

## PROJECT INTRODUCTION

### Objectives

To study the bacterial micro-organisms found in the air in Singapore.

### Project Manager (Grid)

Dr. Patrick Tan  
Genome Institute of Singapore  
tanbop@gis.a-star.edu.sg

### Abstract

Scientists at the Genome Institute of Singapore (GIS) recently completed a pilot grid research programme to study the bacterial microorganisms found in the air in Singapore. The analysis of this data was found to be highly compute intensive, and Grid Computing was tapped to provide the much needed speed up, in collaboration with researchers at the National University of Singapore.

## PROJECT DETAILS

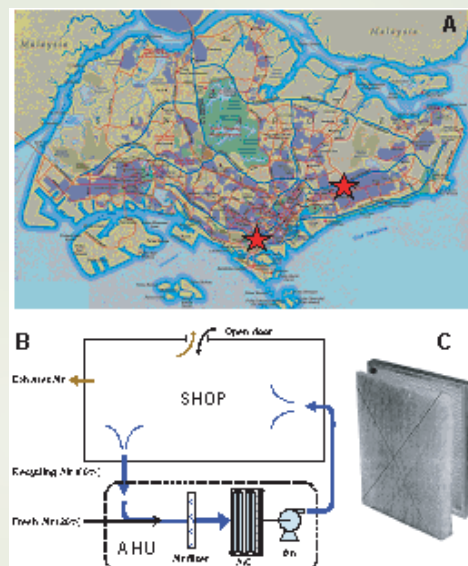
### Description

Air quality plays a significant role in public and environmental health but very little is currently known about micro-organisms in the air of urbanized cities. The research team at GIS used the latest DNA sequencing technology, and overcame the major challenge of analyzing and identifying the thousands of bacterial species through the computing resources made available on the Grid.

The sequencing process resulted in huge amounts of data. Initially, a pilot data set of 20,000 DNA sequences was matched against a large public database of some 3 million known sequences (Genbank). It was estimated that a single run, using isolated computing resources, would have taken an entire month. There is a need to shorten this process significantly and it soon became clear that a totally new type computing infrastructure was required to handle data on this scale.

The analysis was repeated on the Tera-scale Campus Grid at NUS (TCG@NUS), tapping into pre-existing compute resources across the National University of Singapore. The entire query was remarkably shortened to less than two days.

With the success of this pilot grid-enabled trial, the GIS and NUS teams plan to scale up its study on micro-organisms in the air by analysing much larger data sets.



**Collaborating Organisation:**  
National University of Singapore