## Chrome

Purpose: Search by Chromosome Accession number, start and end position

Returns sequence

MongoDB Records

SeqPage {

Chromosome accession

FASTA line (gi|568336023|gb|CM000663.2| Homo sapiens chromosome 1, GRCh38 reference primary assembly)

Start

End

Sequence (zipped binary)

}

Procedure -> Building

Scan directory for all fa files. For each file, extract the accession number

Proceed to read sequence data, in chunks of 10,000 nucleotides. For each 10,000 nucleotides, gzip the sequence and create a SeqPage record, with the correct start/end values.

Procedure -> Web Service

Create a node service (REST) like this:

qdb.ramapo.edu/chrome/accession/start/end that returns the sequence

Look up the correct pages (might be a set), assemble sequence.

# mRNA and Gene

Create mRNA records from the top level file, insert feature data from the feature file.

Do similar for genes – however the features are more limited.

Create a node service (REST) like this:

qdb.ramapo.edu/mrna/accession

qdb.ramapo.edu/gene/geneid

On the UI, embed links to get sequence data (by constructing URL) and to get downstream data (on mRNA).

Note – it would also be nice to make the repository explorable through rest: So the following URLs would return list of all genese or mRNA associated with a species.

Qdb.ramapo.edu/gene/species

Qdb.ramap.edu/mrna/species

# Homologene

Need to develop a listing of all homologs

URI should be

qdb.ramapo.edu/homolog/mrna/accession

qdb.ramapo.edu/homolog/gene/geneid

These would return list of homologous mRNA or genes.

# Alignment

Need service for allignment

# QGRS and QGRS-H for Genes -> use current java implementation for seeding, but use node to serve the site. Utilize services above to perform the seeding (source data).