

Git For Genomes

Version control through sequence graphs

*“The machine code of the genes is
uncannily computer-like”*

R. Dawkins

Yet genetic engineering does not feel like software engineering.

Genetic engineering: *auto-hopped beer*

Task: *insert genes responsible for hop aroma chemistry into the yeast genome.**

1. Design

- a. Source genetic parts from the iGEM parts registry.
- b. Create combinations of variable parts.

2. Build

- a. Plan stepwise cloning strategy
- b. Order or reuse synthetic material (\$)
- c. Assemble intermediate constructs & edit genome



* fictional account, but this has been done before: Denby, C.M., Li, R.A., Vu, V.T. *et al.* Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat Commun* 9, 965 (2018).

Genetic engineering: *auto-hopped beer*

Task: *insert genes responsible for hop aroma chemistry into the yeast genome.**

1. Design

- a. Source genetic parts from the iGEM parts registry.
- b. Create combinations of variable parts.



2. Build

- a. Plan stepwise cloning strategy
- b. Order or reuse synthetic material (\$)
- c. Assemble intermediate constructs & edit genome

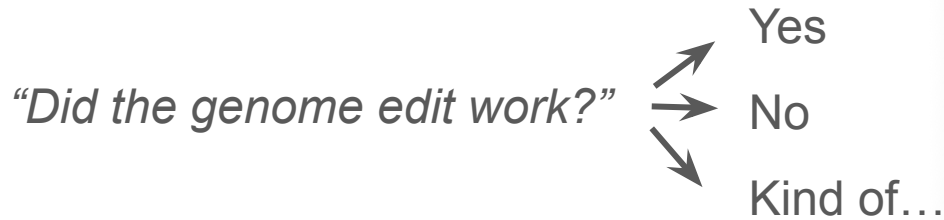
Efficient *Build* strategies
require *Design* coordination

* fictional account, but this has been done before: Denby, C.M., Li, R.A., Vu, V.T. *et al.* Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nat Commun* 9, 965 (2018).

Genetic engineering: *auto-hopped* beer

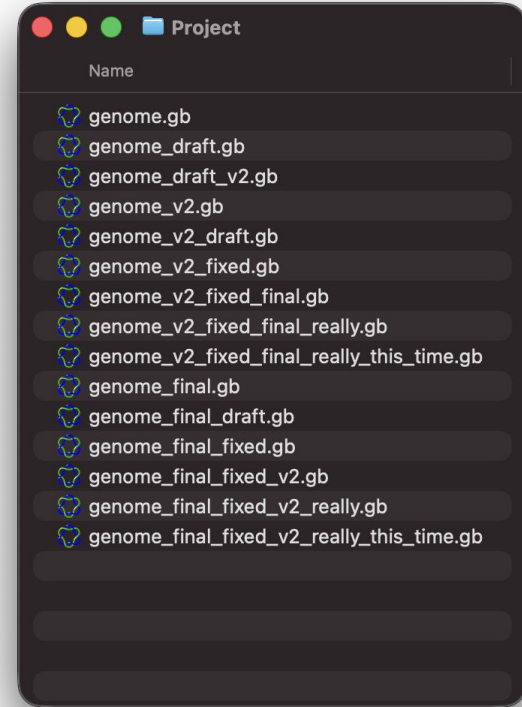
3. Test

a. Read back the genome (DNA sequencing)



b. Evaluate performance of experimental strains

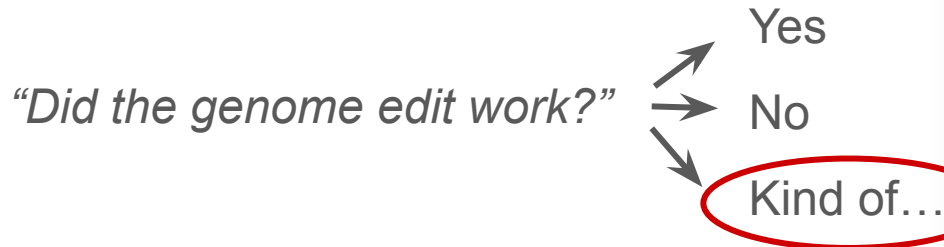
4. GOTO 1



Genetic engineering: *auto-hopped* beer

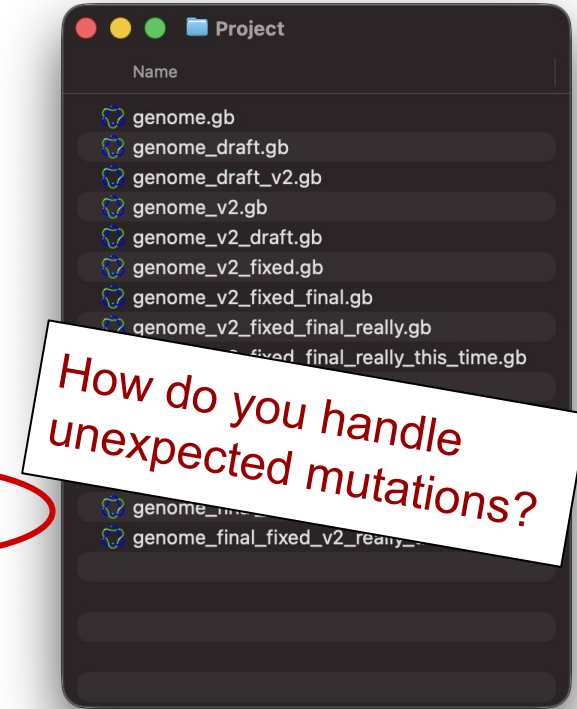
3. Test

- a. Read back the genome (DNA sequencing)



- b. Evaluate performance of experimental strains

4. GOTO 1



Closing the Design-Build-Test loop is much harder than it should be

Introducing the **gen** version control system for biology

- Rust crate with Command Line Interface, Terminal User Interface, Python bindings
- Organize sequences and samples in SQLite-backed repositories
- Changes tracked as operations (~commit)
- Familiar git commands
 - init, checkout, branch, reset, diff, push, pull, ...

The screenshot displays the **gen** version control system's terminal interface. It features a dark theme with a monospaced font. The interface is divided into several panels:

- Operations:** A table listing recent operations with their hashes, change types, and summaries.
- Operation Summary:** A detailed view of the selected operation, showing the sample name and the number of changes.
- Change Graph:** A visual representation of the sequence changes, showing a graph with nodes and edges.

At the bottom of the terminal, a status bar indicates the current state and provides keyboard shortcuts for saving and leaving the panel.

Operation Hash	Change Type	Summary
77ad79f3c03237524057	fasta_addition	m123: 34 changes.
0135c3782dbb15ef70c1	vcf_addition	Sample unknown m123: 4 changes.

Operation Summary
Sample unknown m123: 4 changes.
Sample foo m123: 2 changes.

Change Graph default foo m123

```
graph LR
    ATC[ATC] -- G --> ATCGA[ATCGA ... GAGA]
```

ctrl+s save | esc leave panel

Sequence specific commands (abridged)

Task: insert genes responsible for hop aroma chemistry into the yeast genome.

1. Design

- a. Source genetic parts from the iGEM parts registry. `gen import`
- b. Create combinations of variable parts. `gen import library`

2. Build

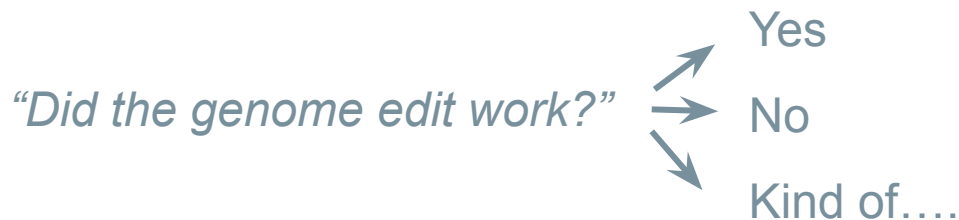
- a. Plan stepwise cloning strategy `gen derive chunks`
- b. Order or reuse synthetic material (\$)
- c. Assemble intermediate constructs & edit genome `gen make stitch`

Sequence specific commands (abridged)

3. Test

- a. Read back the genome (DNA sequencing)

```
gen update vcf  
gen view
```



- b. Evaluate performance of experimental strains

```
gen translate  
gen diff  
gen propagate-annotations
```

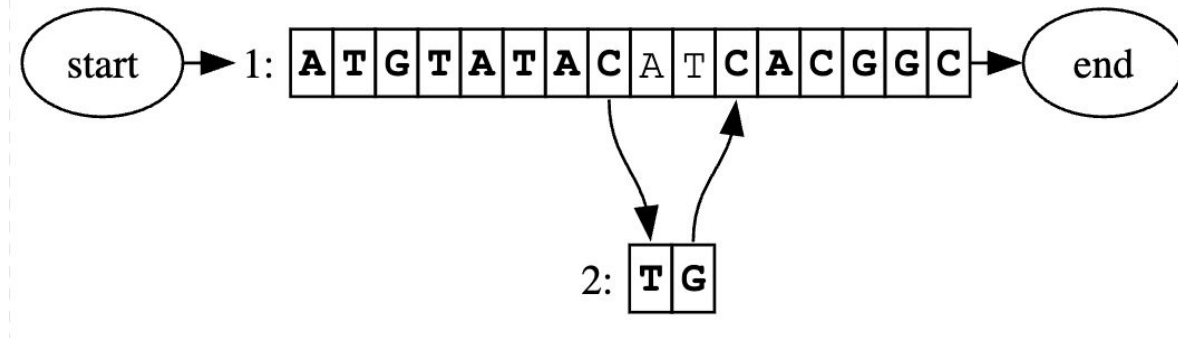
Challenges of working with genetic sequences

- Coordinate frames are very **fragile**
- A genome is neither **uniform** nor **static**.
 - Real samples rarely match “the reference sequence”
- Need to handle both **intended** and **observed** variants
 - Sequencing to confirm a genotype drops information
 - Sequencing without priors loses engineering context



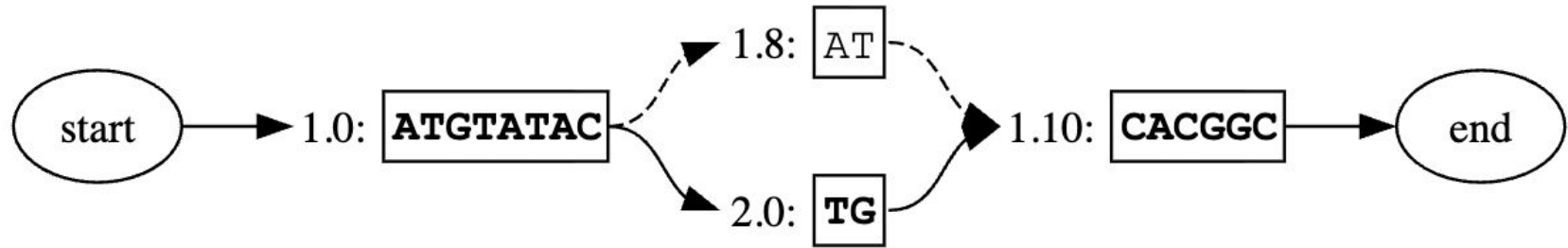
Solution: graph representation of genetic sequences

- Gen models sequences as **walks through a graph**
 - Inspired by the field of pangenomics
- Nodes: sequence fragments
- Edges: variant routes

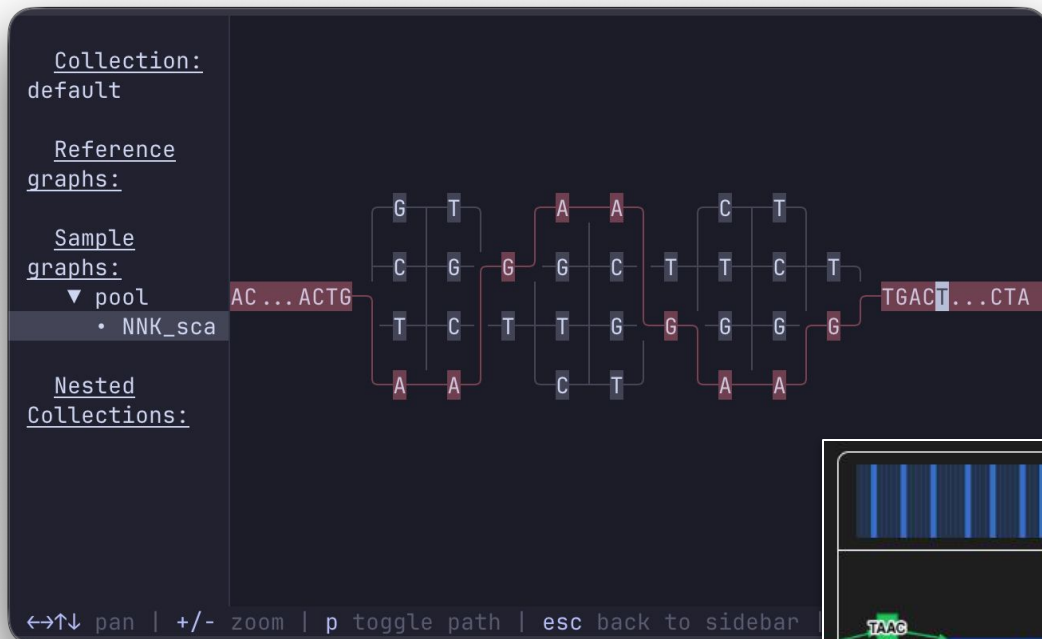


Graph representation of genetic sequences

- Fork: multiple variants exist at this locus
 - Mixed populations, polyploidy
 - Historical variants (changelog)
 - Screening library



“Schrödinger’s molecule”



Tools to work with graphs:
Visualization, subgraph
extraction, stitching, ...

Terminal User Interface ↗

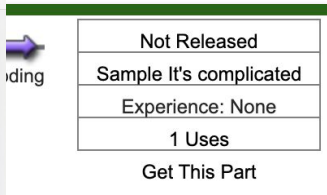
Web interface (genhub.bio) →



Promoting collaborative engineering

- Support for common bioinformatics file formats
 - FASTA, VCF, GFA, GenBank
- Decentralized distribution via patch files
- Synchronize to remote repositories
 - file:// protocol
 - https:// to genhub.bio
- <https://github.com/genhub-bio/gen>

In some embodiments, the one or more amino acids comprising: R at a residue corresponding to the residue corresponding to position 354 in SEQ ID NO: 1; or a combination thereof. In some embodiments, the one or more amino acids comprises one or more amino acid alterations selected from the group consisting of: S299W, A238G, A134E, A134R, M389S, I48V, and combinations thereof. In some embodiments, the engineered



In some embodiments, the one or more amino acids comprising: R at a residue corresponding to the residue corresponding to position 354 in SEQ ID NO: 1; or a combination thereof. In some embodiments, the one or more amino acids comprises one or more amino acid alterations selected from the group consisting of: S299W, A238G, A134E, A134R, M389S, I48V, and combinations thereof. In some embodiments, the engineered

screen

chris-mitchell

Jan 9, 2026, 02:03 PM

here

DIFF

OPERATIONS

Browse samples, collections, and block groups added in this pull request to see how the changes affect your data.

Change scope
Pull request

View overall pull request or a specific operation

Database
default.db

Collections

default

Samples

screen

Block Groups

U00096



C

Chris Mitchell @chris-mitchell · just now

The design looks sensible to me. Could you include the code you used to generate the knockout library too?

[Reply](#) [Resolve](#)

WRITE PREVIEW

Write a general comment...

Jobs Ran

No jobs have been recorded
pull request yet.