# (1) Submission ID#1539274

Antibody cross-reactivity is correlated with phylogenetic distance, supporting efficient vaccine antigen selection with the bioinformatics software Navargator

---


Author(s)

David Curran, PhD

Research Associate

University of Toronto

Dixon Ng, PhD

Postdoctoral Associate

University of Toronto

John Parkinson, PhD

Senior scientist

The Hospital for Sick Children

Scott Gray-Owen, PhD

Professor

University of Toronto

Trevor F Moraes, n/a

Professor

Biochemistry, University of Toronto

Jamie E. Fegan, PhD

Research Associate

University of Toronto



Background

The accessibility to the immune system of bacterial surface proteins makes them attractive targets for subunit vaccines. However, this same property also means they tend to exhibit high sequence variability. Achieving broad cross-protection usually necessitates that antigens from multiple isolates are included, but the choice of sequence variant is a non-trivial problem. Visual inspection of phylogenetic trees is the norm, but this is subjective and can be greatly influenced by the choice of viewing software. This has real-world implications, as groups have shown that the selection of non-optimal antigens likely led to lower cross-protection in the commercially available vaccines against Neisseria meningitidis serogroup B.

Aim/Methods

To address this problem, we have developed Navargator, bioinformatics software that takes a phylogenetic

tree as input and identifies the variants that are the most similar to the greatest number of other sequences. The underlying premise is that cross-reactivity will be correlated with phylogenetic distances extracted from the tree; this was validated by several rodent immunization studies with the proteins transferrin-binding protein B and factor H binding protein from N. meningitidis and N. gonorrhoeae, measuring antibody-based cross-reactivity between an antigen panel using a custom high-throughput ELISA.

Results
Navargator has been made freely available both as an online tool and as source code for local installation. We implemented several different clustering methods, with exact algorithms for smaller datasets, and heuristics suitable for large trees of thousands of sequences. Our immunization studies have shown that this approach is sound, and that cross-reactivity is predicted well by phylogenetic distances in a sigmoidal manner.

Conclusions
The complexity of vaccine development rises sharply with each additional antigen included, so using the minimal number required is an important consideration. Navargator attempts to facilitate this in a systematic and generalizable manner. The user can run the analysis by selecting their desired number of representatives, or they can provide any form of cross-reactivity data and have the program identify a minimum reactivity threshold via correlation with the phylogenetic tree. The program will then identify the smallest number of representatives required to satisfy this threshold.