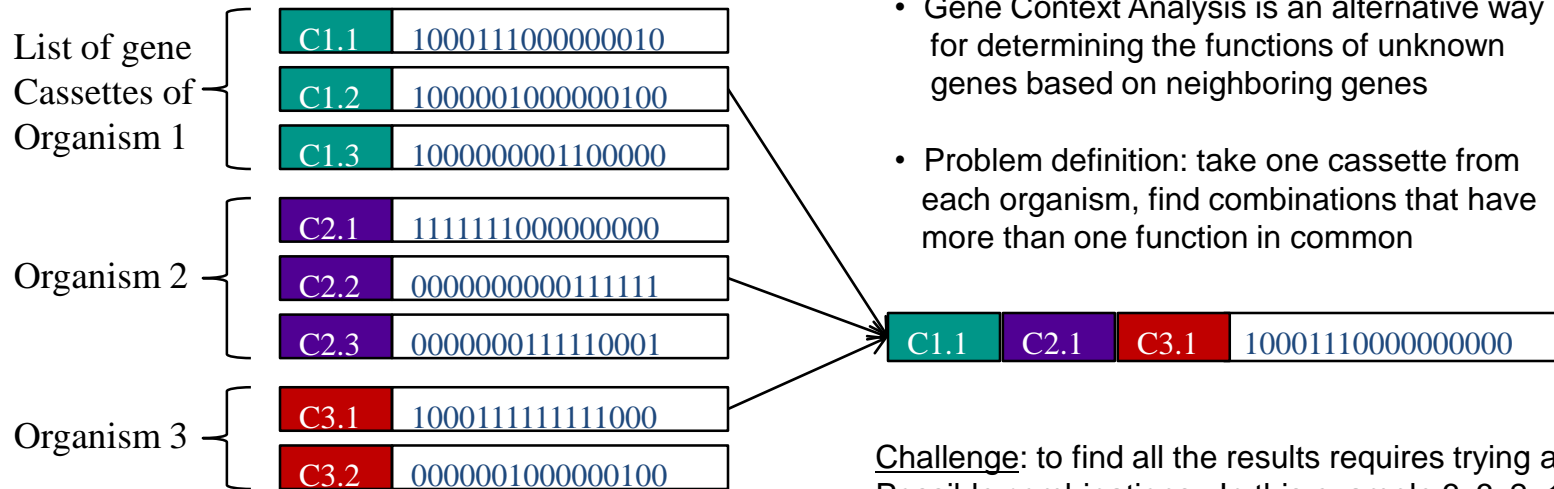


Gene Context Analysis Now Performed in Seconds



The figure shows the bitmap representation of cassette properties, and finding matching combinations

- Gene Context Analysis is an alternative way for determining the functions of unknown genes based on neighboring genes
- Problem definition: take one cassette from each organism, find combinations that have more than one function in common

Challenge: to find all the results requires trying all Possible combinations. In this example $3 \times 3 \times 2 = 18$. For millions of cassettes this search is **exponential**.

APPROACH

- Reorganize the list of functions per gene cassette into bitmaps
- Use FastBit to compress the bitmaps
- Re-structure the query processing algorithm into bitwise logical operations
- Remove solution entries that are contained in other entries (maximal solutions only)
- Progressive pruning the possible solutions based using bitmaps as keys for comparisons

IMPACT

- Providing interactive exploration
- When more than 5 organisms are involved in a query, the previous system based on a commercial database system takes too long and the GUI times out
- Using the new solution, queries involving 600 organisms took **less than 10 seconds**
- New solution deployed in IMP system at <img.jgi.doe.gov> since May 2013