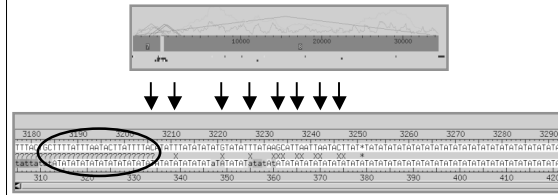


## Common Finishing Errors in GEP Submissions

.....

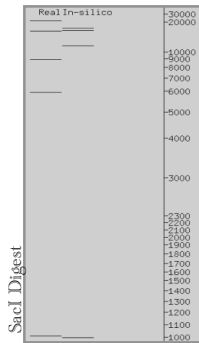
Projects from the *D. grimshawi* dot chromosome  
By Jeannette Wong

### Improper Force Join



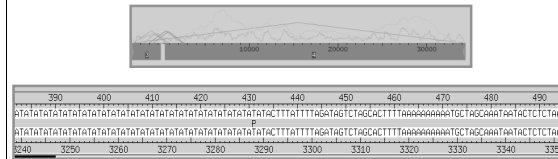
- A student's attempt at a force join fails to consider the high quality discrepant bases between the two consensus sequences.

### Digests after Improper Force Join...



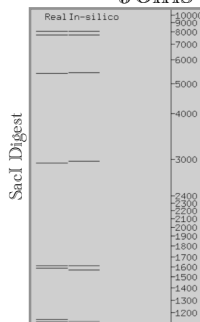
- real vs. *in-silico* digest comparison
- red bands > 2% difference in size
- poor matches between the two digests
- does not support assembly, indicates a misassembly

### Proper Force Join



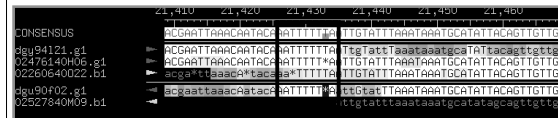
- proper force join contains no high quality discrepancies between the consensus sequences

### Digests after Proper Joins and Correct Assemblies



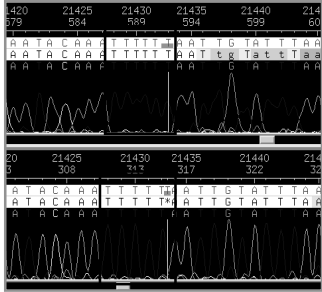
- exact match between real and *in-silico* fragment sizes
- verifies proper force join and assembly of project

### Miscalled Insertion/Deletion Sites



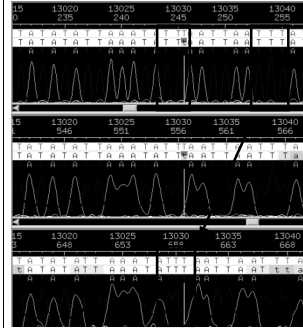
- navigator issue in the 21 kb region concerning the number of T's in the mononucleotide run
- high quality consensus suggests six T's, some reads suggest only five
  - student tagged this as an insertion/deletion site

## Checking the Quality of the Reads...



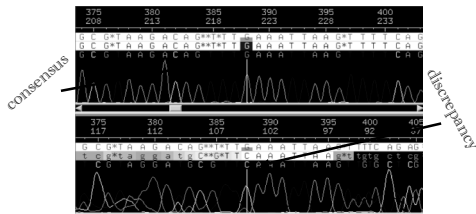
- 6 T's is low quality
- narrower spacing pattern among 6 T's compared to adjacent nucleotides
- 5 T's is of higher quality and has regular nucleotide spacings
- therefore, mononucleotide run should only contain 5 T's
- not an indel site, but a sequencing artifact

## Improper Manual Edits



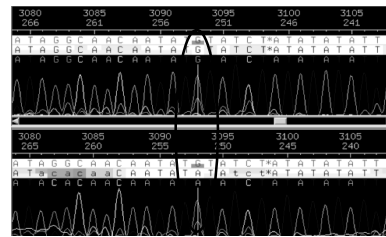
- student manually edited T-douplet to T-triplet
- a single low quality read justifies T-triplet
- T-triplet spacing is narrower in comparison to adjacent triplet in consensus sequence
- high quality reads support doublet with regular spacing
- ergo, not a T triplet: student should have comment tagged sequencing artifact from one read

## Miscalled Single Nucleotide Polymorphism



- student miscalled high quality discrepancy as possible SNP
- poor quality sequence, not reliable enough to call SNP

## Identified Single Nucleotide Polymorphism



- both high quality, same sequence with one nucleotide difference not present anywhere else on fosmid
  - correctly tagged by student