The Remap Service from NCBI

Mapping genomic coordinates from one genome assembly to another for selective organisms

https://www.ncbi.nlm.nih.gov/genome/tools/remap/

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope

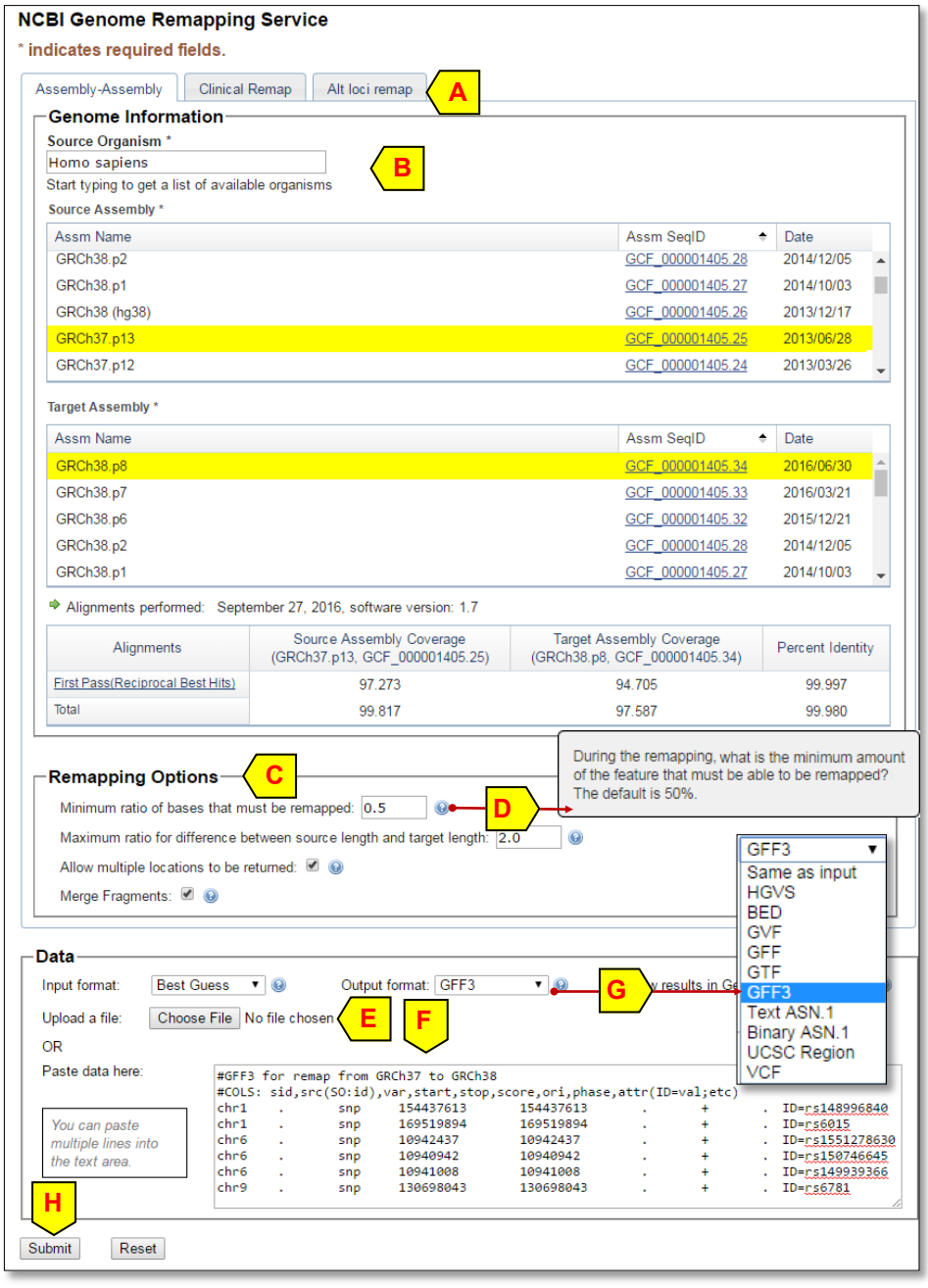
Updates of genome assemblies or the availability of multiple assemblies for a single organism present an annotation comparison challenge. Performing a *de novo* annotation, especially for some datasets, can be computationally and time intensive. One way to quickly address this issue is to align the assemblies to each other and then 'map' the features from one coordinate system to the other using these alignments as a guide. While not a replacement for *de novo* annotation, this 'remap' process can be useful when looking at individual regions. When performed on annotations across the genome, this process can provide useful information on the degree of difference between the two assemblies. The Remap Service is a tool from NCBI that makes remapping features/annotations simple and straightforward.

Access to the service

The Remap Service is accessible online through the genome tools collection at https://www.ncbi.nlm.nih.gov/genome/tools/remap. The Assembly-Assembly Remap Service allows for remapping of features from one assembly to another. Clinical Remap allows for the remapping of features (including HGVS-defined sequence variations) to RefSeqGene sequences, including the underlying transcript and protein sequences. It also maps from the RefSeqGene sequences to an assembly.

The input form

The Remap Service takes input from a web form (as shown). The type of remap to be performed can be selected using tabs at the top (**A**). For Assembly-Assembly remap, organism and source/target genomes can be selected using the lists in the "Genome Information" section (**B**). Mapping criteria can be adjusted using parameters in the "Remapping Options" (**C**). Increasing the "minimum ratio of bases" and decreasing the "maximum ratio of difference" improve the stringency of the mapping process. Popups linked from question mark icons (**D**) provide context-specific help.



The actual list of coordinates to be remapped can be uploaded as a text file using the "Choose File" button (**E**) or pasted directly in the textbox (**F**). In the example, a set of features from GRCh37 in GFF3 format is pasted in the input box, see page 2 for details. Multiple formats for input/output coordinates are supported with specific formats selectable using the pull-down list (**G**). A message box at the bottom (not shown) identifies fields needing input data. It disappears when all requirements have been satisfied. The "Submit" button (**H**) will then become available.

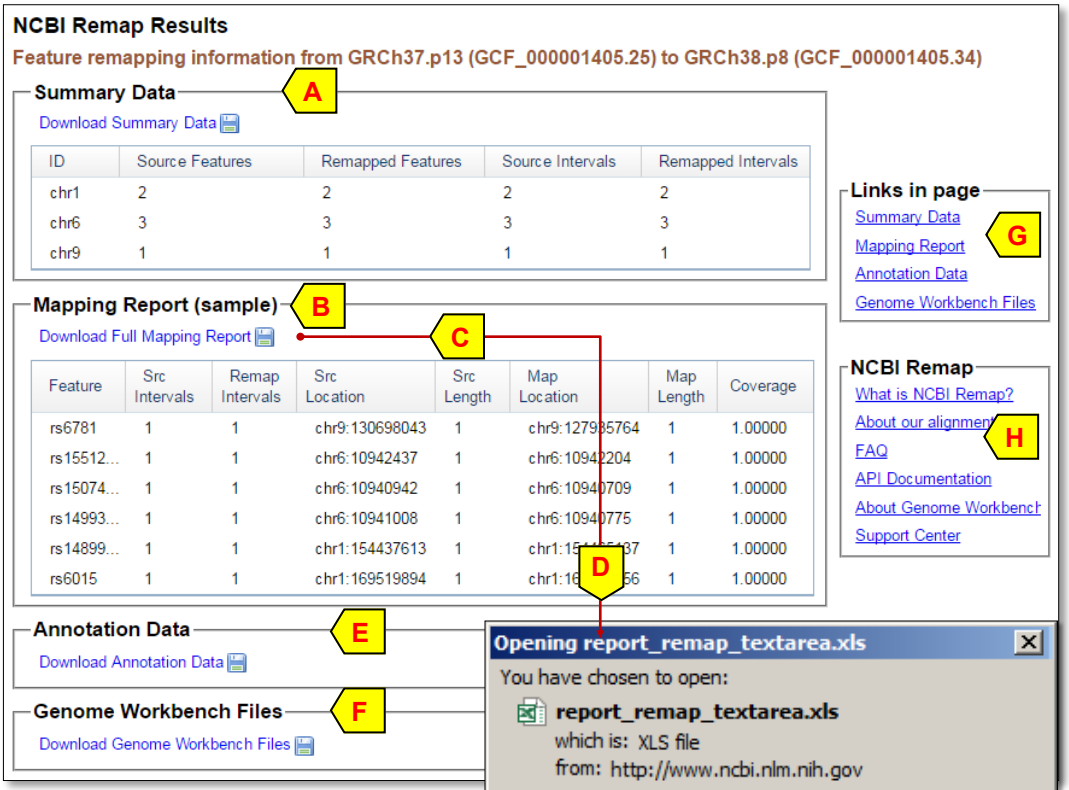
Example input format

The Remap Service supports several input formats. The example shown below is a list of SNPs from GRCh37p13 (hg19) in GFF3 format, which can be used as test input to the Remap Service to map them forward to GRCh38p2. For testing purposes, source and target assemblies can be switched.

```
#GFF3 for remap from GRCh37 to GRCh38
#COLS:sid,src(SO:id),var,start,stop,score,ori,phase,attr(ID=;etc)
chr1    .       snp     154437613       154437613       .       +       .       ID=rs148996840
chr1    .       snp     169519894       169519894       .       +       .       ID=rs6015
chr6    .       snp     10942437        10942437        .       +       .       ID=rs1551278630
chr6    .       snp     10940942        10940942        .       +       .       ID=rs150746645
chr6    .       snp     10941008        10941008        .       +       .       ID=rs149939366
chr9    .       snp     130698043       130698043       .       +       .       ID=rs6781
```

Displaying and downloading remap results

A completed remap request is displayed in a summary format with four main sections: Summary Data, Mapping Report, Annotation Data, and Genome Workbench Files.



The Summary section (**A**) provides a synopsis of the remapped features grouped by chromosome. Complete summary is downloadable using the disk icon. Up to ten remapped entries are shown in the Mapping Report section (**B**), with the complete set in the downloadable file marked by the disk icon (**C**). Clicking the icon activates a download window (**D**) from where the result can be saved or opened in Excel (under Firefox, other browsers may behave differently).

All features are in this report whether they successfully remapped or not. Additionally, this report provides interval-by-interval comparison of the features on the source and the target assemblies. Lastly, the coverage score is calculated by dividing the length of the feature in the target assembly by the length of the feature in the source assembly. A value of "1.00000" is perfect, while a value of less than 1 indicates a deletion in the target assembly and a value of greater than 1 indicates an insertion in the target assembly. This sort of metric does not assess single base changes within the region.

The link in the Annotation Data section (**E**) provides the remap data for the target assembly only. The link in the Genome Workbench Files section (**F**) provides the data in a format that can be readily imported into the standalone Genome Workbench tool (https://www.ncbi.nlm.nih.gov/projects/gbench/) from NCBI. Online links to help documents are also available from the result page (**G**) to provide more detailed explanation on the file format. Additional requests for technical assistance and feature additions to this tool can be emailed to NCBI using the "Write to the Help Desk" link (**H**).

API access

The Remap Service also provides API access. An example Perl script along with example command line are also provided. Detailed explanation is available in the online document:
https://www.ncbi.nlm.nih.gov/genome/tools/remap/docs/api?page=result