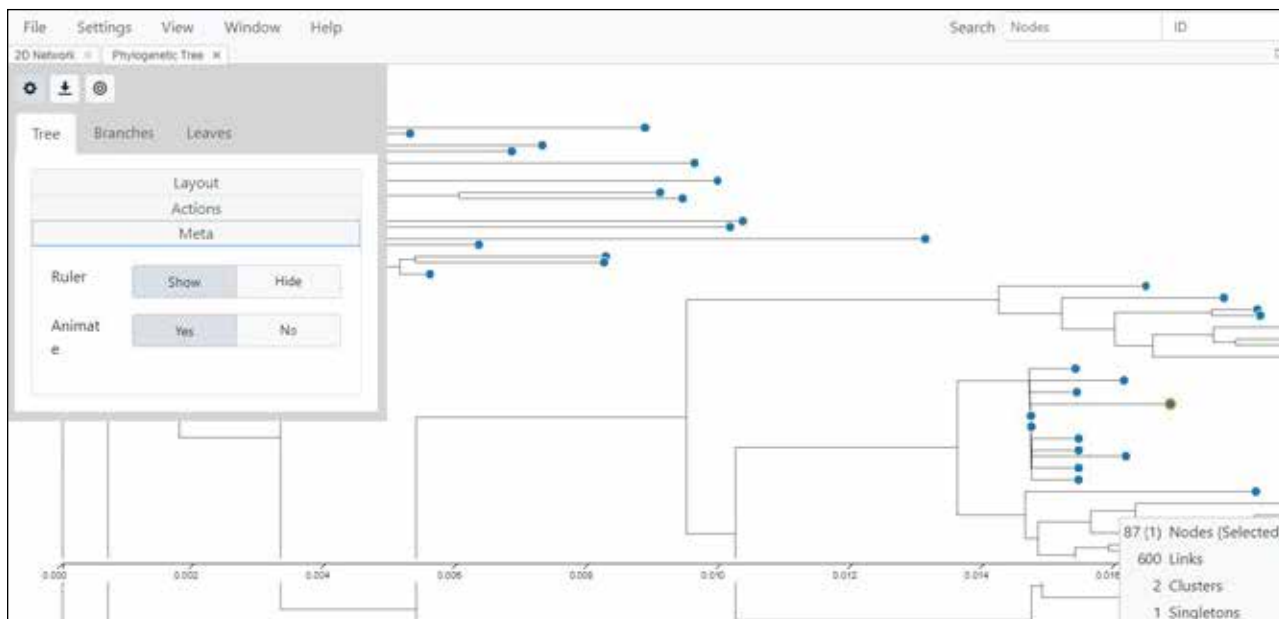
**巩固:**将距离为零的分支折叠成组合分支。

图 109. 操作(Actions)选项卡允许您修改后代分支的显示方式。



**元:**使用按钮(图 110)选择标尺显示和树配置更改的动画。

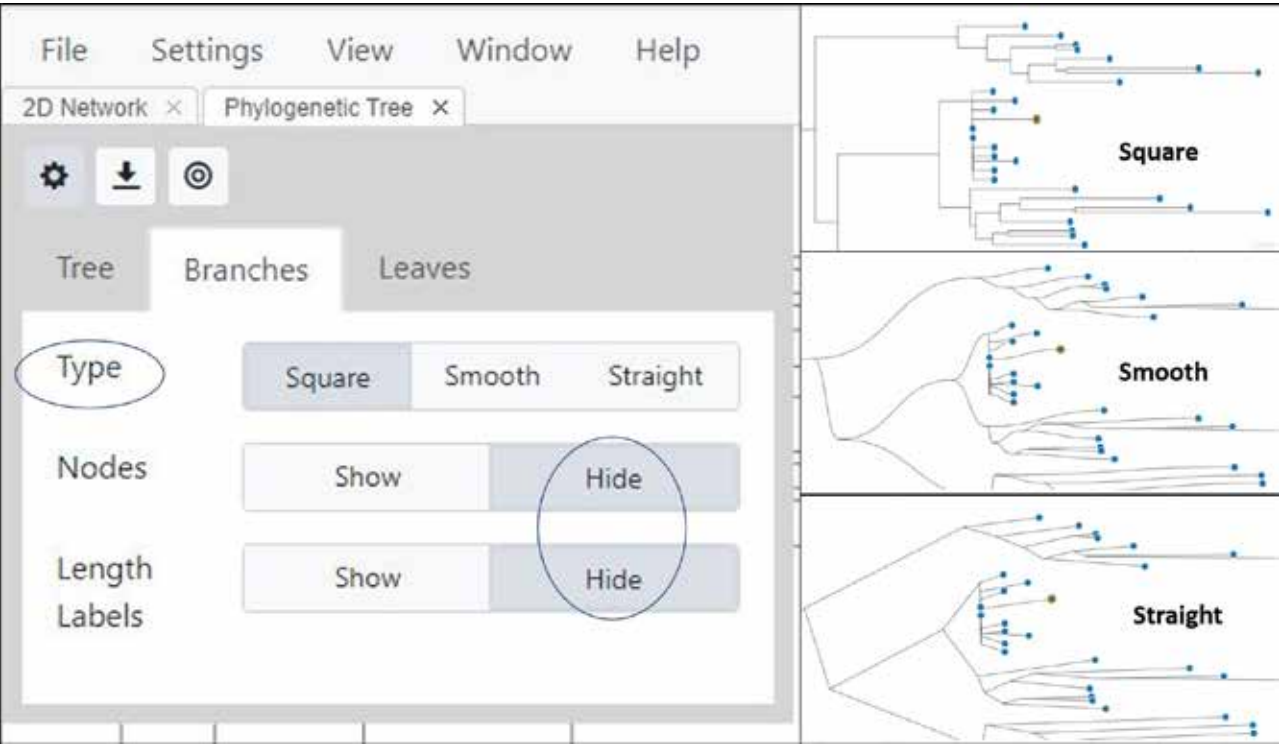
图 110. 元选项卡允许您更改标尺显示和动画



分支选项卡

此选项允许分支的外观被更改为方形、平滑或笔直(图 111)。您还可以选择是否显示或隐藏节点以及分支长度。这两个参数默认是隐藏的。

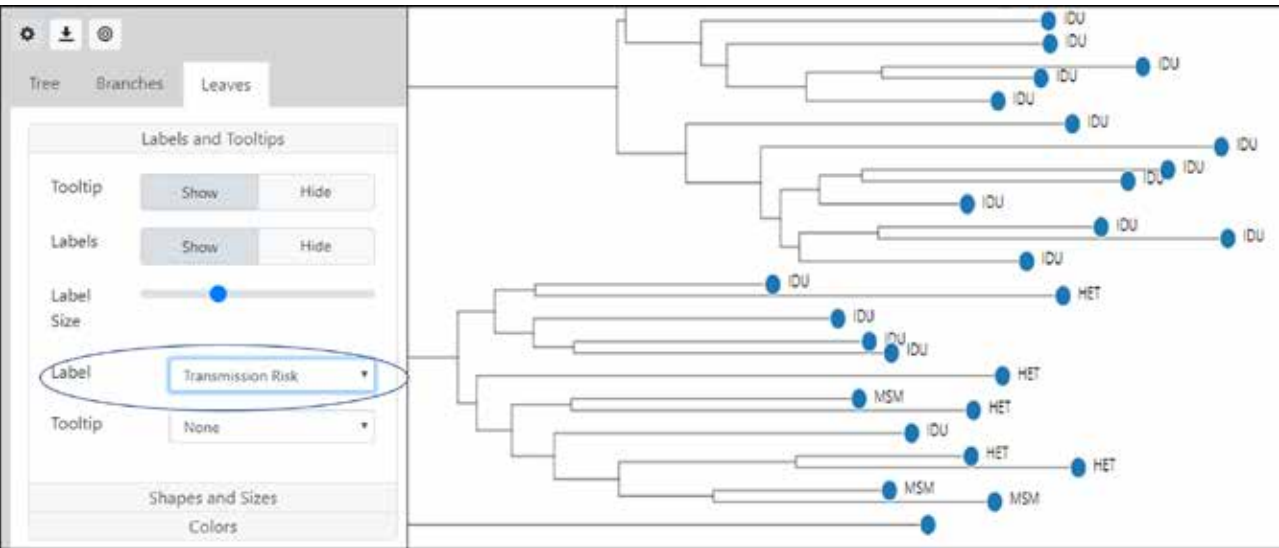
图 111. 分支选项卡选项



叶选项卡

**标签和工具提示:**此选项允许您选择是否显示标签和工具提示, 并允许您更改标签大小(图 112)。

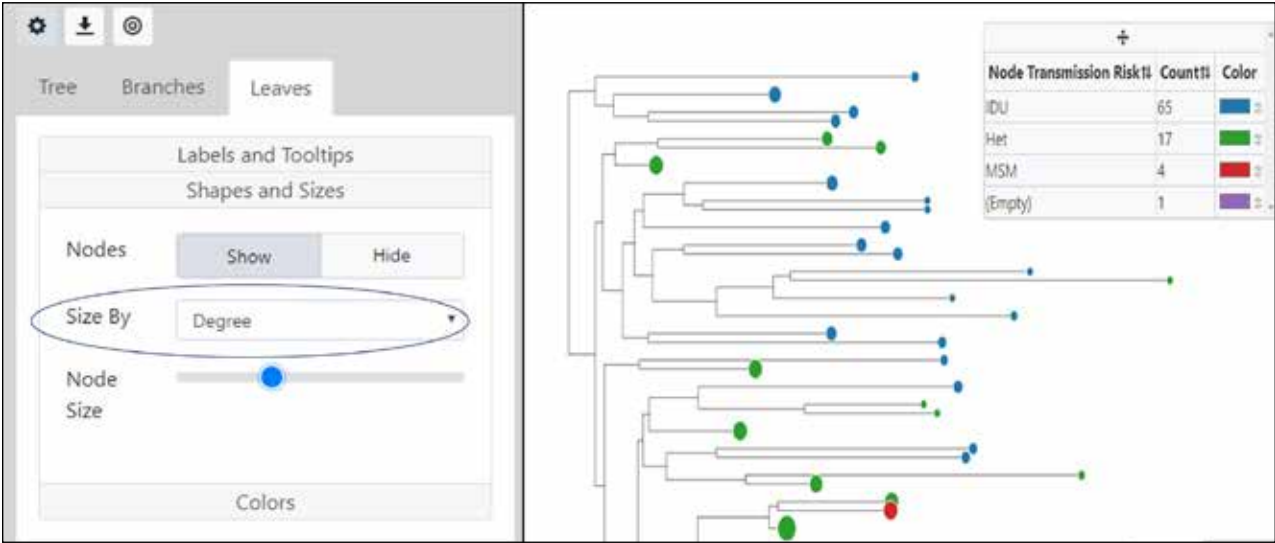
图 112. 叶选项卡、标签和工具提示。叶节点被标记为传输风险。



**形状和尺寸:**此选项卡允许您通过节点列表中的任何变量来调整节点大小, 可从下拉菜单中获得(图 113)。

**颜色:**将您带到全球设置的样式选项卡, 您可以在其中按您选择的变量进行着色。

图 113. 节点按度数大小, 按传输风险着色; 节点标签被隐藏



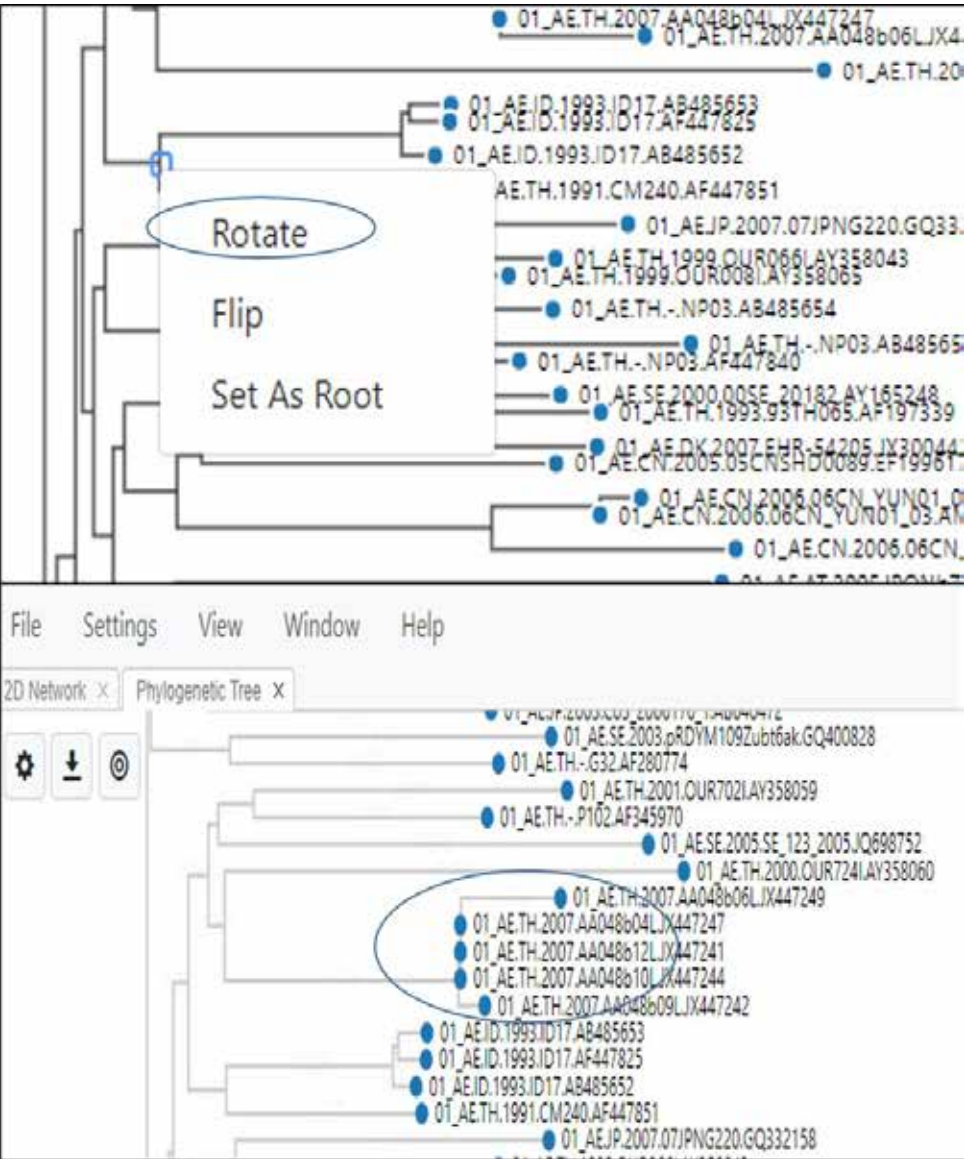
**颜色:**将您带到全球设置的样式选项卡, 您可以在其中按您选择的变量进行着色。系统发育视图 (Phylogenetic View) 右上角的搜索框允许通过分类单元 (序列) 名称或 ID 对树进行搜索。您可以在树中搜索下拉菜单中可用的任何变量。然后分类单元名称在树上以蓝色突出显示。如果您在一棵大树中有许多分类单元, 并且需要定位单个分类单元, 则此功能很有用。

节点选项

右键点击树中的节点, 以查看可用的各种选项, 包括旋转、重新生根和删除分支。有关这些功能的示例, 请参见下图。

**旋转:**此选项用于沿树中选定节点的轴旋转分支。

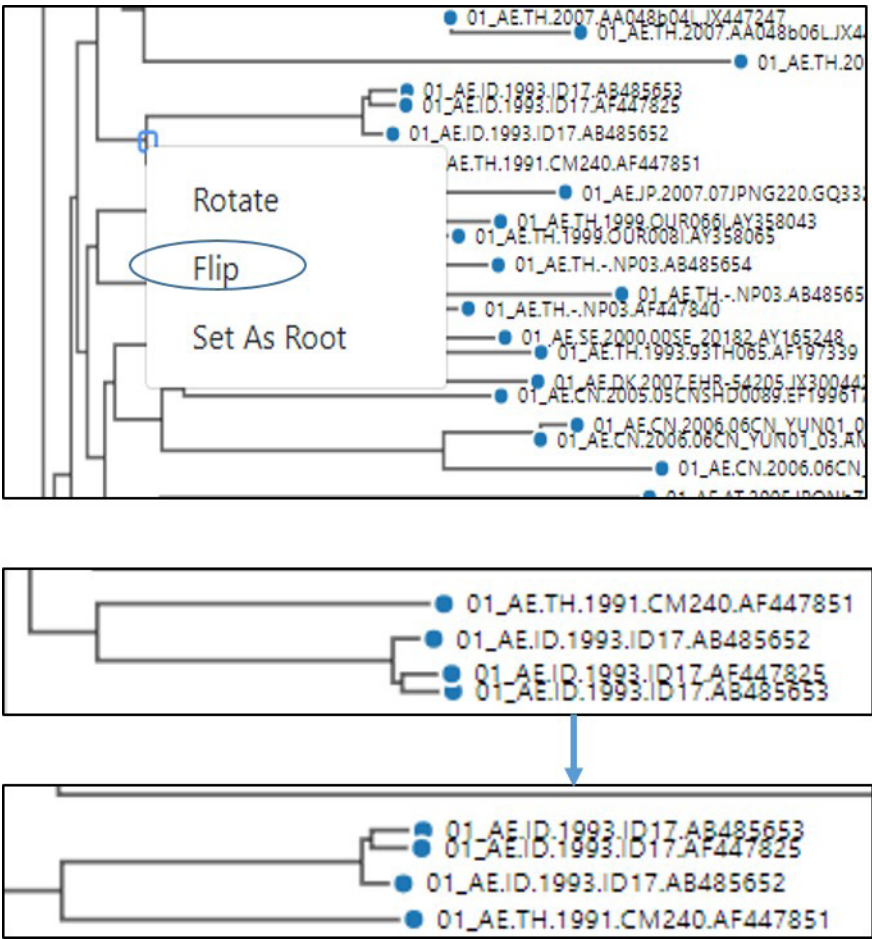
图 114. 系统发育树视图显示如何沿选定节点的轴旋转进化枝 (序列簇)。





**翻转:**这会沿节点的轴翻转进化枝, 因此位于底部的进化枝现在将翻转到顶部。

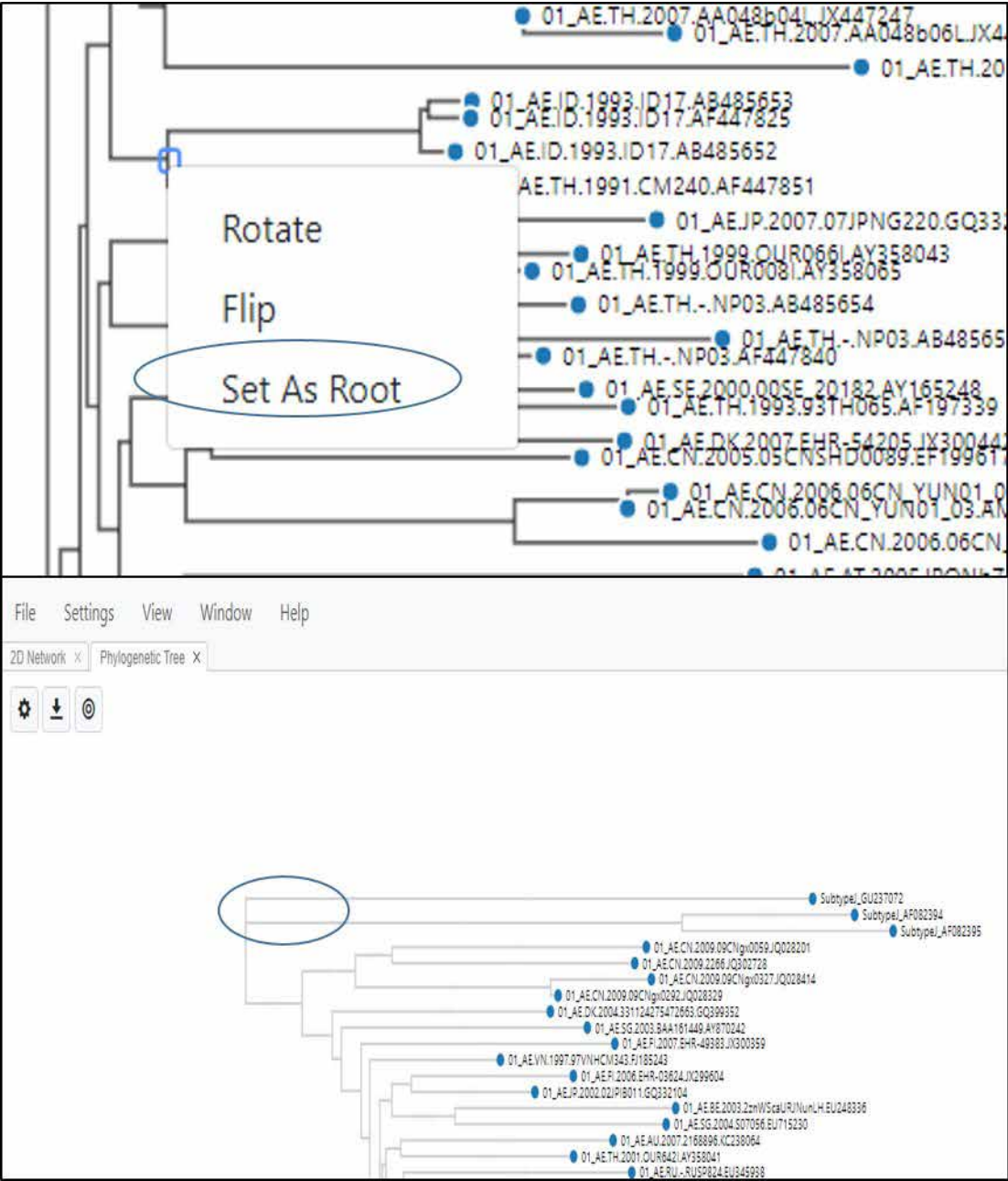
图 115. 系统发育树视图显示如何沿选定节点的轴翻转进化枝 (序列簇)。



**设置为根:**仅当您具有有关分类群潜在进化历史的先验信息或分析中包含外群时, 才使用此选项。外群用于为树“植根”。外群是一组与您要分析的分类群接近但又不同的分类群。例如, HIV-1 亚型 J 可用作分析 HIV-1 亚型 CRF01\_AE 序列的外群 (图 116)。请注意, 没有用于重新扎根的撤消选项。如果您在某个节点上重新扎根, 并想回到原来的树状结构, 请关闭窗口, 从查看 (View) 菜单中选择系统发育树 (Phylogenetic Tree), 以显示原来的树状结构。



图 116. 系统发育树视图显示如何在选定节点处重新生根树。

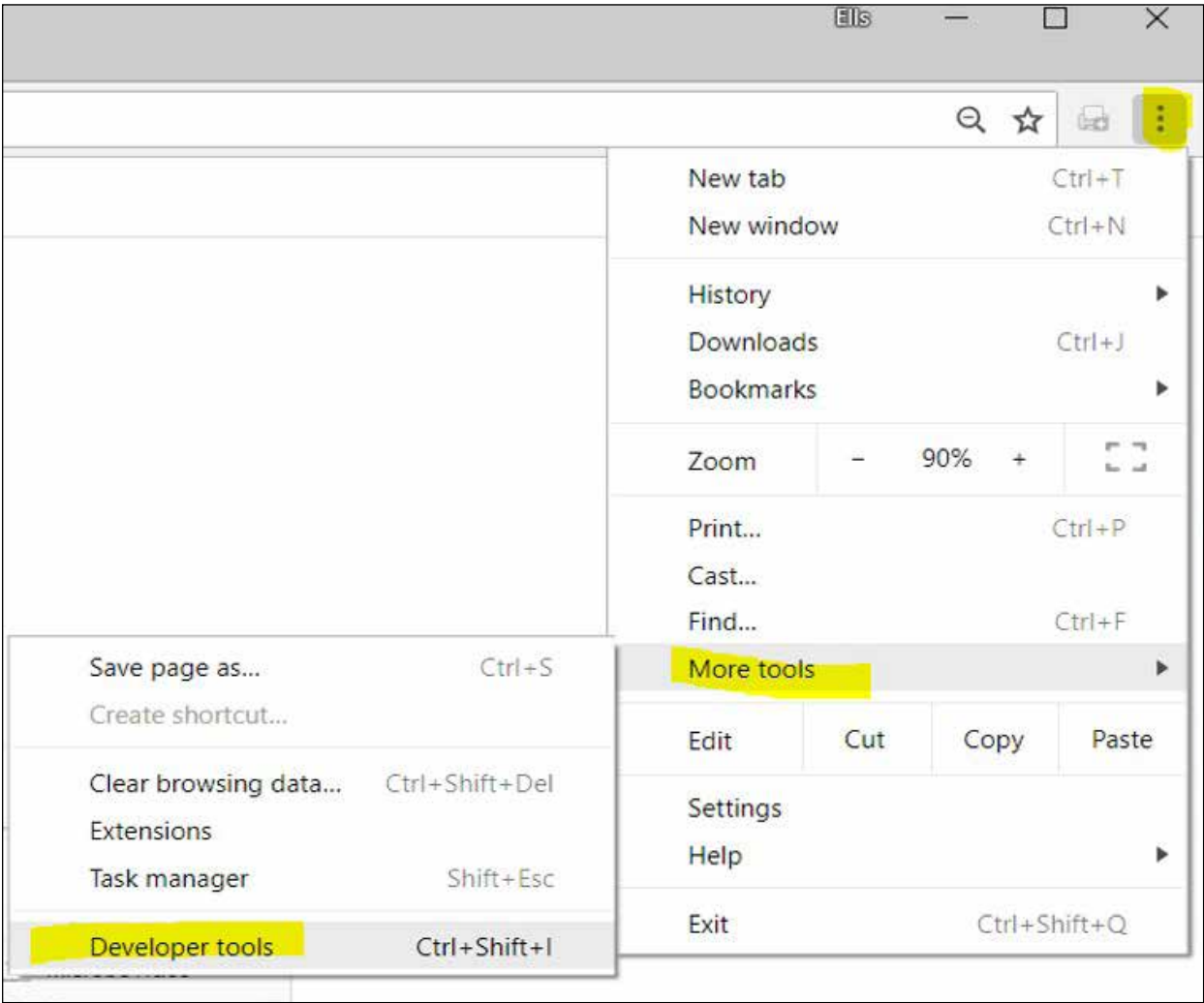


# 故障排除

如果您在加载数据文件时遇到问题, 如果您观察到异常行为, 或者如果 MicrobeTrace 冻结或挂起, 您可能需要通过使用以下步骤清除浏览器缓存。

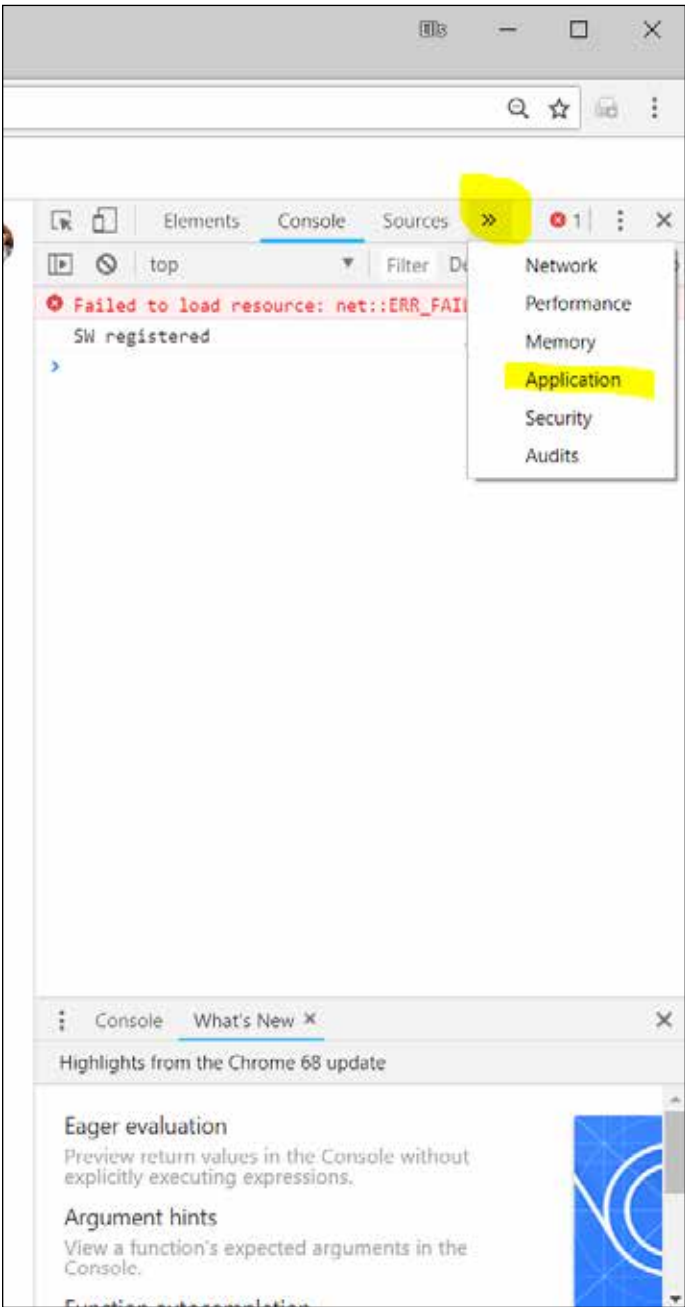
在 Chrome 中, 通过点击三个垂直点打开开发者工具(图 117); 选择**更多 (More)** 工具, 然后选择开发人员 (Developer) 工具。

图 117. 在开发者 (Developer) 工具窗口中清除浏览器缓存



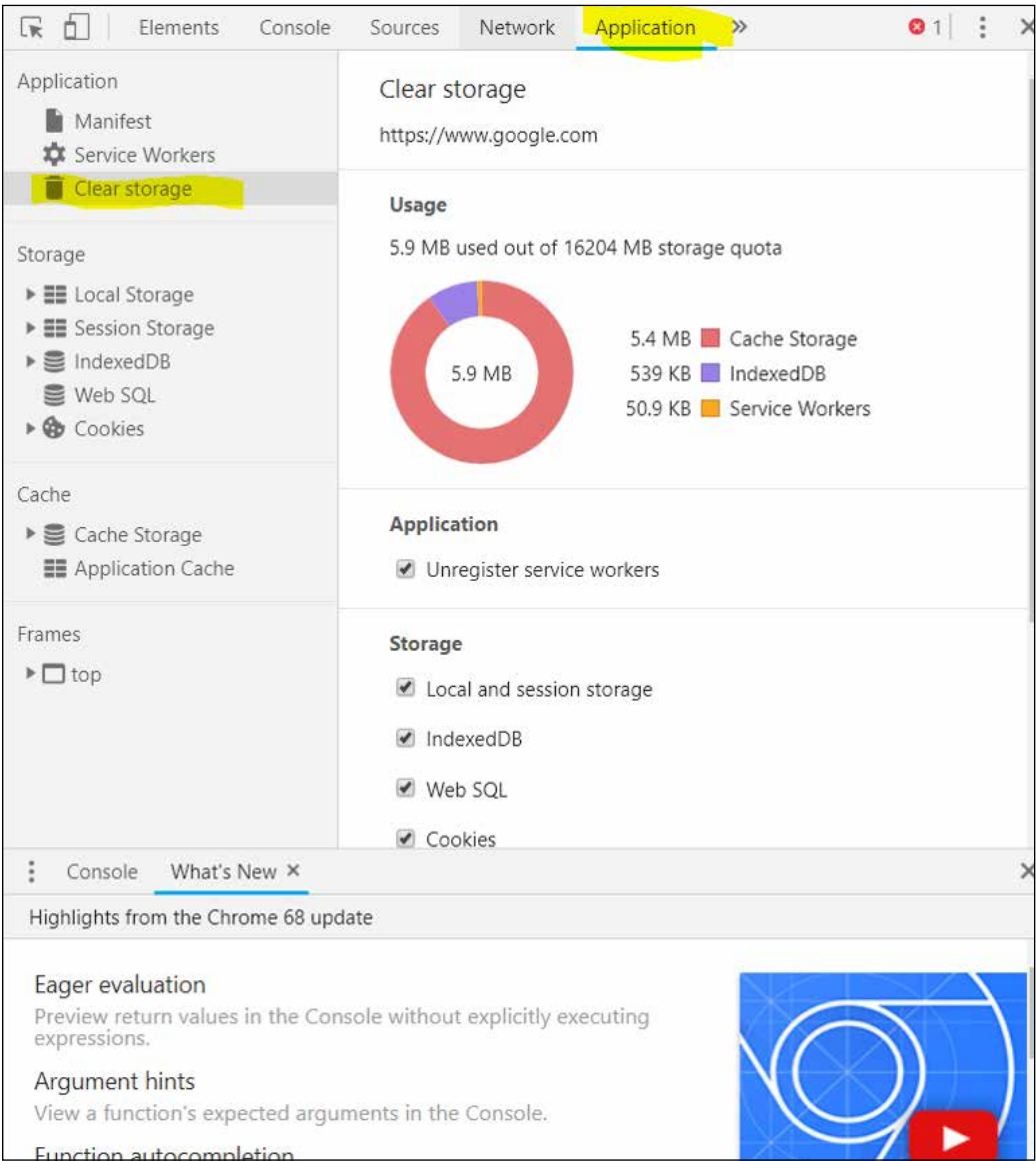
选择 >> (图 118), 然后从下拉菜单中选择应用程序。

图 118. 通过应用程序选项清除浏览器缓存



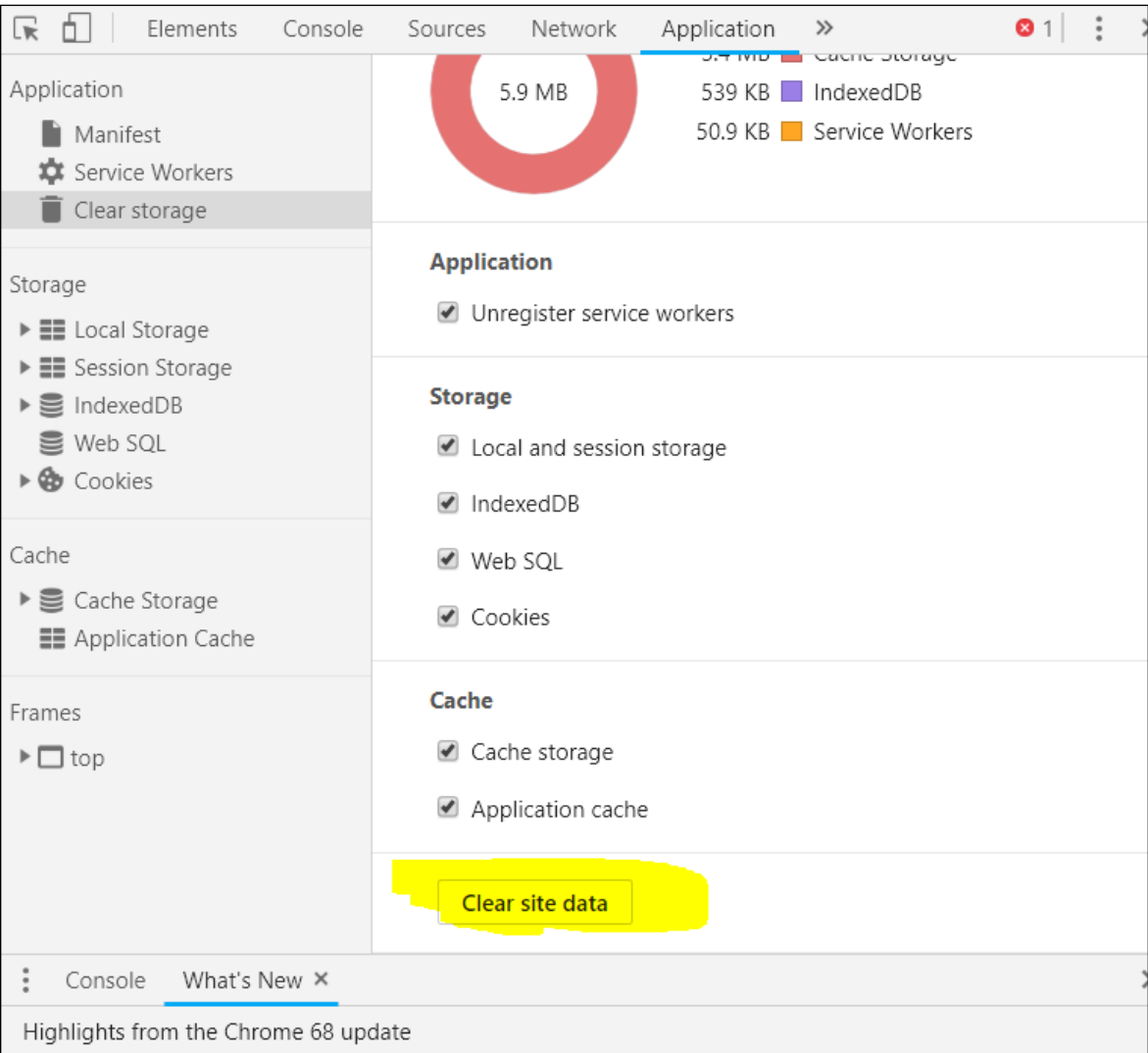
进入应用程序选项卡后, 选择清除存储(图 119)

图 119. 使用清除存储选项清除浏览器缓存



最后, 滚动到该子选项卡的底部, 然后选择清除站点数据(图 120)。

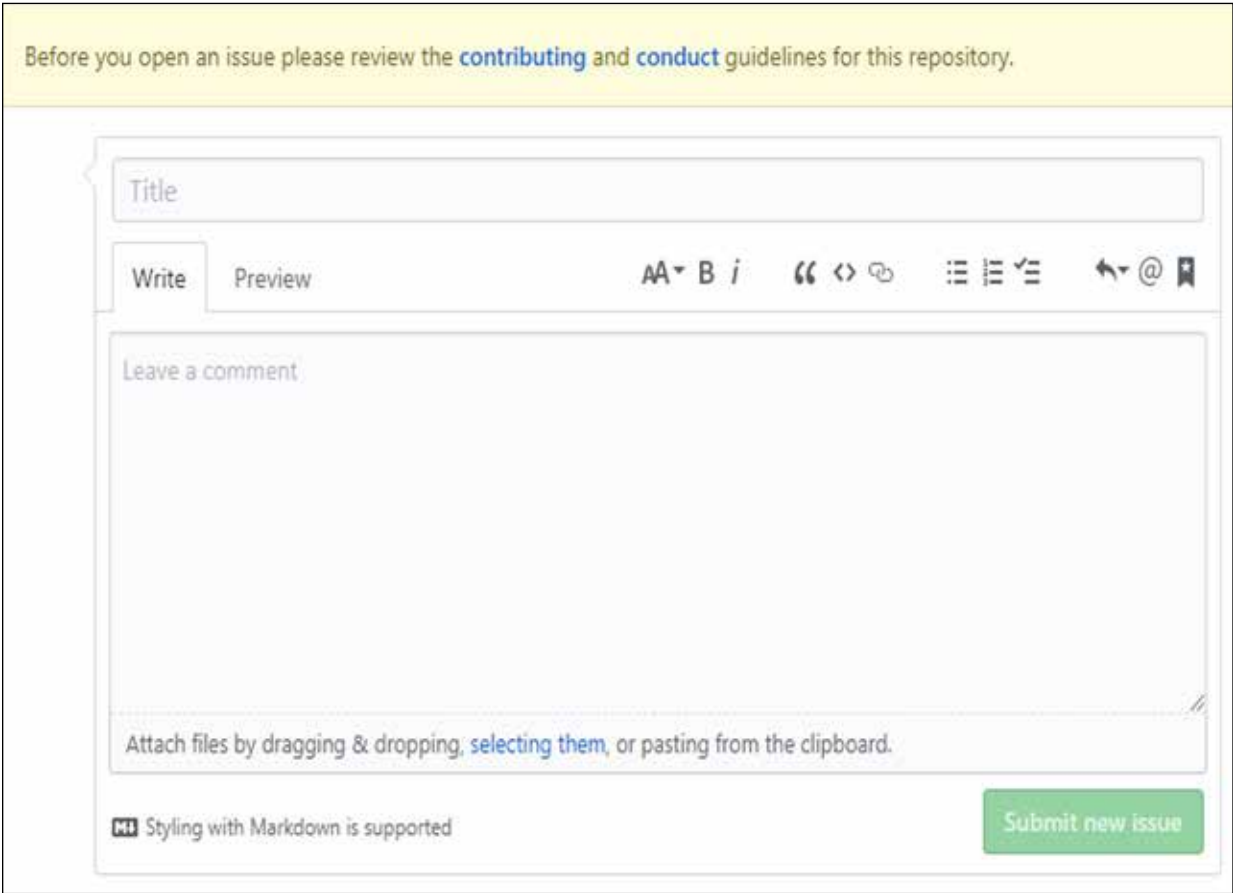
图 120. 清除浏览器缓存, 以清除站点数据



这一系列步骤将清除缓存。如果清除浏览器缓存不能解决您的问题, 您可以使用 MicrobeTrace GitHub 站点报告和追踪任何软件问题。要报告任何 MicrobeTrace 问题, 请使用链接访问 MicrobeTrace“新问题”网站 <https://github.com/CDCgov/MicrobeTRACE/issues/new>。访问此站点需要 GitHub 帐户。如果您没有 GitHub 帐户, 系统会提示您创建一个。GitHub 帐户是免费的, 访问只需要一个电子邮件地址和一个自行生成的密码。

在 MicrobeTrace 问题网站上, 系统会提示您填写一份表, 要求提供问题的标题和描述(图 121)。

图 121. 将问题报告给 MicrobeTrace 团队



The screenshot shows the GitHub 'New issue' form for the repository 'CDCgov/MicrobeTRACE'. At the top, a yellow banner reads: 'Before you open an issue please review the [contributing](#) and [conduct](#) guidelines for this repository.' Below this is a form with a 'Title' input field. Underneath the title field are two tabs: 'Write' (selected) and 'Preview'. To the right of the tabs is a rich text editor toolbar with icons for bold (AA), italic (i), link (link icon), unlink (unlink icon), list (bulleted list icon), ordered list (numbered list icon), and quote (quote icon). Below the tabs is a large text area with the placeholder text 'Leave a comment'. At the bottom of the text area, it says 'Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.' In the bottom left corner, there is a small icon and the text 'Styling with Markdown is supported'. In the bottom right corner, there is a green button labeled 'Submit new issue'.

在标题中包含您正在使用的软件程序中的视图。请在评论部分提供尽可能多的有关问题的详细信息, 包括使用的操作系统。您也可以使用此表告诉我们如何改进软件。

例如, 第 27 期的标题为“组的钥匙表无法滚动到页面底部”。虽然清楚该错误是如何违反了用户期望的, 但该声明并未描述“组密钥表”是什么, 或者在使用 MicrobeTrace 时可以在哪里找到它。相比之下, 第 #125 问题“将距离矩阵从 TN93 切换到 SNP 的选项”描述了在哪里遇到问题(在距离矩阵中), 以及需要什么(在用于分析的 TN93 核苷酸替换模型和可用但未使用的 SNP 之间切换的选项)。

您还可以发送电子邮件, 通过 [microbetrace@cdc.gov](mailto:microbetrace@cdc.gov) 向 CDC MicrobeTrace 支持代表请求帮助。在起草您的电子邮件时, 请尽可能详尽, 列出您在遇到问题之前在程序中采取的每项操作。您提供的细节越多, 我们解决问题的可能性就越大。



## 参考文献

1. Ellsworth M. Campbell, Anthony Boyles, Anupama Shankar, Jay Kim, Sergey Knyazev, William M. Switzer. MicrobeTrace: Re-tooling Molecular Epidemiology for Rapid Public Health Response, PLOS Computational Genomics. 手稿已接受。预印本在此: <https://www.biorxiv.org/content/10.1101/2020.07.22.216275v1>
2. Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRANsmiSSion Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol Biol Evol.* 2018 Jul 1;35(7).
3. Guy Yachdav, Sebastian Wilzbach, Benedikt Rauscher, Robert Sheridan, Ian Sillitoe, James Procter, Suzanna E. Lewis, Burkhard Rost, Tatyana Goldberg; MSAViewer: interactive JavaScript visualization of multiple sequence alignments, *Bioinformatics*, Volume 32, Issue 22
4. Hightower GK, May SJ, Perez-Santiago J, Pacold ME, Wagner GA, Little SJ, et al. HIV-1 clade B pol evolution following primary infection. *PLoS ONE.* 2013; 8(6):e68188. doi: 10.1371/journal.pone.0068188.
5. Oster AM, France AM, Mermin J. Molecular Epidemiology and the Transformation of HIV Prevention. *JAMA.* 2018;319(16):1657–1658. doi:10.1001/jama.2018.
6. Barré-Sinoussi F, Abdool Karim SS, Albert J, Bekker LG, Beyrer C, Cahn P, Calmy A, Grinsztejn B, Grulich A, Kamarulzaman A, Kumarasamy N, Loutfy MR, El Filali KM, Mboup S, Montaner JS, Munderi P, Pokrovsky V, Vandamme AM, Young B, Godfrey-Faussett P. Expert consensus statement on the science of HIV in the context of criminal law. *J Int AIDS Soc.* 2018;21(7):e25161. doi: 10.1002/jia2.25161.

## 致谢

MicrobeTrace 团队由 Tony Boyles、Ellsworth Campbell、Anupama Shankar、Jay Kim、Evan Moscoso、Francis Ambrosio、Roxana Cintron、Sergey Kniazev 和 Bill Switzer 组成。我们感谢 HIV-TRACE 的开发人员在 HIV 传播研究和软件开发方面的开创性工作, 包括 Joel Wertheim、Sergei Kosakovsky Pond 和 Steven Weaver。我们还要感谢 MicrobeTrace 所依赖的所有开源项目的开发者。最后, 我们感谢众多 beta 测试人员的投入。特别感谢 CDC 的结核病消除流行病学小组 (Sarah Talarico、Kathryn Winglee、Ben Silk) 为我们提供了非常宝贵的意见, 从而促成了许多新功能和改进功能的开发。我们非常感谢 CDC 的高级分子检测计划和 HIV 预防部门为 MicrobeTrace 的开发提供资金。

