

SAtoolkit V1.0

用户手册

一、 软件简介

SAtoolkit V1.0 软件是一款借助 C 语言和 Python3 语言编写的生物序列（核酸序列和氨基酸序列）比对的工具包，通过自己的实现逻辑，集成三个主流的生物序列比对算法。用户通过简单的操作，即可将自己的生物序列通过 SAtoolkit 软件快速得到分析，并输出相似序列的位置等信息，并提供丰富的图形展示。SAtoolkit V1.0 软件是一个集高效与便捷于一体的软件。

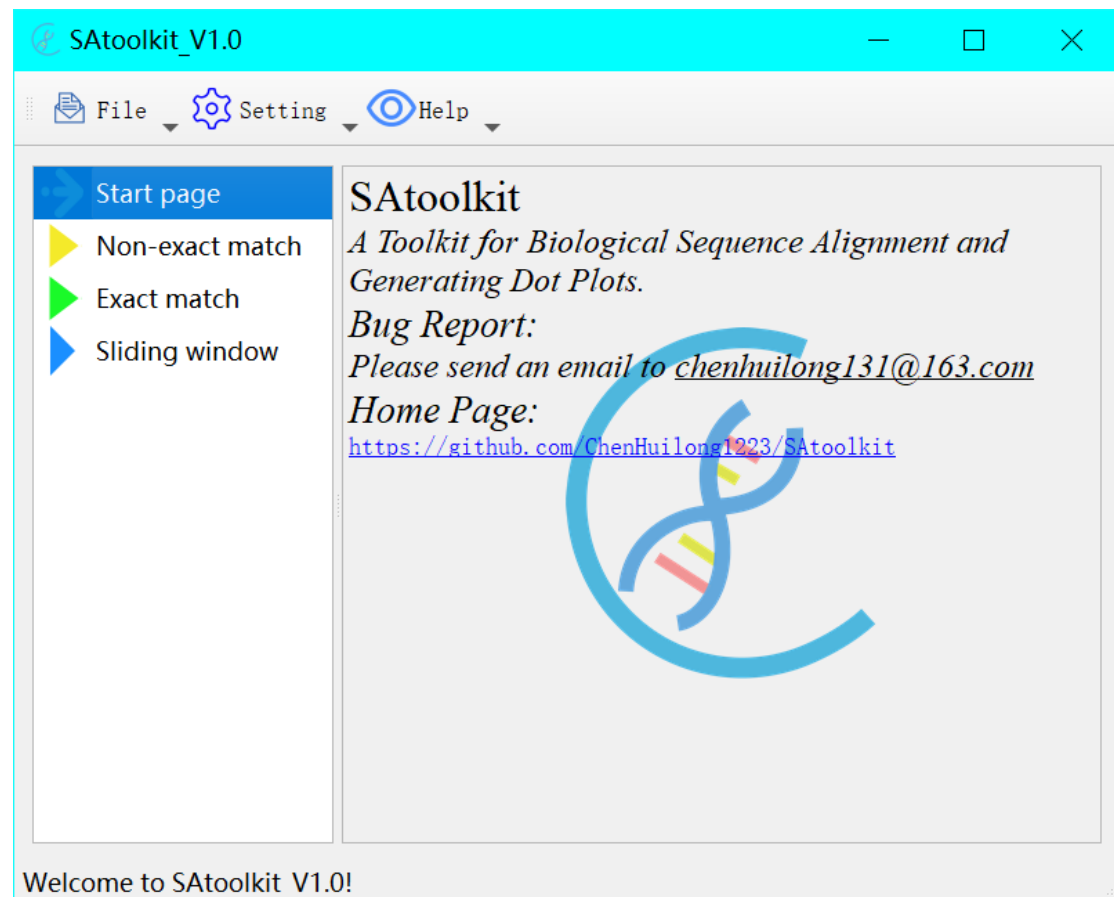
二、 编写目的

该用户手册是用户使用 SAtoolkit V1.0 软件的操作说明，用户通过阅读用户手册，能够直接对 SAtoolkit V1.0 软件的功能、操作有一定的了解，按照用户手册上的说明，通过实际操作，用户能够迅速掌握 SAtoolkit V1.0 软件的使用方法。

三、 系统操作

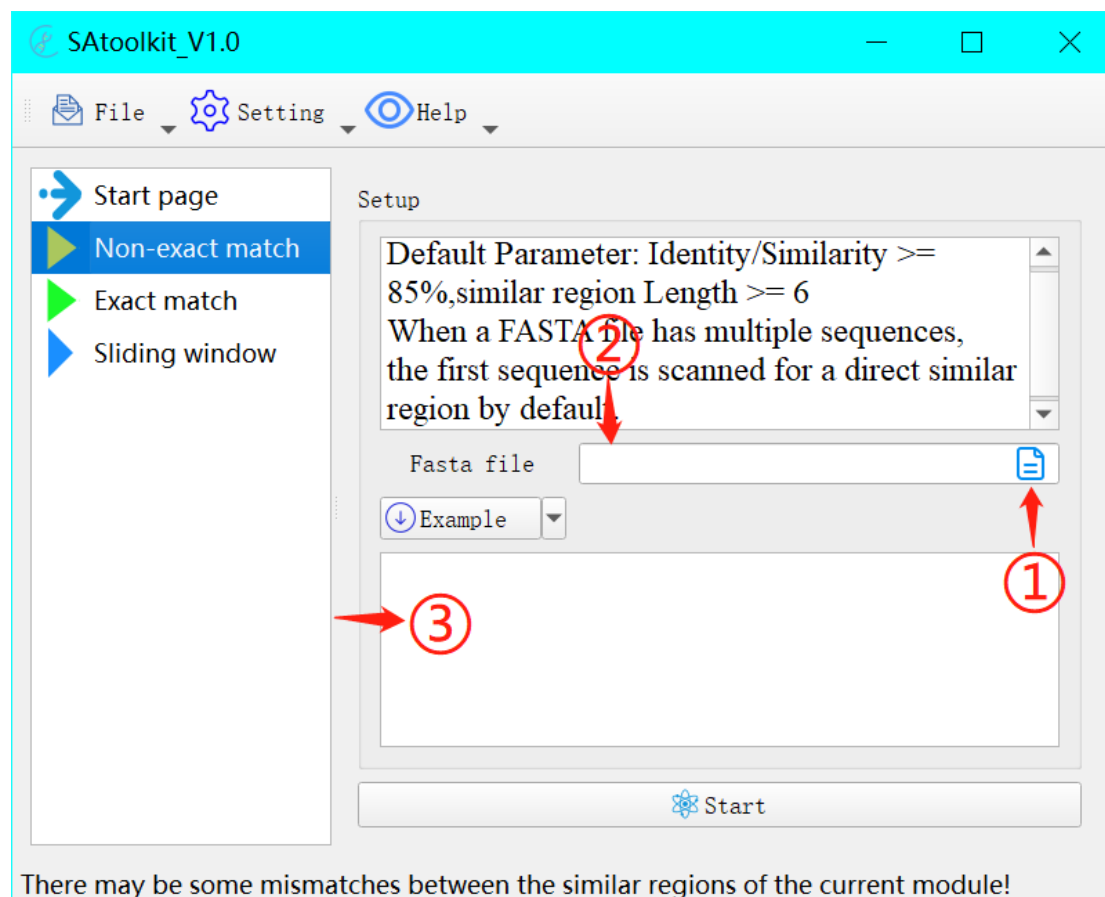
3.1 首页

点击桌面上 SAtoolkit V1.0 的图标即可进入到 SAtoolkit 软件的首页界面中如下图所示：



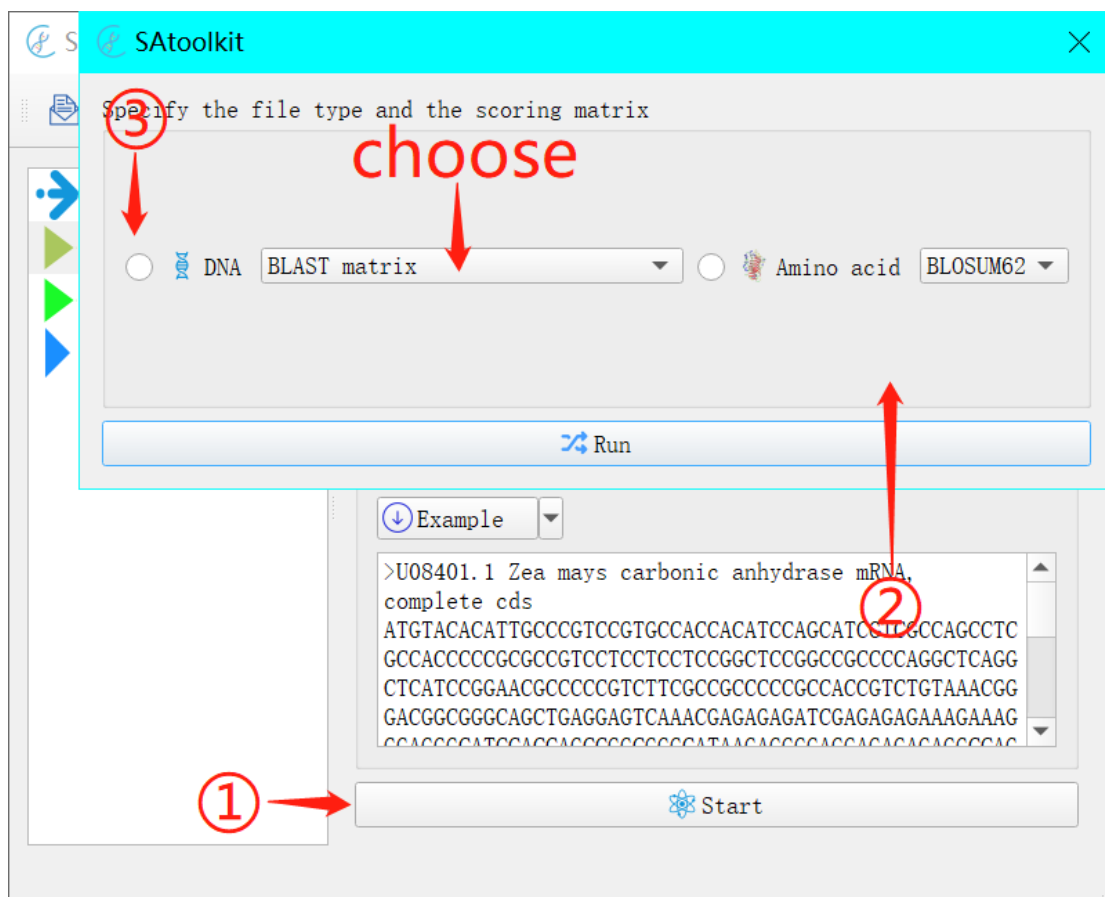
3.2 操作

使用 SAtoolkit V1.0 的文件格式为标准的 fasta 文件。正如下图所示，SAtoolkit V1.0 有三种输入文件的方式：①点击文件图标，在用户电脑对应的位置进行选择。②可以把文件路径直接输入到文本框中，或者直接把用户想要分析的文件拖进文本框中，自动识别文件路径并添加到文本框中。③直接把用户想要分析的 fasta 格式序列粘贴进 Example 文本框中。

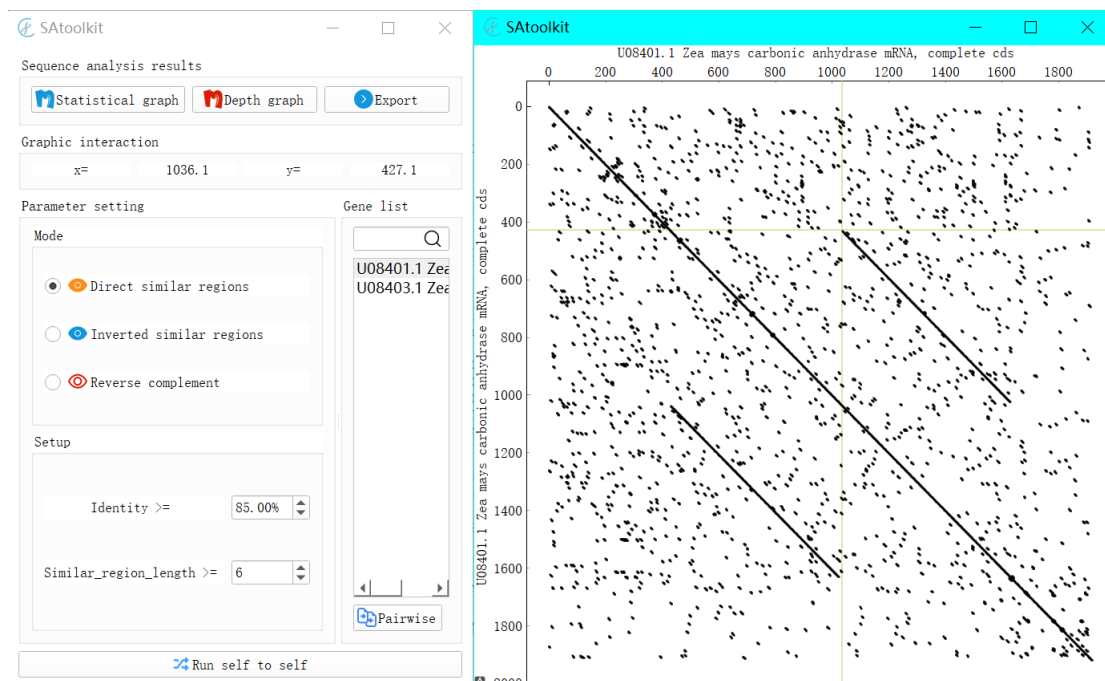


SAtoolkit 集成了 CyDotian 算法（Non-exact-match 和 Exact match）和 sliding window 算法进行序列比对。全部基于自己的逻辑进行实现。分为三种模式。以下以第一种 Non-exact match 为例进行软件操作的详细讲演。

该模式是基于 CyDotian 算法实现的允许序列比对有错配的序列比对效果，用户根据自己的喜好输入正确的 fasta 文件后，点击 Start 按钮，进入到指定用户序列类型的对话框。根据用户输入的序列类型指定对应的类型（DNA 或者氨基酸）。DNA 支持两种矩阵（BLAST 和 Transition-transversion 矩阵）。氨基酸支持 BLOSUM45、BLOSUM62、BLOSUM80、BLOSUM90 、 PAM30 、 PAM70 和 PAM250 矩阵。



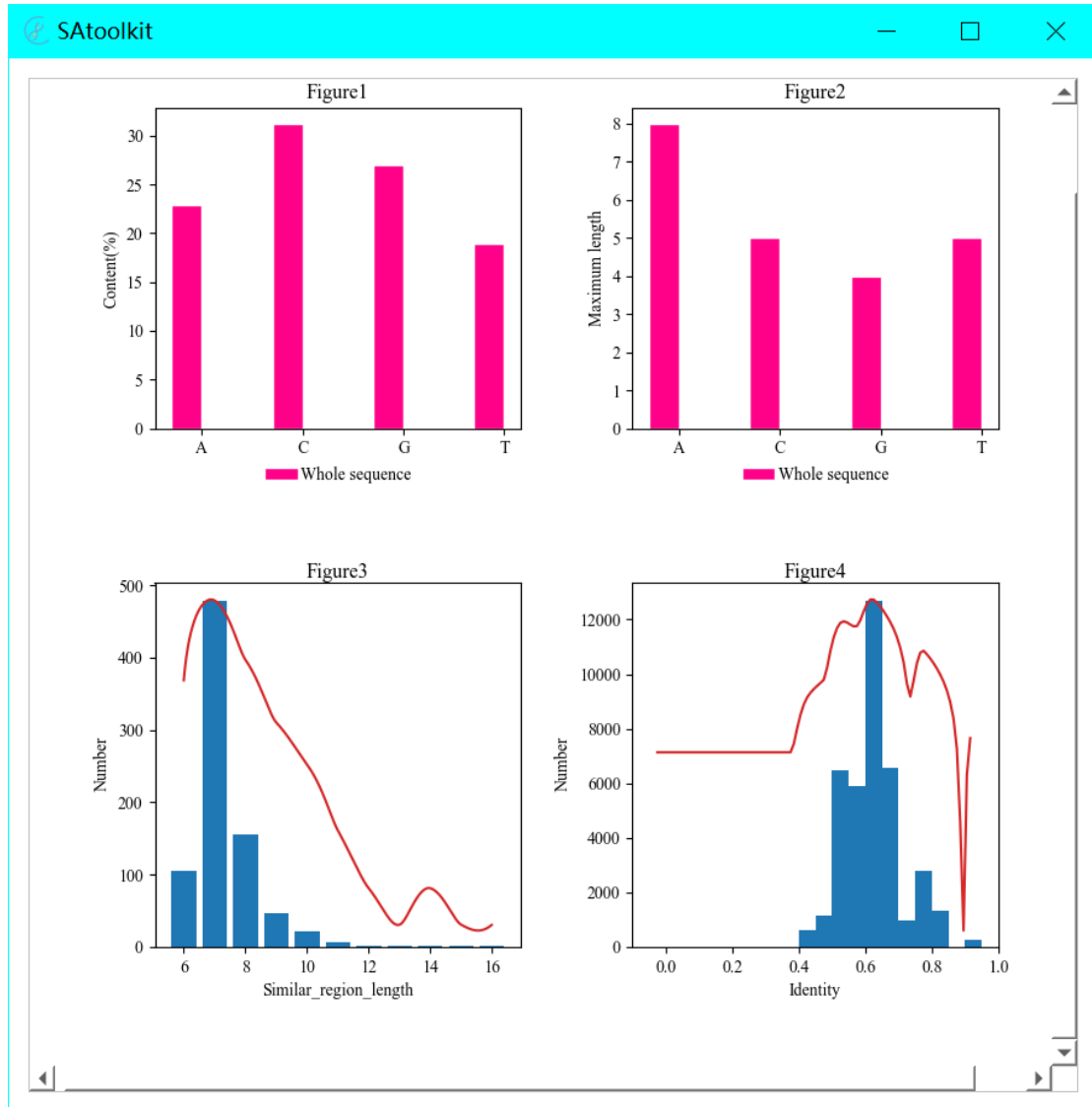
选择好对应的矩阵之后，点击 **Run** 按钮，将会弹出一个控制面板（左边）和一个点阵图窗口（右边）。



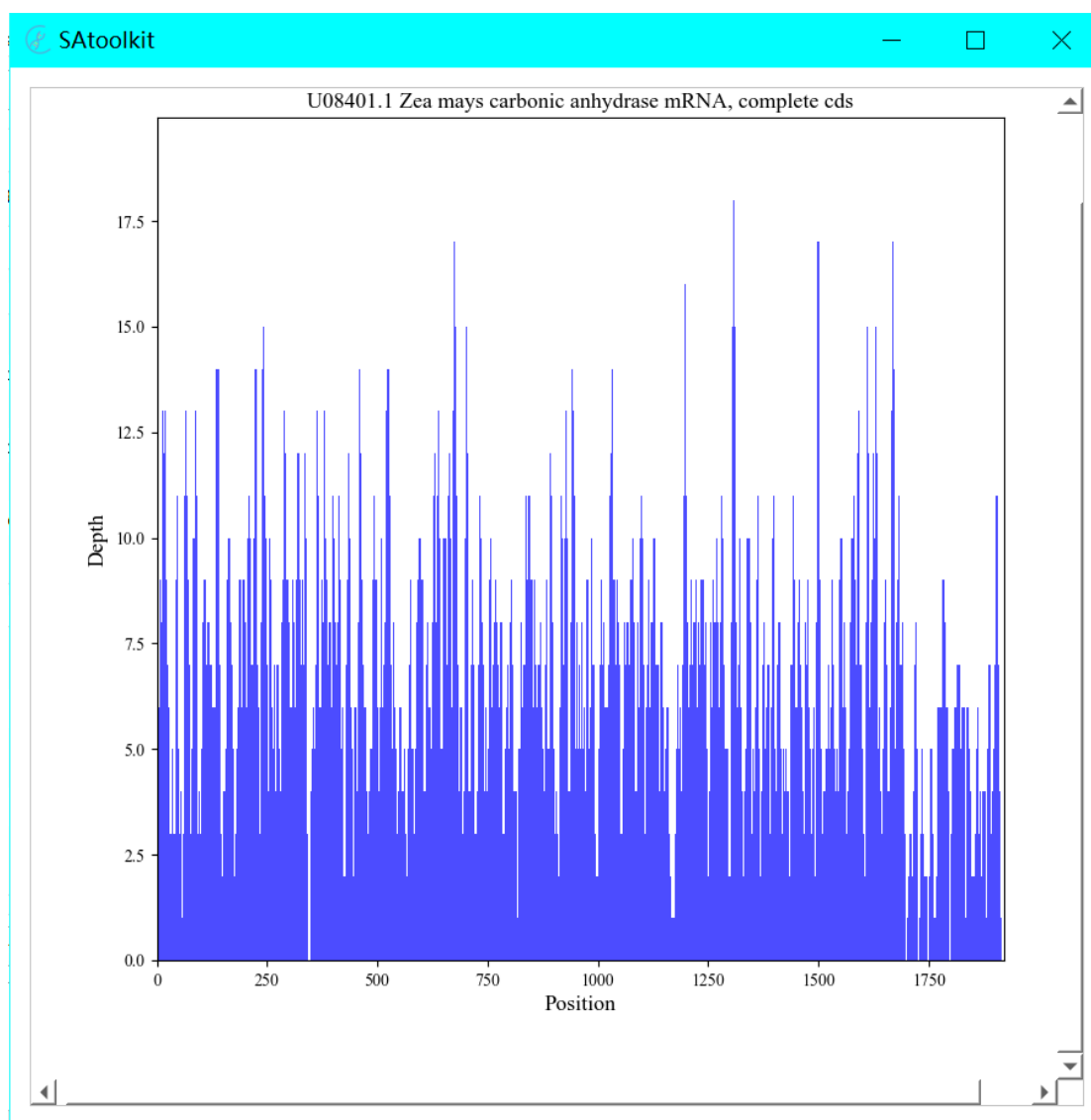
默认的结果是基因列表（Gene list）中第一条序列的自己对自的比对结果。控制面板一共划分为 4 个部分：Sequence analysis results、Graphic interaction、Parameter setting、Gene

list。

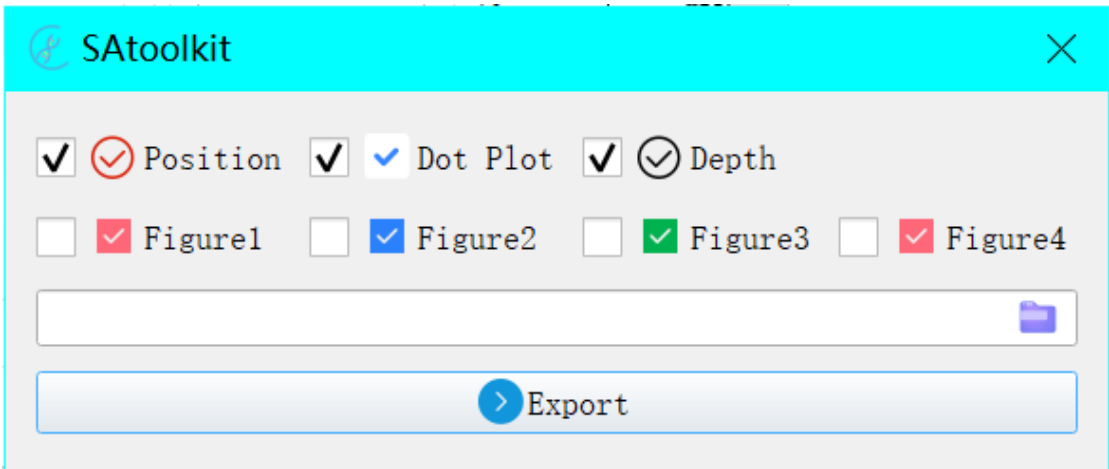
Sequece analysis results 部分一共三个按钮，Statistical graph 按钮点击之后将会弹出一个显示该序列整体情况的四种统计图。



Depth graph (如果可用的情况下) 按钮点击之后将会弹出一个显示序列整体内部点图深度的一条形图。



Export 按钮点击之后将会弹出一个布有用户可选择想要输出对应结果的勾选控件的对话框，用户选择之后，即可选择对应的输出文件夹路径并点击 Export 按钮输出即可。



SAtoolkit

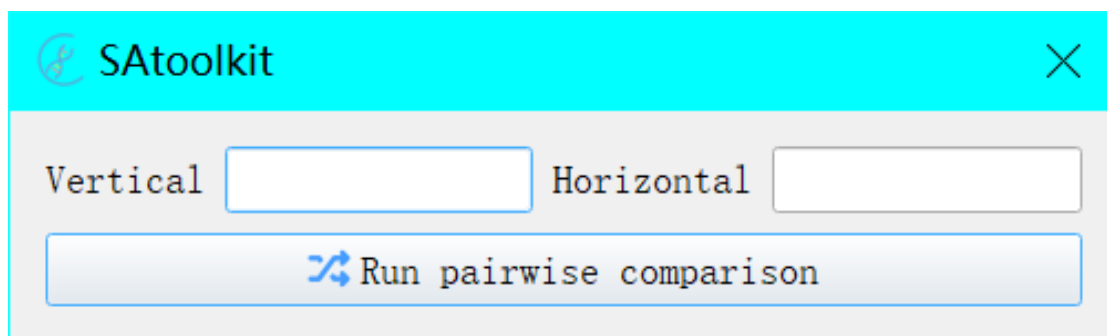
☒ ☒ Position ☒ ☒ Dot Plot ☒ ☒ Depth

☐ ☒ Figure1 ☐ ☒ Figure2 ☐ ☒ Figure3 ☐ ☒ Figure4

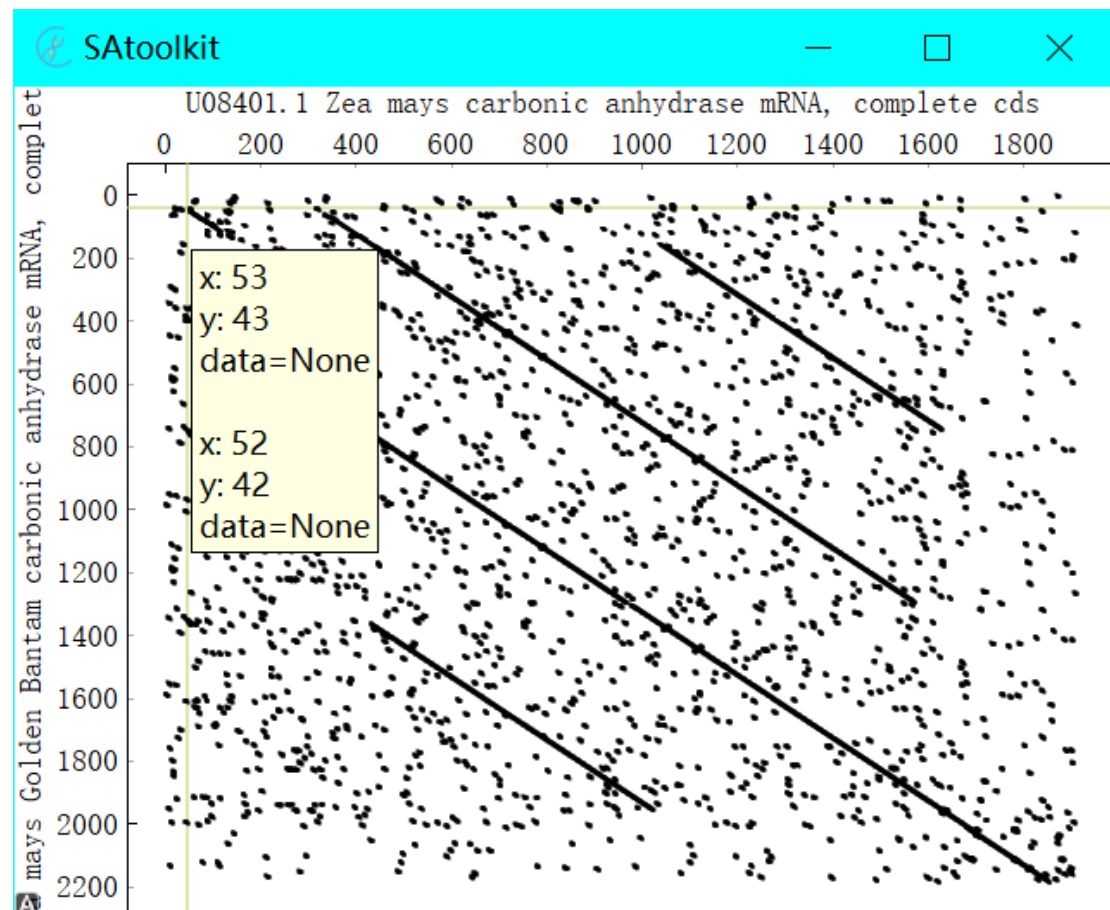
Graphic interaction 部分用于显示右边点阵图交互的十字光标位置坐标实时信息。

Parameter setting 是影响输出结果的一个重要部分，区域中 Mode 部分的三个控件从上到下分别表示 Direct repeats（寻找正向相似的区域）、Inverted repeats（寻找反向相似的区域）和 Reverse complement（寻找反向互补的相似区域）——当用户输入序列类型为 DNA 的时候，才会出现 Reverse complement 的控件。Setup 部分是表示用户设置的阈值，表示输出结果为 Identity/Similarity 大于等于当前阈值和 Repeat length 大于等于该长度的结果。

Gene list 部分用于显示用户输入序列的 ID 列表，用户选中那个 ID，在 Parameter setting 部分设置完参数之后即可点击 Run self to self 按钮，之后输出的结果将是这个 ID 对应序列的分析结果。该部分的 Pairwise 按钮的作用分别表示两两比对模式。用户点击 Pairwise 按钮之后，表示进入两两比对模式，用户通过选择 Gene list 里的基因名字，将用户选择的基因 ID 显示在 Vertical 和 Horizontal 知识的文本框中（用户最后一次点击的 ID 放置在 Horizontal（水平），倒数第二次点击的 ID 放置在 Vertical（垂直））。



确认无误后，点击 Run pairwise comparision 运行即可出相应的结果。



另外，SAtoolkit 还有很多简单方便的小功能，用户通过字符即可快速上手。总而言之，SAtoolkit 对于生物序列的比对，是一个十分便捷的工具。