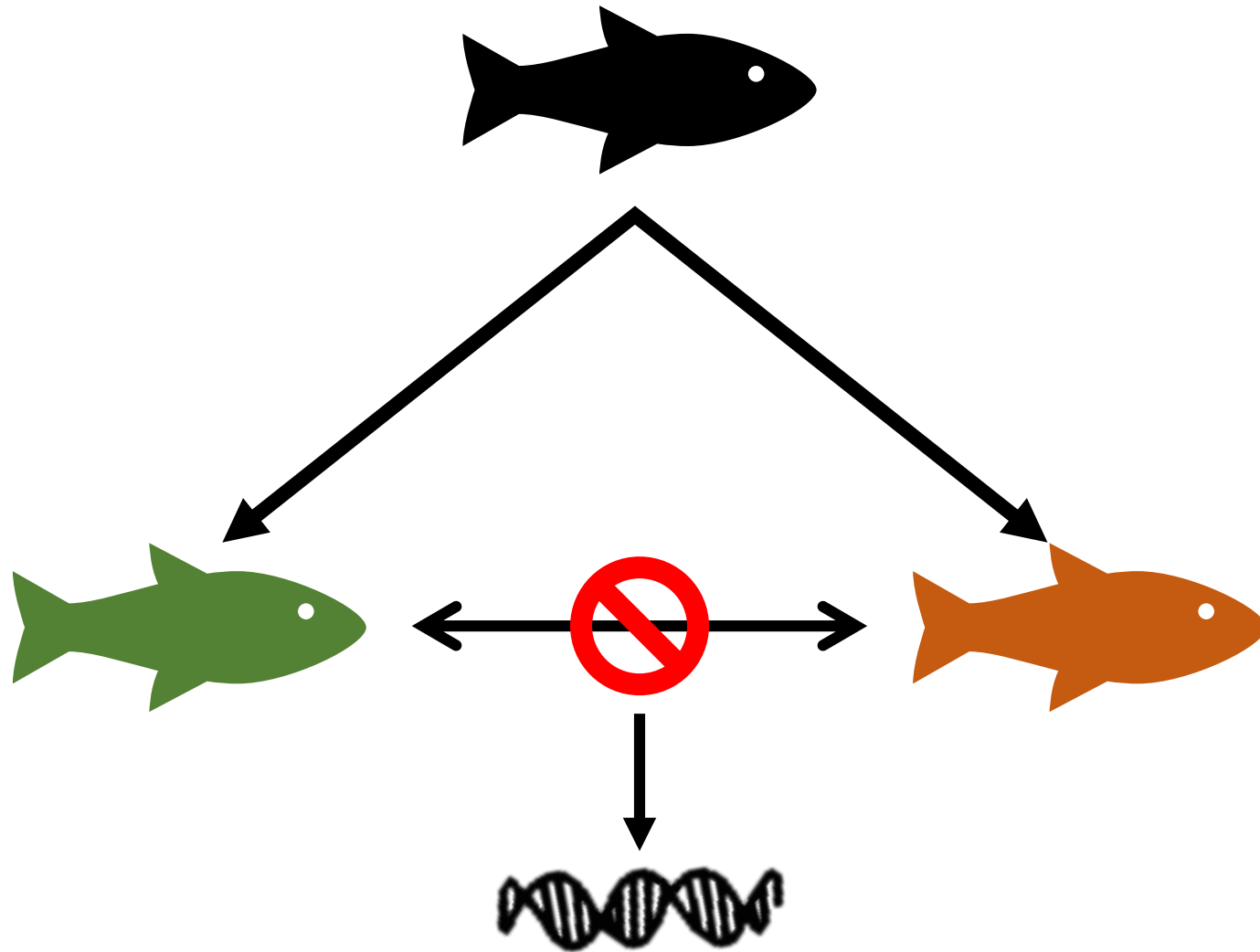


# Using HTC for genomic ancestry analysis

Megan Frayer  
Ph.D. Student, Genetics

July 21, 2017

# Genetics of Speciation



The house mouse hybrid zone can tell us about how speciation is proceeding between these subspecies

*M. m. domesticus*

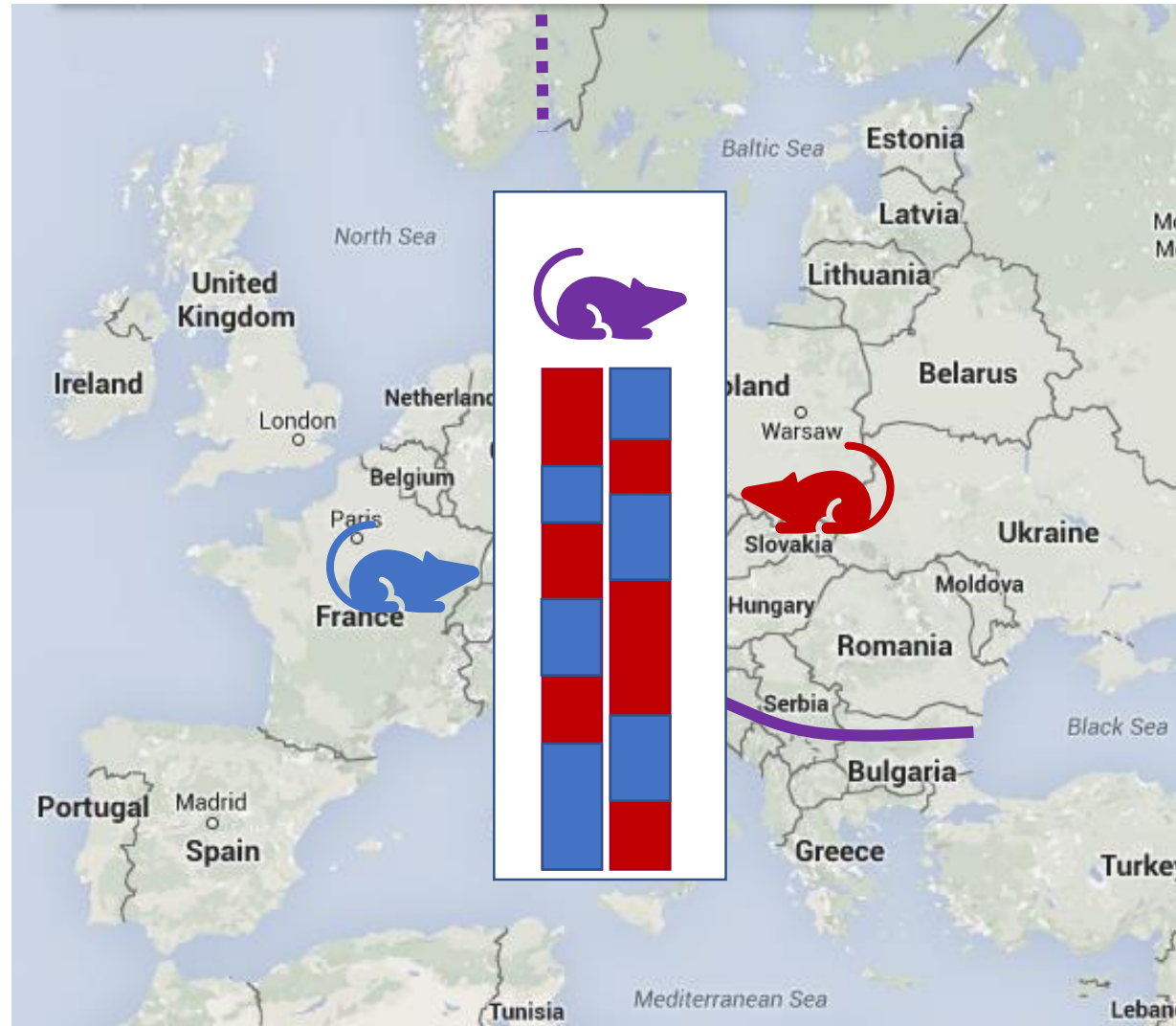


*M. m. musculus*



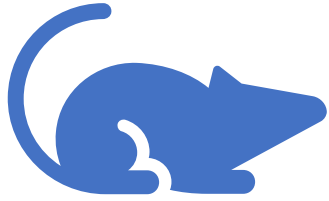
The house mouse hybrid zone can tell us about how speciation is proceeding between these subspecies

*M. m. domesticus*

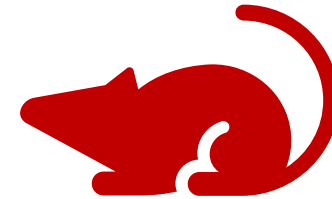


*M. m. musculus*





ATCGTCAGTCAGTCGATCGATACGTA GCATGCAGTACGATGCAGTACGATGATACG  
TAGCAGTCAGACACGTAGCTATGCAT CGTACGTCATGCTACGTCATGCTACTATGC



# DAG for Junction Inference

Phase the  
source  
population  
panels

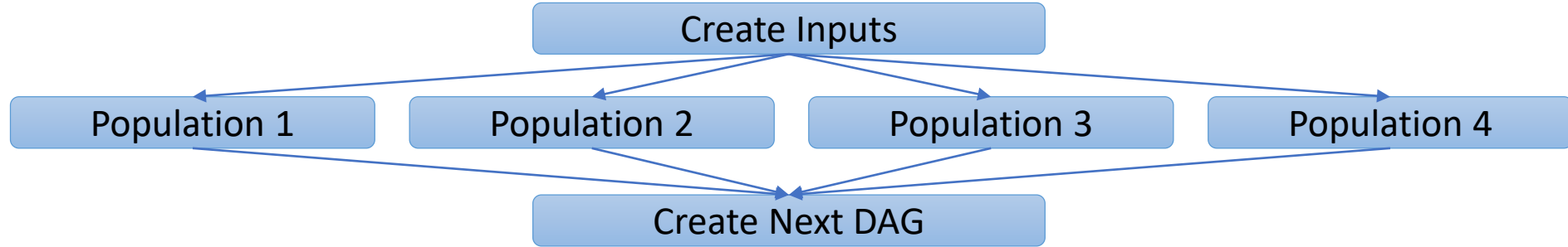
Create the  
input files

Run the  
inference  
and  
analysis

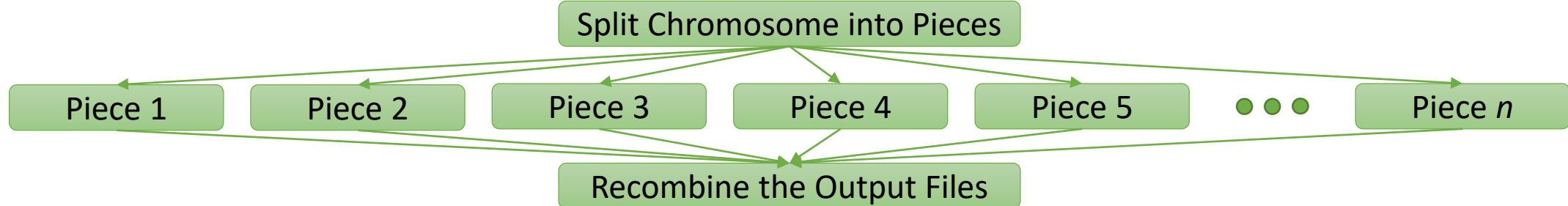


chr1.dag

SPLICE beagle.dag



SUBDAG\_EXTERNAL inputs.dag

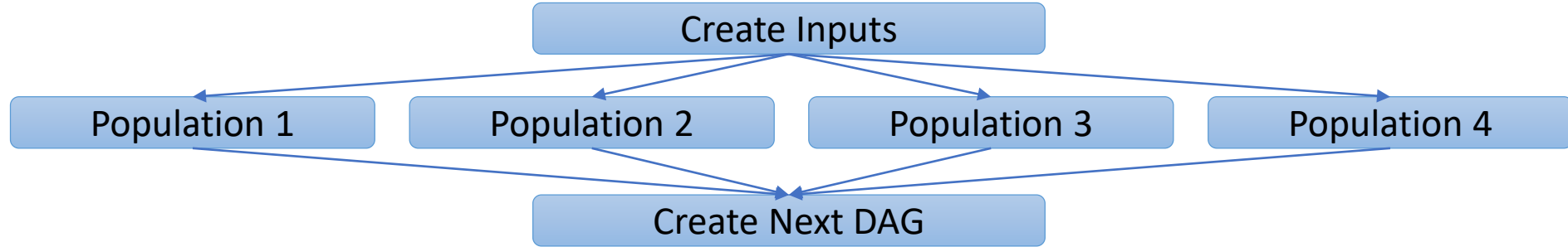


Run Inference Program

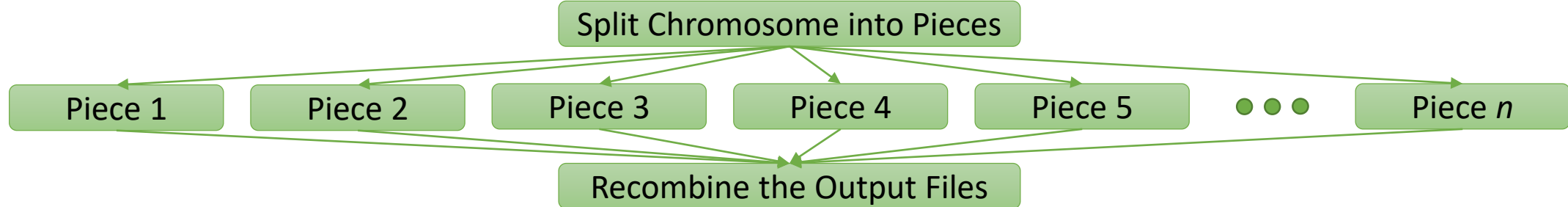
Analyze the Output

chr1.dag

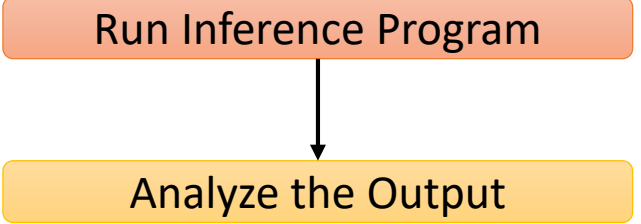
SPLICE beagle.dag



SUBDAG\_EXTERNAL inputs.dag



Before HTC: 72 hours/test  
57 days/19 tests  
With HTC: 14 hours/test  
4 days/19 tests  
**57 days → 4 days**





# Parameter grid search

- What is the combination of input parameters with the highest likelihood?

# Parameter grid search

| Parameter               | Values to be tested |         |         |         |         |         |         |         |         |         |
|-------------------------|---------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| <b>defaultRate</b>      | 0.8                 | 0.86    | 0.99    | 1.15    |         |         |         |         |         |         |
| <b>timeSince</b>        |                     |         |         |         |         |         |         |         |         |         |
| <b>Admixture</b>        | 1000                | 3750    | 6500    | 9250    | 12000   | 14750   |         |         |         |         |
| <b>ancestryProp1</b>    | 0.4                 | 0.5     | 0.6     |         |         |         |         |         |         |         |
| <b>ancestralRate1</b>   | 41000               | 69250   | 97500   |         |         |         |         |         |         |         |
| <b>ancestralRate2</b>   | 14000               | 23650   | 33290   | 20815   | 35158   | 49500   |         |         |         |         |
| <b>mutation1</b>        | 1E-04               | 1E-05   | 1E-06   | 1E-07   | 1E-08   |         |         |         |         |         |
| <b>mutation2</b>        | 3.4E-05             | 3.4E-06 | 3.4E-07 | 3.4E-08 | 3.4E-09 | 5.1E-05 | 5.1E-06 | 5.1E-07 | 5.1E-08 | 5.1E-09 |
| <b>miscopyRate</b>      | 0.01                | 0.001   | 1E-04   | 1E-05   | 1E-06   |         |         |         |         |         |
| <b>Miscopy Mutation</b> | 0.01                | 0.001   | 1E-04   | 1E-05   | 1E-06   |         |         |         |         |         |

108,000 combinations of parameters to be tested

# Parameter grid search

Create  
input files

Run  
parameter  
tests

Compile  
and  
analyze  
results



parameter\_test.dag

Create Input  
Files

Examples of files to print:

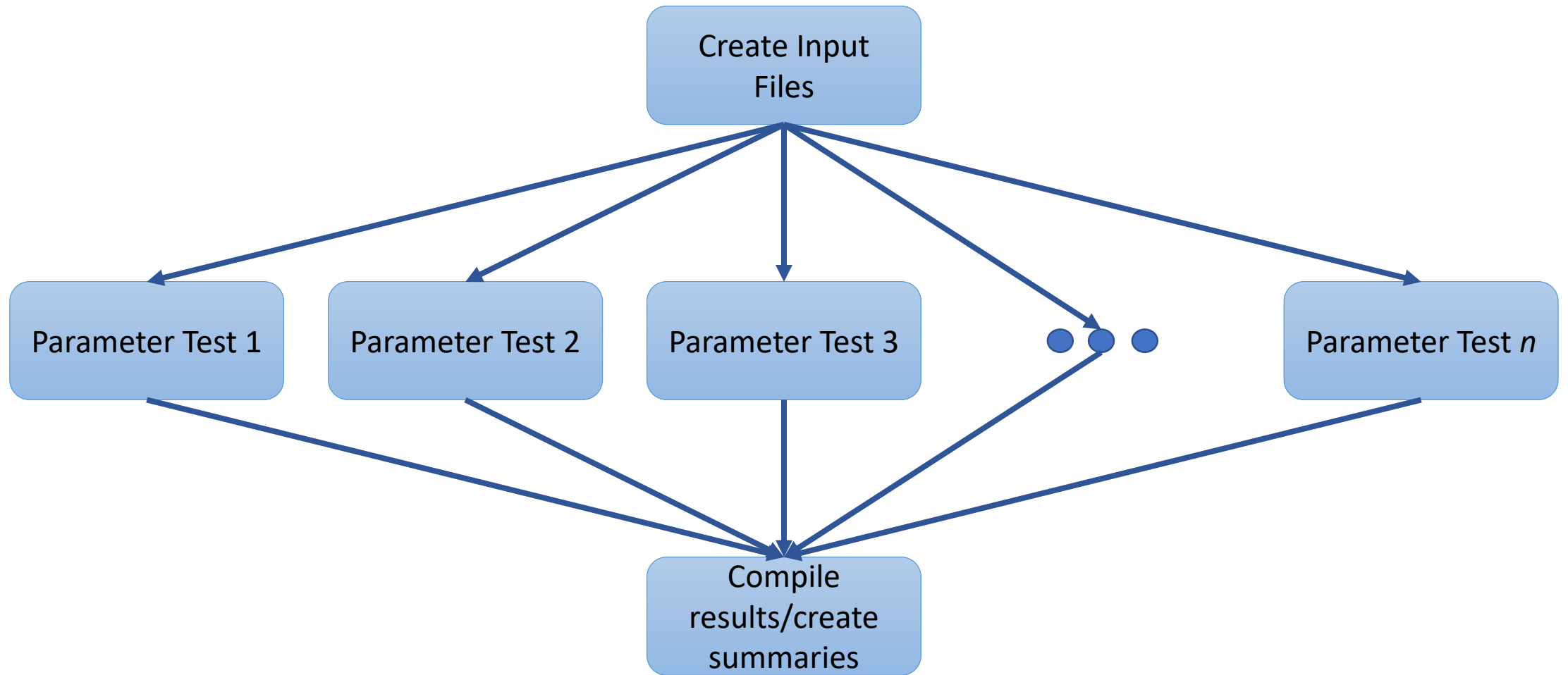
Submit files

Executables

Input for programs being run

Scripts that will need to be run

parameter\_test.dag



parameter\_test.dag

Create Input  
Files

SUBDAG\_EXTERNAL

Parameter Test 1

Parameter Test 2

Parameter Test 3

...

Parameter Test *n*

Before HTC: 2 hours/test  
24.6 years/108,000 tests  
With HTC: 2 hours/test  
10 days/108,000 tests  
**24.6 years → 10 days**

Compile  
results/create  
summaries

# Testing with Simulated Chromosomes

- How well is the program performing?

# Testing with Simulated Chromosomes

Simulate  
Chromosomes

Determine the  
true junction  
map

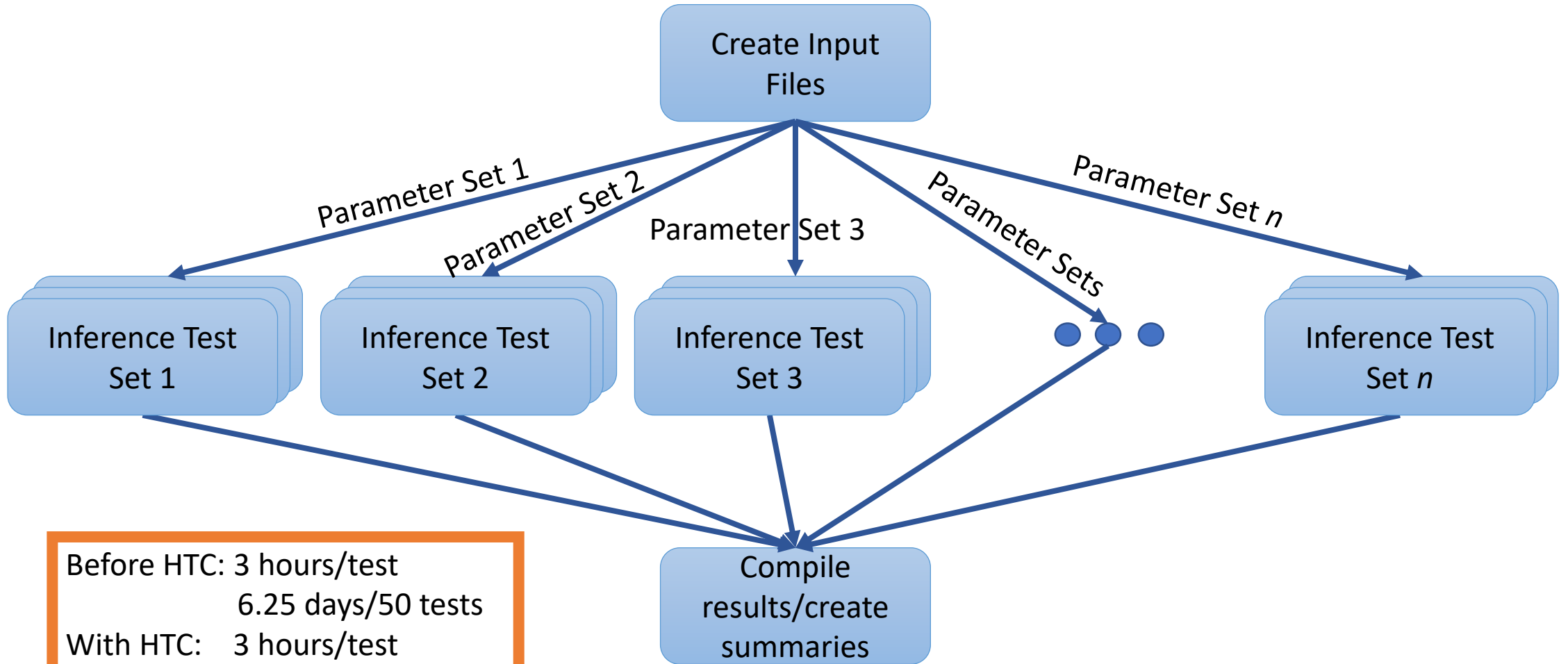
Infer junctions  
using the  
method to be  
tested

Compare the  
true and  
inferred maps





inference\_testing.dag



Before HTC: 3 hours/test  
6.25 days/50 tests  
With HTC: 3 hours/test  
10 hours/50 tests  
**6.25 days → 10 hours**

# Conclusion

- HTC can improve research in biological sciences
- Even simple DAGs can make a big impact on your research
- DAGs can also improve reproducibility

HTC has shortened my Ph.D. by 39.8 years so far.