PClouds vocabulary update

Stephen Pollard

12/17/2013

The vocabulary used in the code is different from the vocabulary we use now to talk about clouds. This document is my understanding of the old terms and tries to connect them to the terms we use now.

# Files

I renamed all the “.c” files to “.cpp” because they all are C++ source files and not simply C source files. Things should be named what they are.

For some reason, the main() function was in “genomepro.cpp”. I have no idea why but I have renamed “genomepro.cpp” to “PClouds.cpp”.

# PCloud

When the code says ‘Pcloud’ it actually means ‘kmer’.

These struct are define in “macrodefine.h”. If you look at the members of these structs, you’ll find different combinations of ‘index’, ‘number’, ‘cloud’, and ‘extension’.

The code uses ‘extend’ to describe expanding a core kmer. Now we usually use the word ‘expand’.

|  |  |  |
| --- | --- | --- |
| Name in code | Better/newer name | Examples/description |
| Pcloud | Kmer | \_Pcloud, \_Pcloud1, \_Pcloud2, \_Pcloud3 |
| index | sequence\_as\_int | \_Pcloud.index |
| number | count | \_Pcloud.number |
| cloud | assigned\_cloud\_id | \_Pcloud1.cloud |
| extension | has\_expanded | keep track of whether this core kmer has been expanded around or not |
| pattern | sequence | patterntoindex() |
|  |  |  |
| main clouds | core kmers | kmers above the core threshold |
| accessory pclouds | outer kmers | kmers above the outer threshold and below the core threshold |
| main clouds A, B, C | Doesn’t need a name | core kmers above step 1, 2, 3 thresholds |
| number A, B, C | Doesn’t need a name | number of core kmers above step 1, 2, 3 thresholds |
|  |  |  |
| copy threshold | outer threshold | first value in “C parameters” |
| end threshold | core threshold | second value in “C parameters” |
| step 1,2,3 threshold | ??? | third, fourth, and fifth values in “C parameters” |
|  |  |  |
| repeat file | kmer\_counts\_file |  |
|  |  |  |
| prefix “pch” on a variable name | unnecessary | pointer to character (often to a character array that should be a string) |
| prefix “m\_n” on a variable name | unnecessary | integer |
|  |  |  |