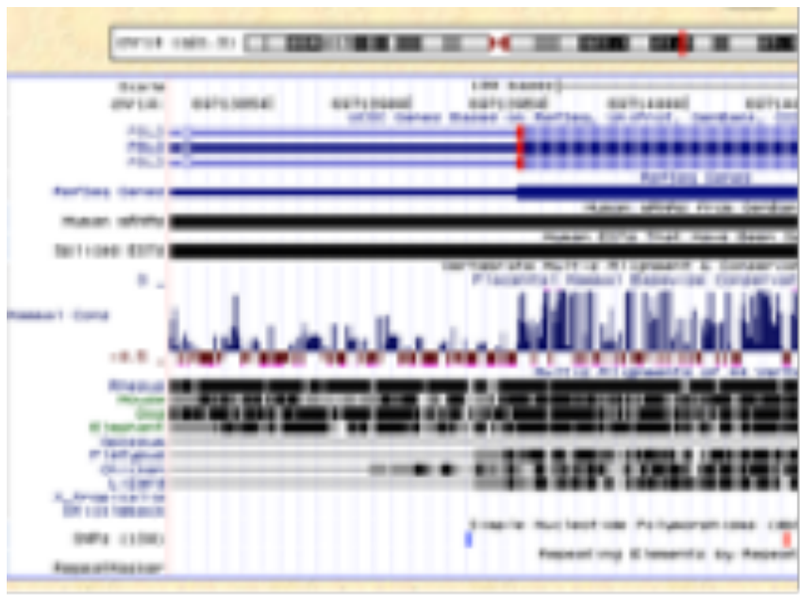
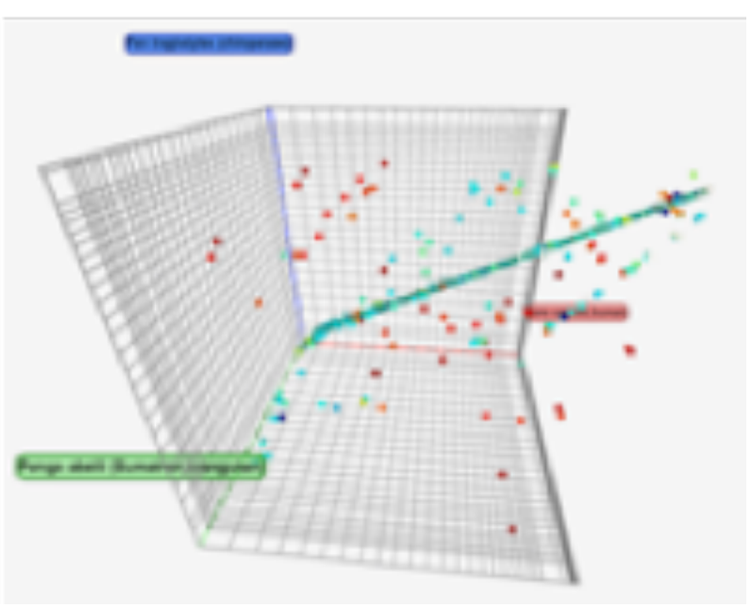
Computational Genomics

Introduction To Genome Annotation

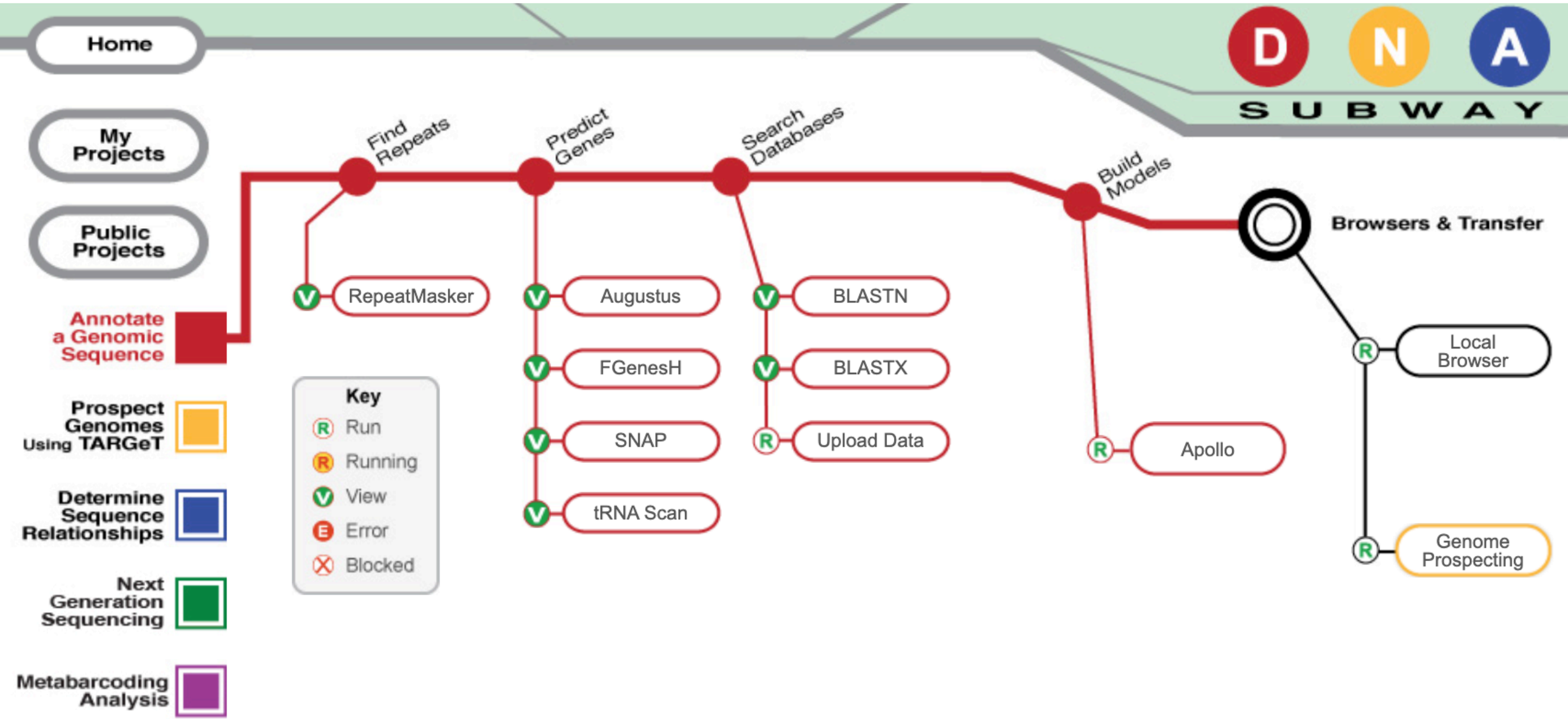
Cyverse User Portal



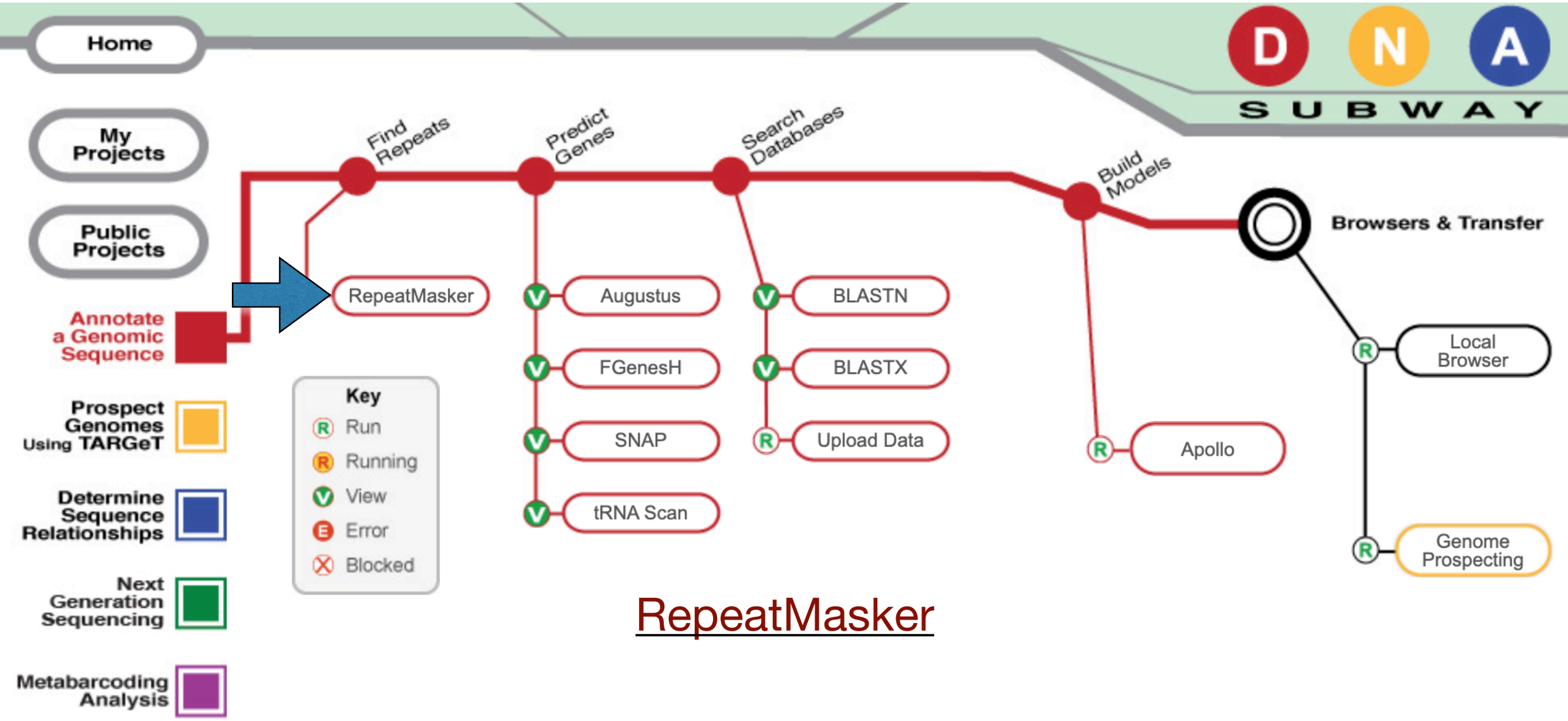


Cyverse User Portal

DNA SUBWAY


Home
My Projects
Public Projects
Annotate a Genomic Sequence
Prospect Genomes Using TARGeT
Determine Sequence Relationships
Next Generation Sequencing
Metabarcoding Analysis
Find Repeats
Predict Genes
Search Databases
Build Models
Browsers & Transfer
RepeatMasker
Augustus
FGenesH
SNAP
tRNA Scan
BLASTN
BLASTX
Upload Data
Apollo
Local Browser
Genome Prospecting
Key
Run
Running
View
Error
Blocked
D N A
SUBWAY

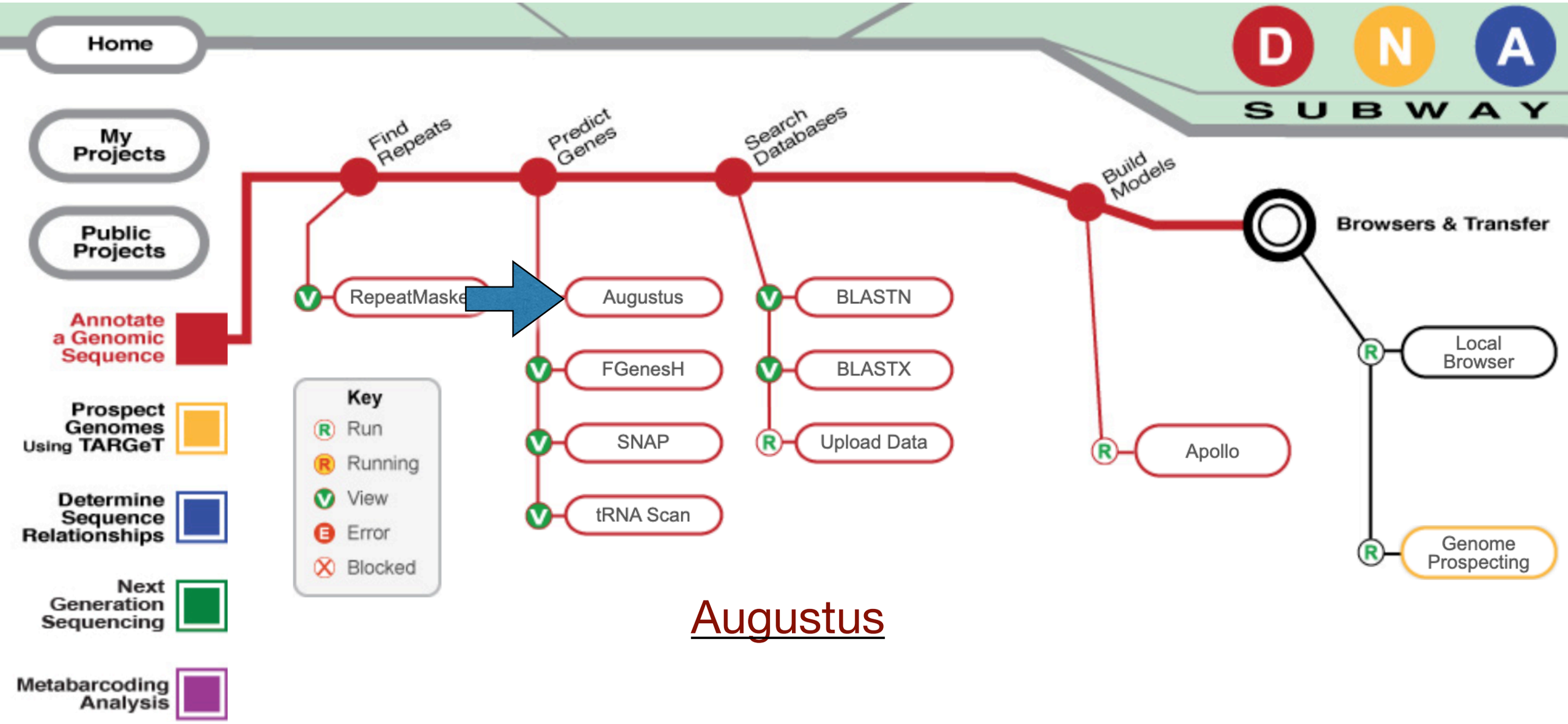
DNA SUBWAY



RepeatMasker

RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences. The output of the program is a detailed annotation of the repeats that are present in the query sequence as well as a modified version of the query sequence in which all the annotated repeats have been masked (default: replaced by Ns). Currently over 56% of human genomic sequence is identified and masked by the program. Sequence comparisons in RepeatMasker are performed by one of several popular search engines including nhmmer, cross_match, ABBlast/WUBlast, RMBlast and Decypher. RepeatMasker makes use of curated libraries of repeats and currently supports Dfam (profile HMM library derived from Repbase sequences) and Repbase, a service of the Genetic Information Research Institute.

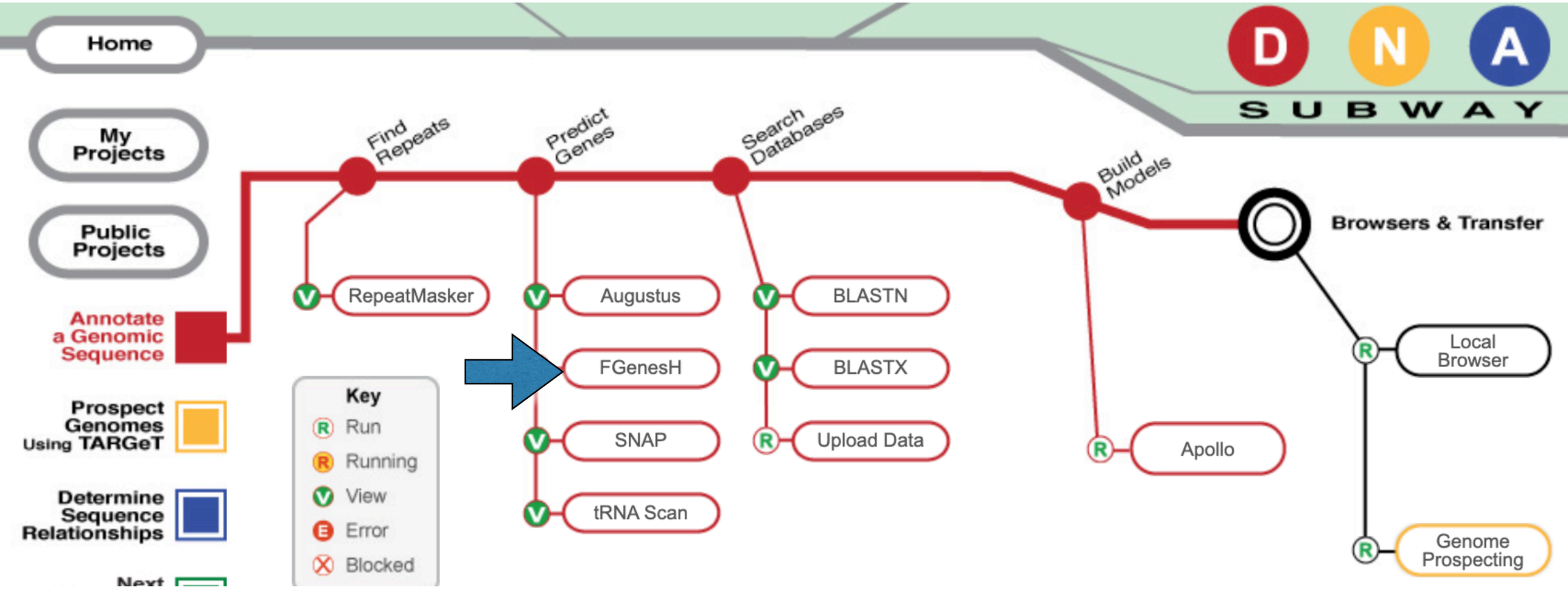
DNA SUBWAY



Augustus

AUGUSTUS is a program that predicts genes in eukaryotic genomic sequences. It can be run on this web server, on a new web server for larger input files or be downloaded and run locally. It is open source so you can compile it for your computing platform. You can now run AUGUSTUS on the German MediGRID. This enables you to submit larger sequence files and allows to use protein homology information in the prediction. The MediGRID requires an instant easy registration by email for first-time users.

DNA SUBWAY



FGenesH

- Fgenesh - most accurate and fastest HMM-based gene prediction program;
- Fgenesh+ - gene prediction program that uses similar protein information;
- gene finding parameters for Fgenesh and Fgenesh+;
- est_map - program for mapping known mRNAs to genome (genome alignment with splice sites identification) and mapping a set of available ESTs to improve the gene prediction accuracy and add 5' and 3'- noncoding sequences
- prot_map - mapping protein database to genomic sequences;

Fgenesh++ pipeline main steps

1. Predict gene models using full length mRNAs.
(Map full-length mRNAs - from RefSeq or other sources – to a genome using Est_Map program, select good mappings. Mapped regions are excluded from further gene mapping process.)

2. Predict gene models using known proteins from NR.

2a. Map known proteins (from NR database or its subsets) to a genome using Prot-Map program. Reconstruct gene structures.

2b. Refine gene models with mapped protein sequences using Fgenesh+ program.

2c. Select reliable gene models through blast2 alignments between predicted and homologous proteins.

3. Predict genes *ab initio* on the rest of genome by Fgenesh program. Use genefinding parameters trained on query genome or its close relative.

4. Predict genes *ab initio* in large introns.

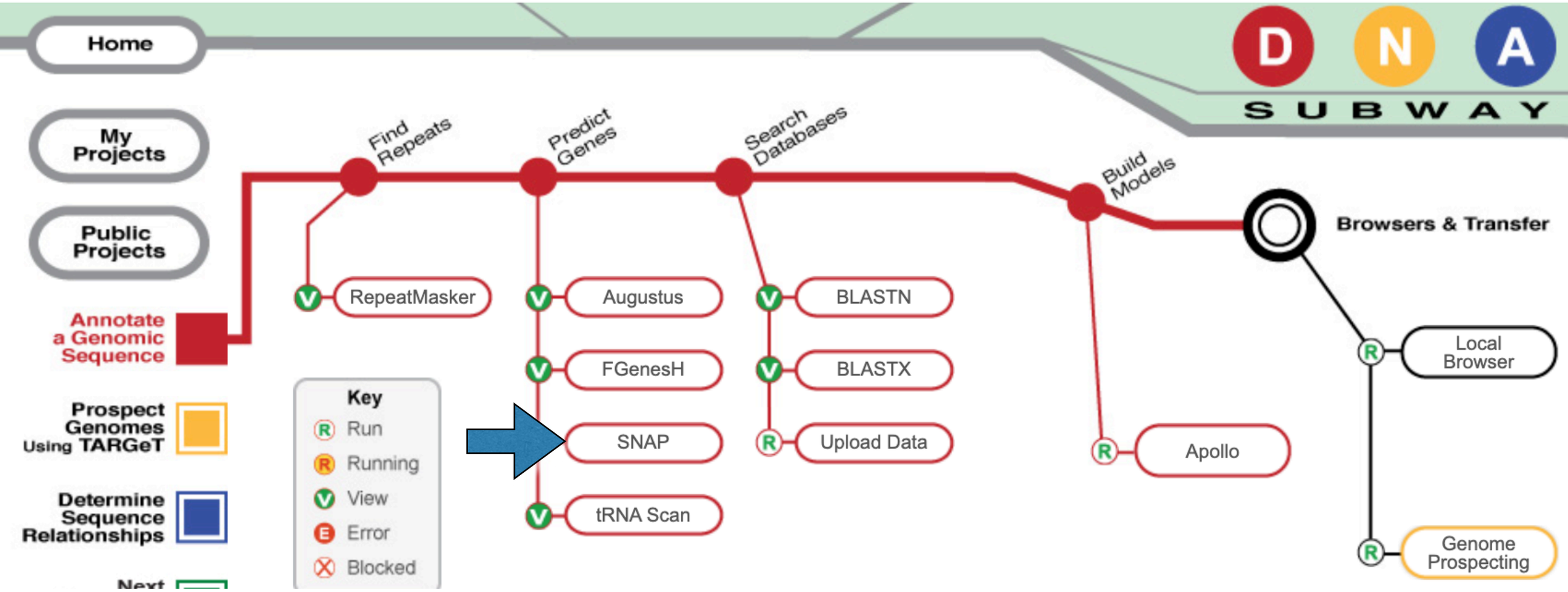
Using transcript reads
Align transcript reads to genomic sequence by Reads_Map program

Using ESTs*
Align ESTs to genomic sequence by Est_Map program

Lists of potential splice sites and introns are used as additional evidence

*With ESTs, 5'- and 3'- untranslated parts of first and last CDS exons can also be predicted.

DNA SUBWAY



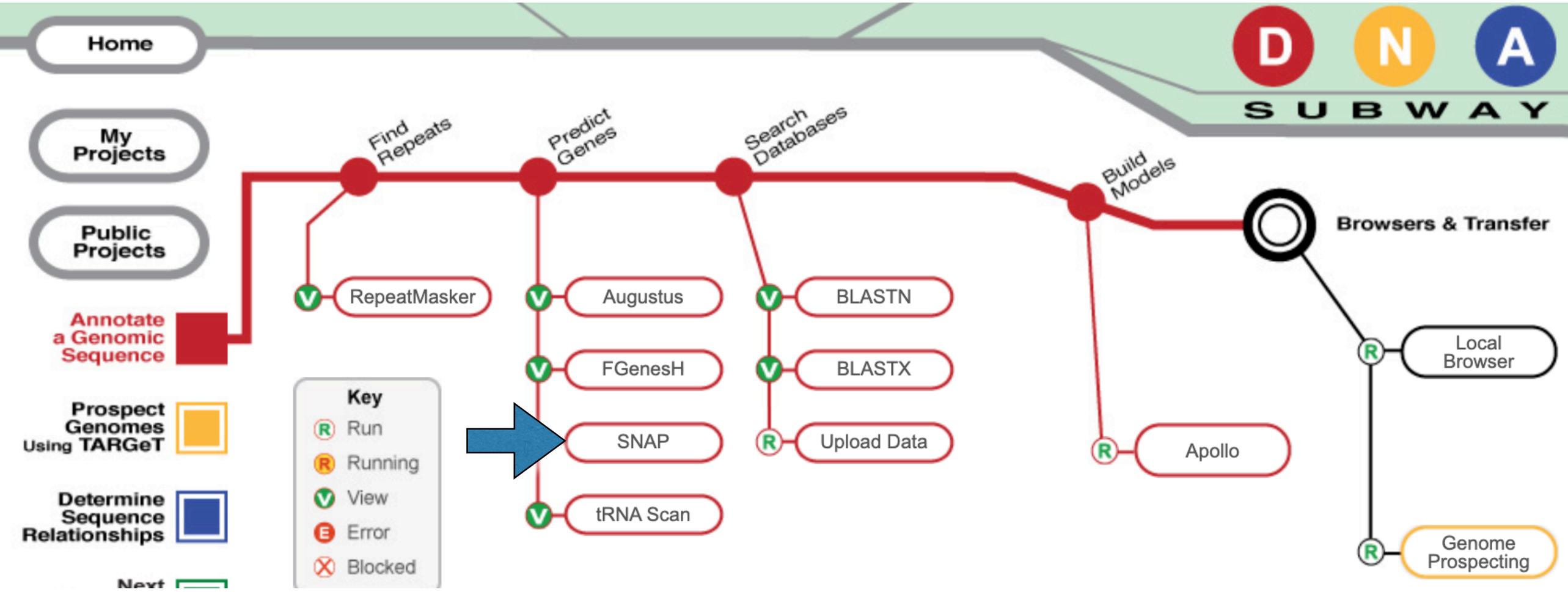
SNAP

The Korf Lab

ATAGCGAAT
TATCGCTTA

Making sense of sequences

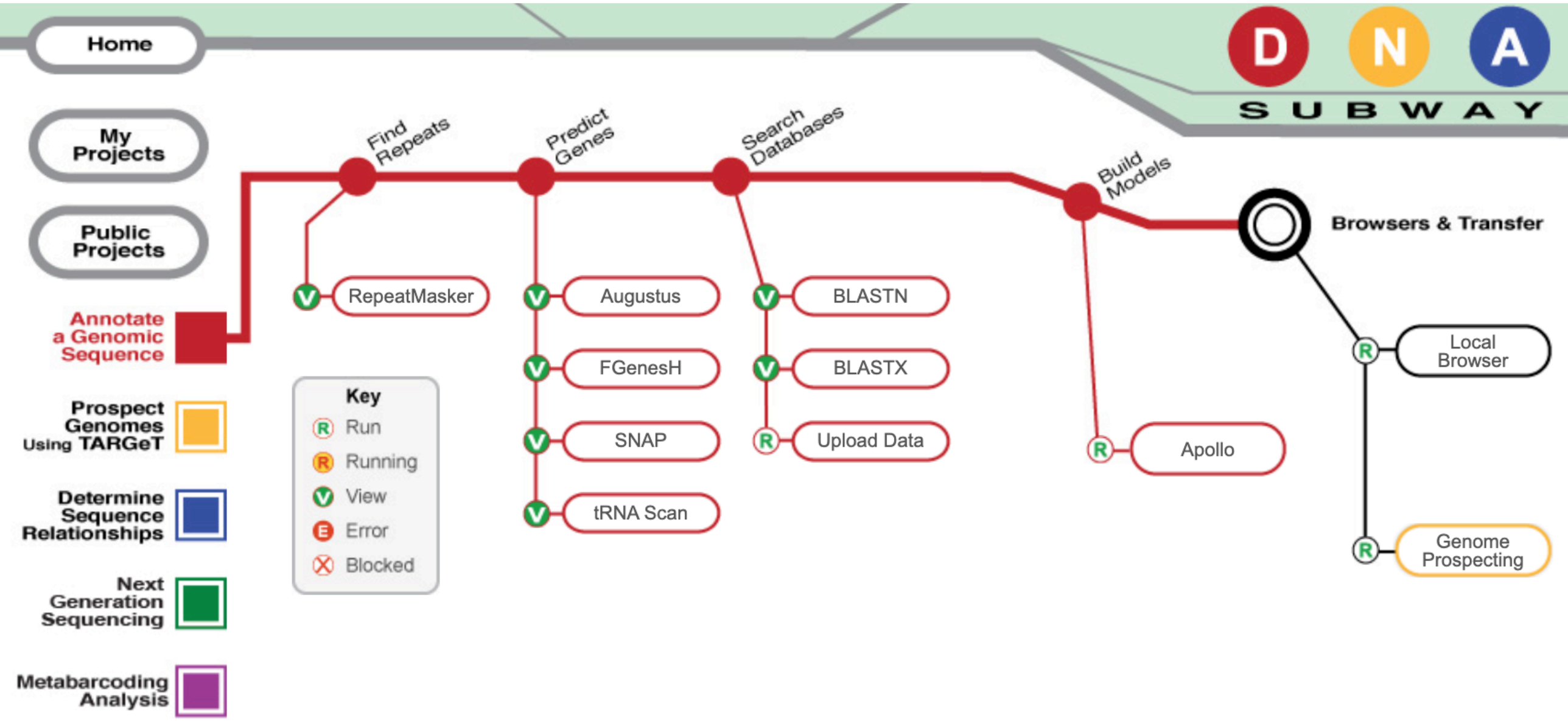
DNA SUBWAY



tRNAscan-SE

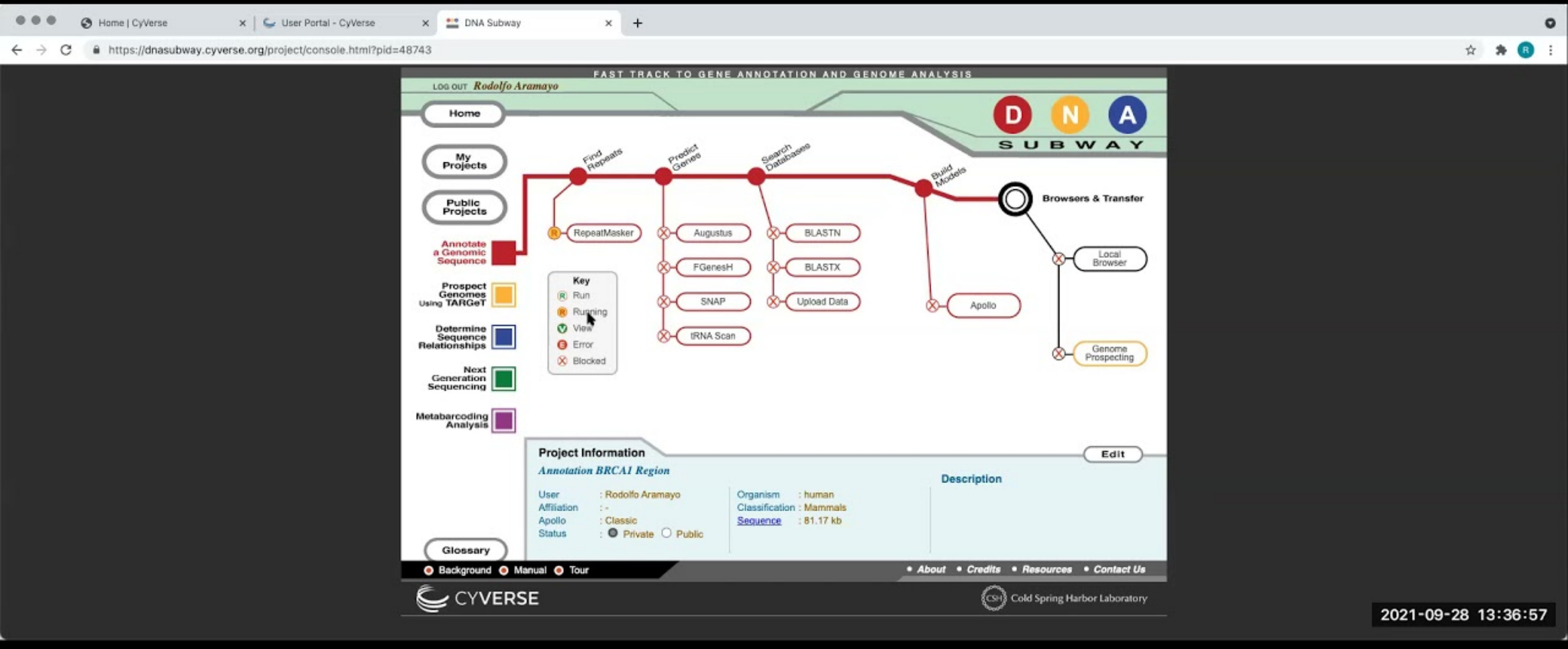
tRNAscan-SE detects ~99% of eukaryotic nuclear or prokaryotic tRNA genes, with a false positive rate of less than one per 15 gigabases, and with a search speed of about 30 kb/second. It was implemented for large-scale human genome sequence analysis, but is applicable to other DNAs as well. It applies our COVE software (see below) with a carefully built tRNA covariance model, while getting around COVE's speed limitations by using two tRNA finding programs from other research groups as fast first-pass scanners (Fichant and Burks', and an implementation of an algorithm from A. Pavesi's group). It runs on any UNIX system with Perl and a C compiler installed.

DNA SUBWAY


Home
My Projects
Public Projects
Annotate a Genomic Sequence
Prospect Genomes Using TARGeT
Determine Sequence Relationships
Next Generation Sequencing
Metabarcoding Analysis
Find Repeats
Predict Genes
Search Databases
Build Models
Browsers & Transfer
RepeatMasker
Augustus
FGenesH
SNAP
tRNA Scan
BLASTN
BLASTX
Upload Data
Apollo
Local Browser
Genome Prospecting
Key
Run
Running
View
Error
Blocked
D N A
SUBWAY

DNA SUBWAY

BRACA1 DNA Region



2021-09-28 13:36:57