Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information

Presented by: Thomas Cowell
November 29, 2018

Janssen, S. et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems* 3, e00021-18 (2018).
Outline

• Background

• SEPP method

• Results

• Conclusions
Studying the Microbiome

• Gut Microbes are known to influence health
• Short amplicons are obtained from the bacteria present in patient fecal samples
• The population of microbes present in the patient are inferred and associated with disease states
Short Amplicons Contain Weak Phylogenetic Signal

16S

V4

read recruitment

pro: reference phylogeny
con: losing sOTUs

de novo

con: no reference
pro: keep all sOTUs
The Phylogenetic Placement Problem

Input:
- A reference tree and alignment on the set of full-length sequences
- A “query sequence”

Output:
- The original tree with the query sequence added as a leaf

Current Methods:
- Step 1. Merge the query sequence into the full alignment
- Step 2. Add the query sequence, optimizing some tree criterion

Mirarab S, Nguyen N, Warnow T. 2012. SEPP: SATé-enabled phylogenetic placement. Pac Symp Biocomput 247–258
SATé-Enabled Phylogenetic Placement (SEPP)

SAT-é decomposes the full reference tree into small closely related subsets

1. HMMs extend the subset alignment to include the query sequence
2. Pplacer adds the query sequence to the subtree optimizing likelihood

pro: reference phylogeny
pro: keep most sOTUs
SEPP compared to De Novo Phylogeny

• Data Set: Amplicons of the V4 region of the 16S ribosomal subunit from 599 men studied for osteoporosis
• De Novo: MSA obtained via MAFFT, Phylogeny reconstruction using FastTree
• SEPP: HMMER + pplacer
• Individuals were clustered by the phylogenetic relatedness of their gut microbiome
SEPP compared to De Novo Phylogeny

mafft & fasttree
mafft & fasttree shorter Methanobrevibacter
SEPP
SEPP Provides Increased Resolution

• Amplicon sequences were obtained from fecal samples of 179 children in Malawi
• Growth was measured simultaneously (height by age) and grouped into good and poor
• SEPP was compared against closed reference and open reference OTU picking methods.
SEPP Provides Increased Resolution

- SEPP distinguishes groups with the highest significance
- subOTUs provide higher taxonomic resolution
SEPP Improves Phylogenies

- 10,000 fragments were selected from a large reference tree
- A de novo phylogeny was reconstructed on the fragments
- SEPP was used to reinsert the fragments into the reduced reference tree
**Accuracy of Reinsertion using SEPP**

- Short fragments were created from each of the full sequences in the reference database.
- SEPP successfully reinserted fragments with 5 or fewer ambiguities with species level resolution.

---

**Diagram Description:**

- **Part A:** A bar chart showing the mean tip distance for different categories like Kingdom, Phylum, Class, Order, Family, Genus, and Species.
- **Part B:** A box plot depicting insertion distance for different fragment ambiguities (1 to >100) with varying levels of confidence intervals.
- **Part C:** A bar graph illustrating the number of fragments (log scale) for each fragment ambiguity category.
Accuracy of Reinsertion using SEPP

- Unambiguous fragments were randomly mutated 1 to 10 times
- SEPP reinsertion was resolved below the species level for up to 3 mutations
- A typical 200 bp read can be expected to contain 2 errors which can still be resolved by this method
SEPP is Highly Parallelizable

- The Phylogenetic Placement Problem is separate for each query sequence
- SEPP implementation is largely parallelizable
Conclusions

• SEPP uses divide and conquer approach to improve taxonomic assignment
• subOTU methods improves taxonomic resolution and accuracy by accommodating the full amplicon sequences
• De novo approaches can provide misleading results
• Clinical variables are recovered with greater statistical power using SEPP methods
Questions?