Genomic Ancestry Analysis in Wild Hybrid House Mice

Megan Frayer

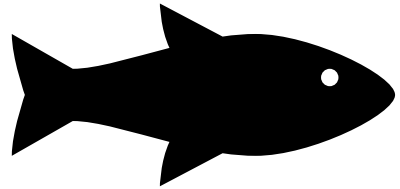
Ph.D. Student, Laboratory of Genetics

UW-Madison

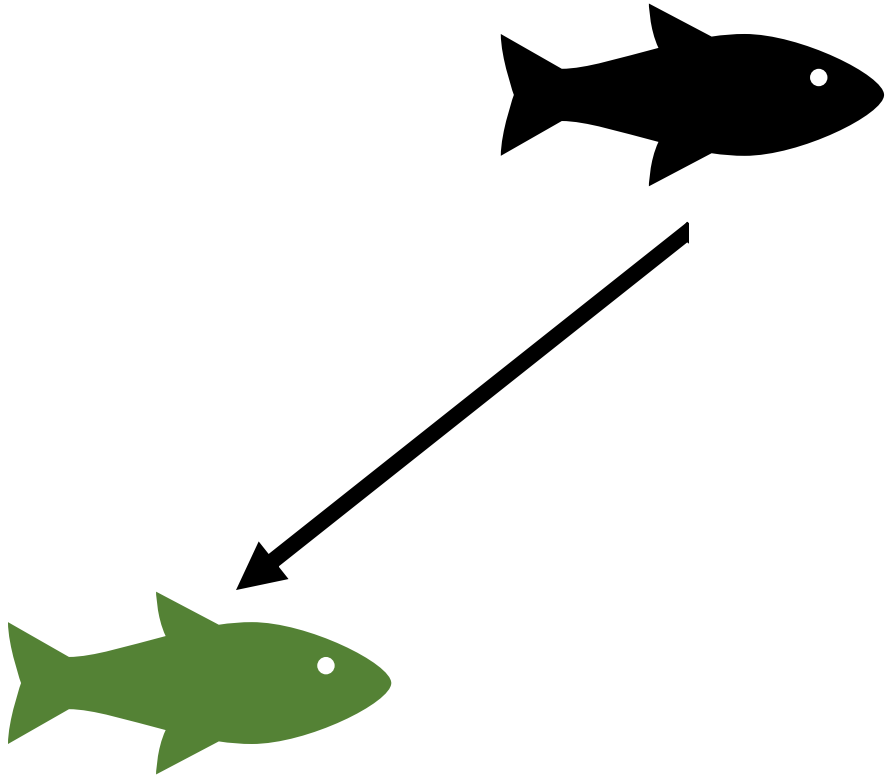
HTCondor Week 2022

Genetics of Speciation

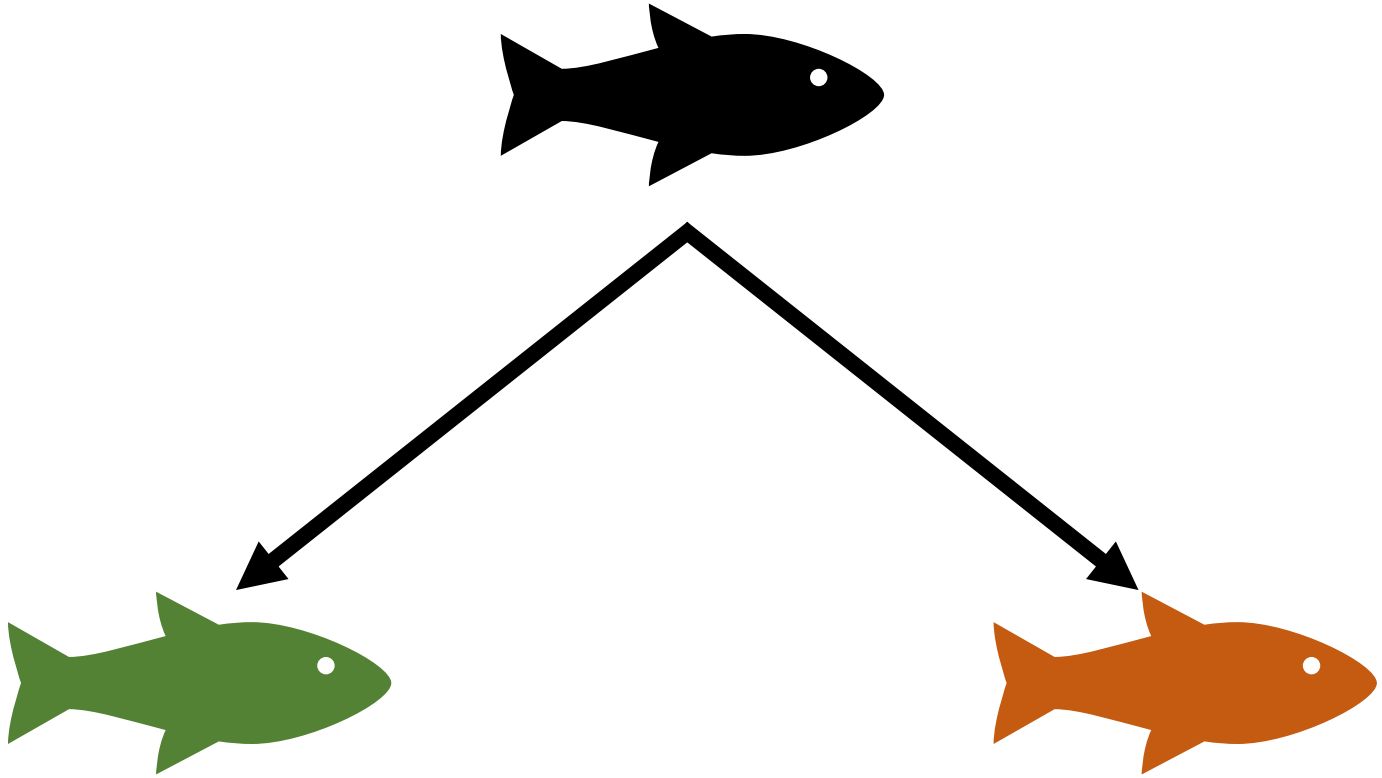
Genetics of Speciation



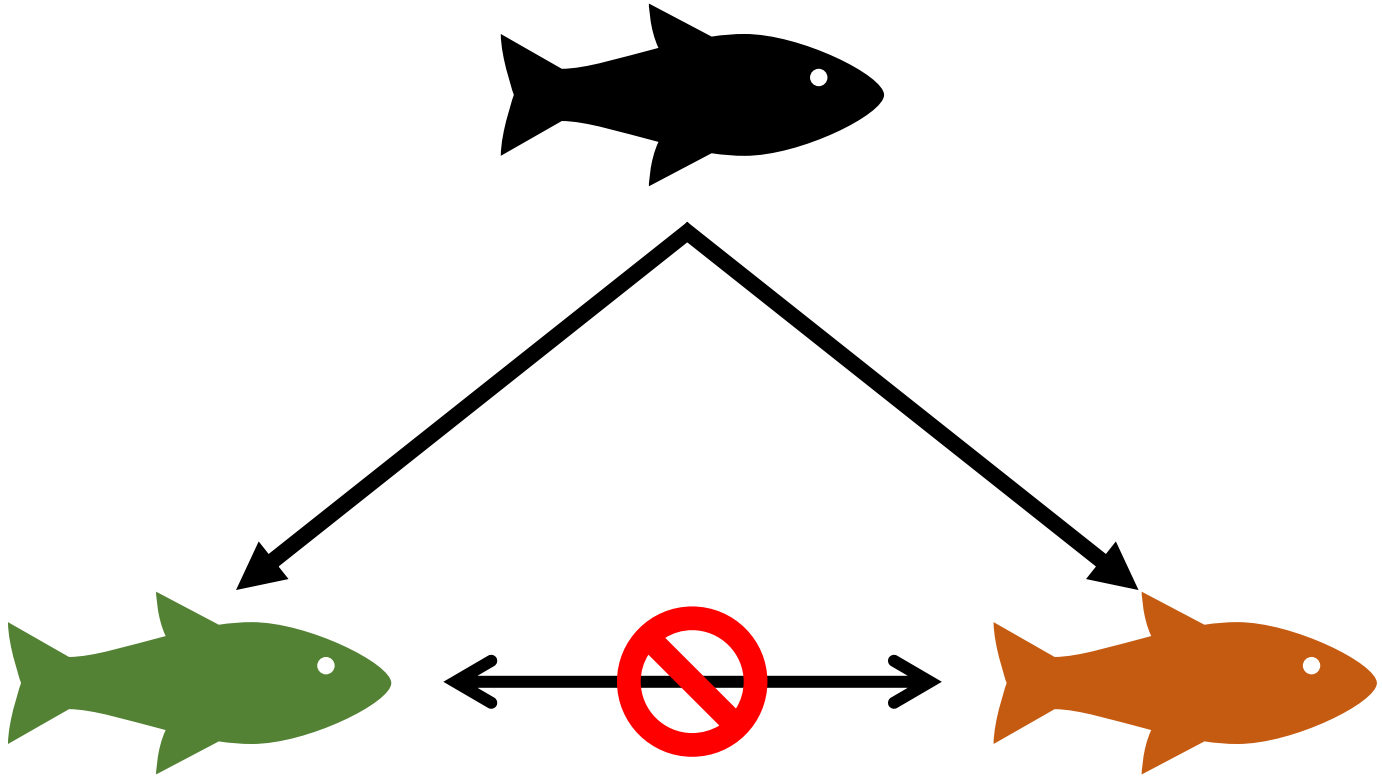
Genetics of Speciation



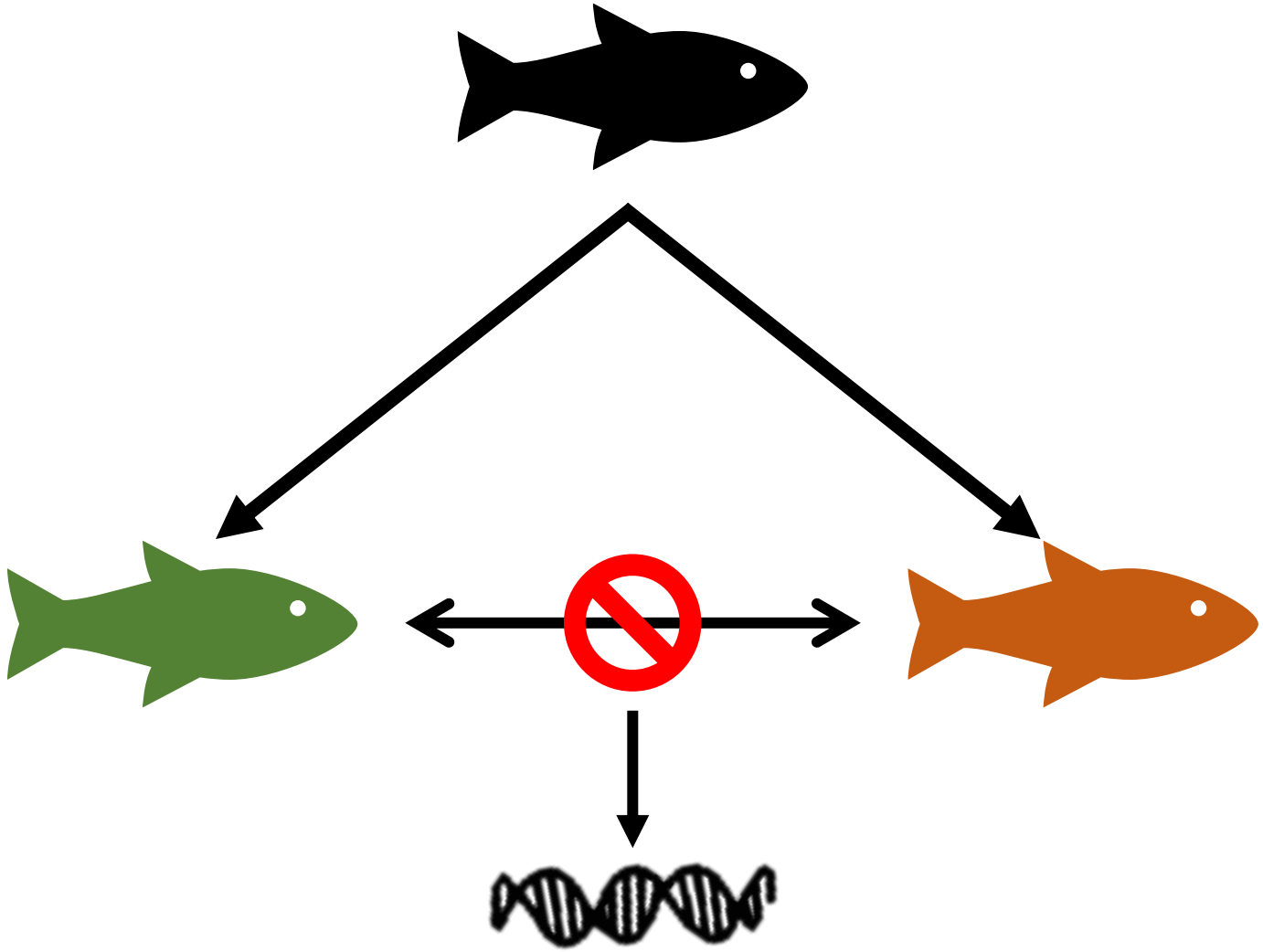
Genetics of Speciation



Genetics of Speciation

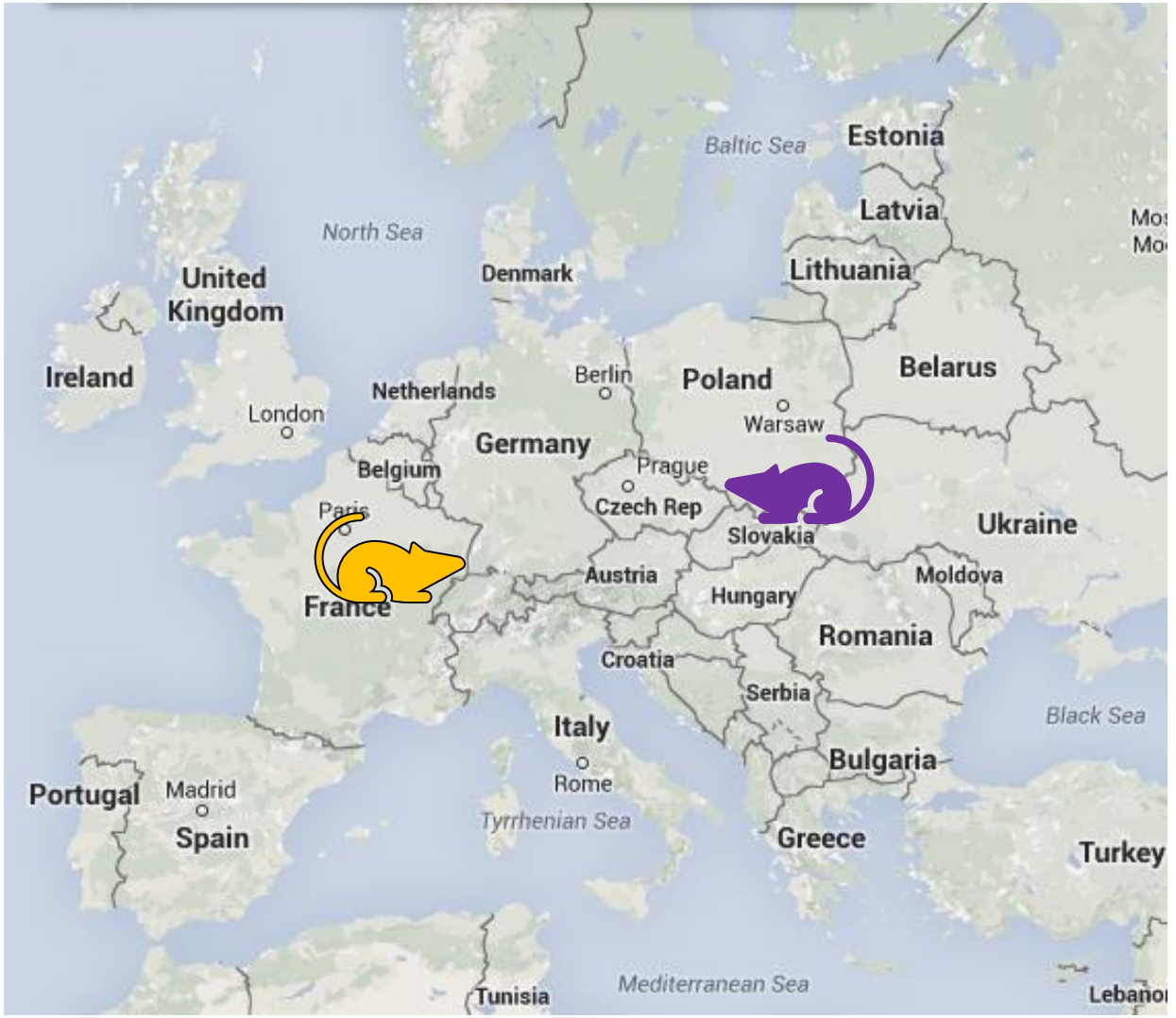


Genetics of Speciation



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

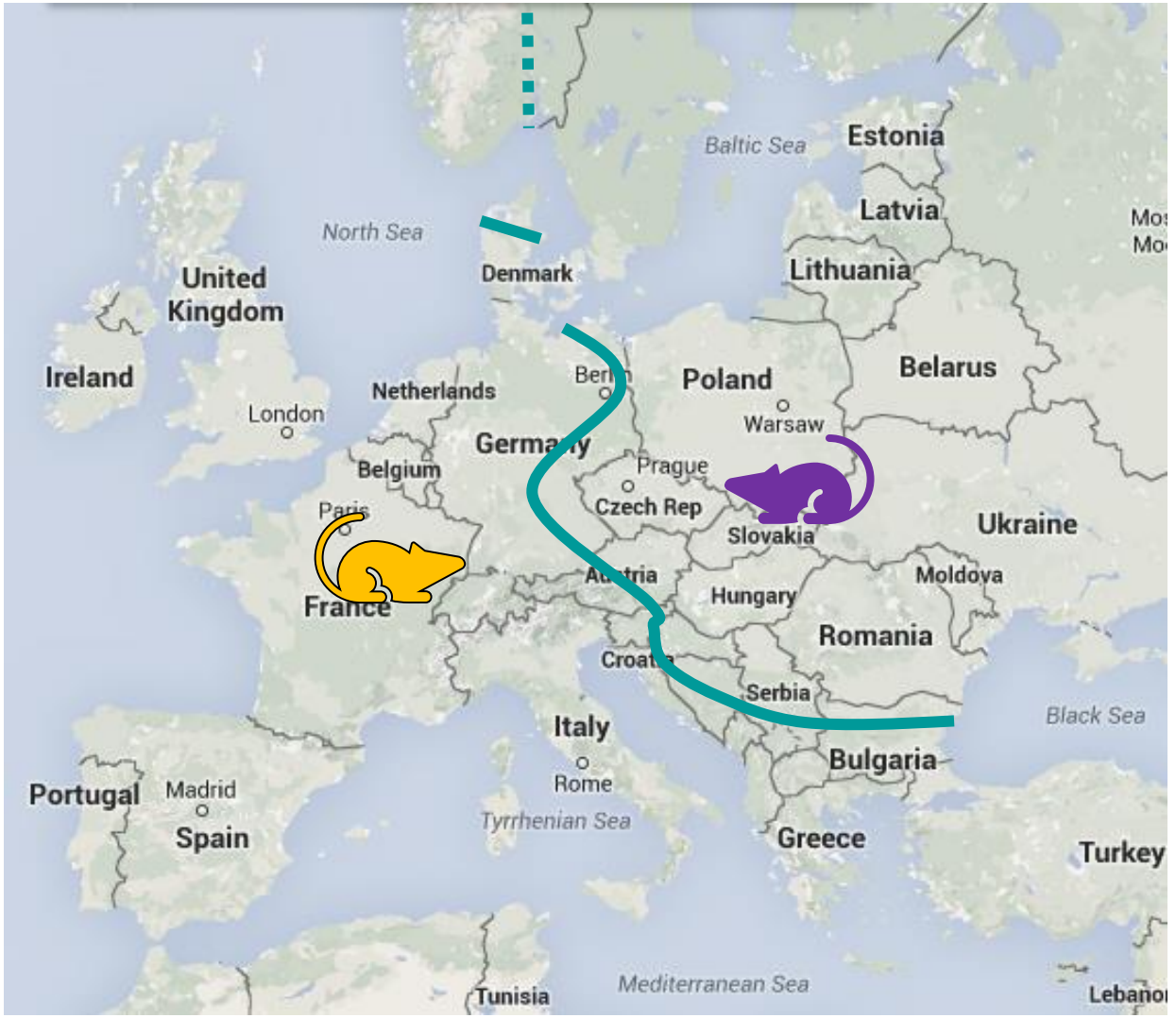
M. m. domesticus



M. m. musculus

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

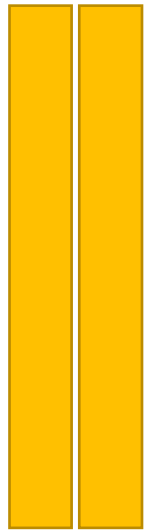
M. m. domesticus

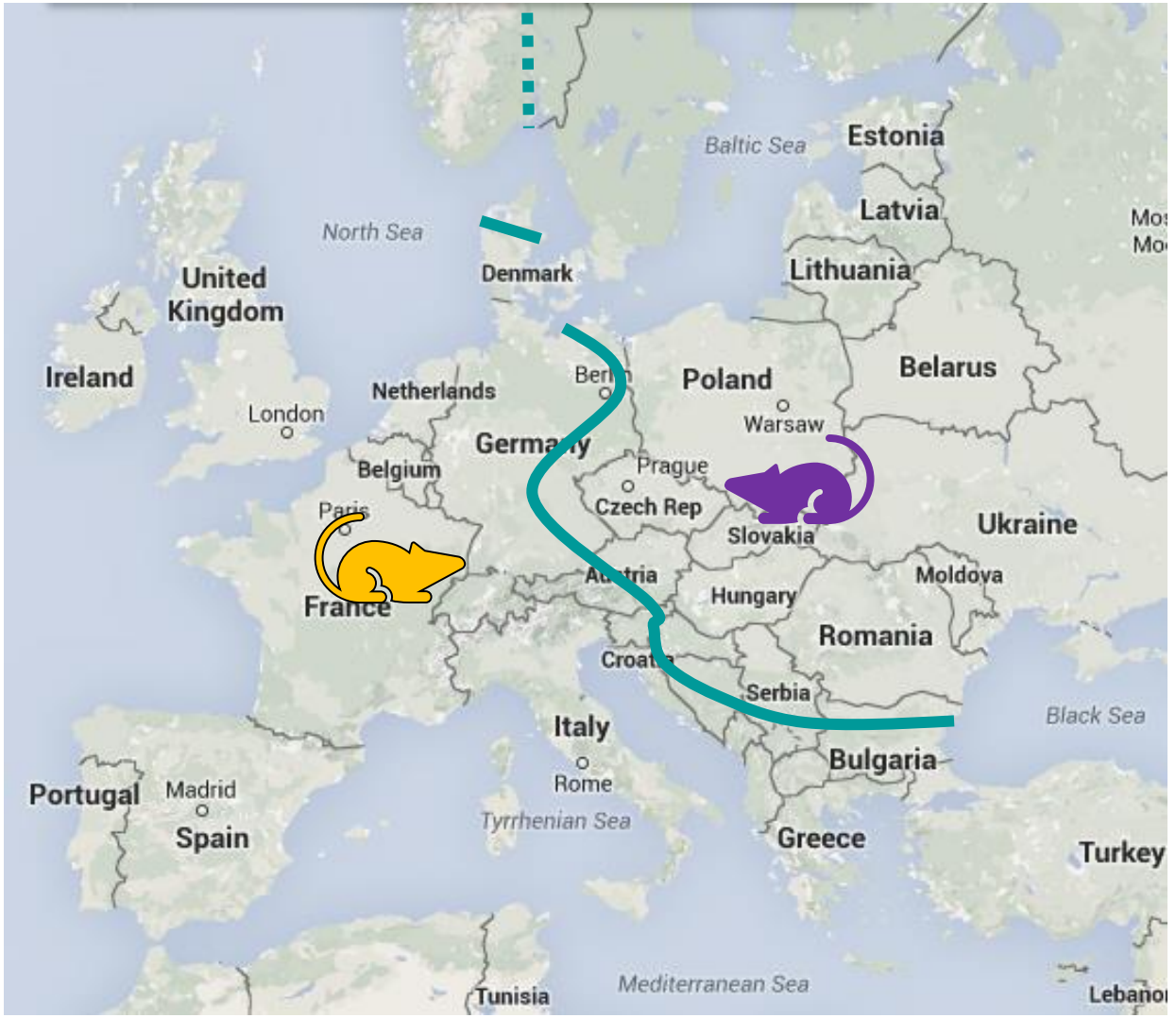


M. m. musculus

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

M. m. domesticus

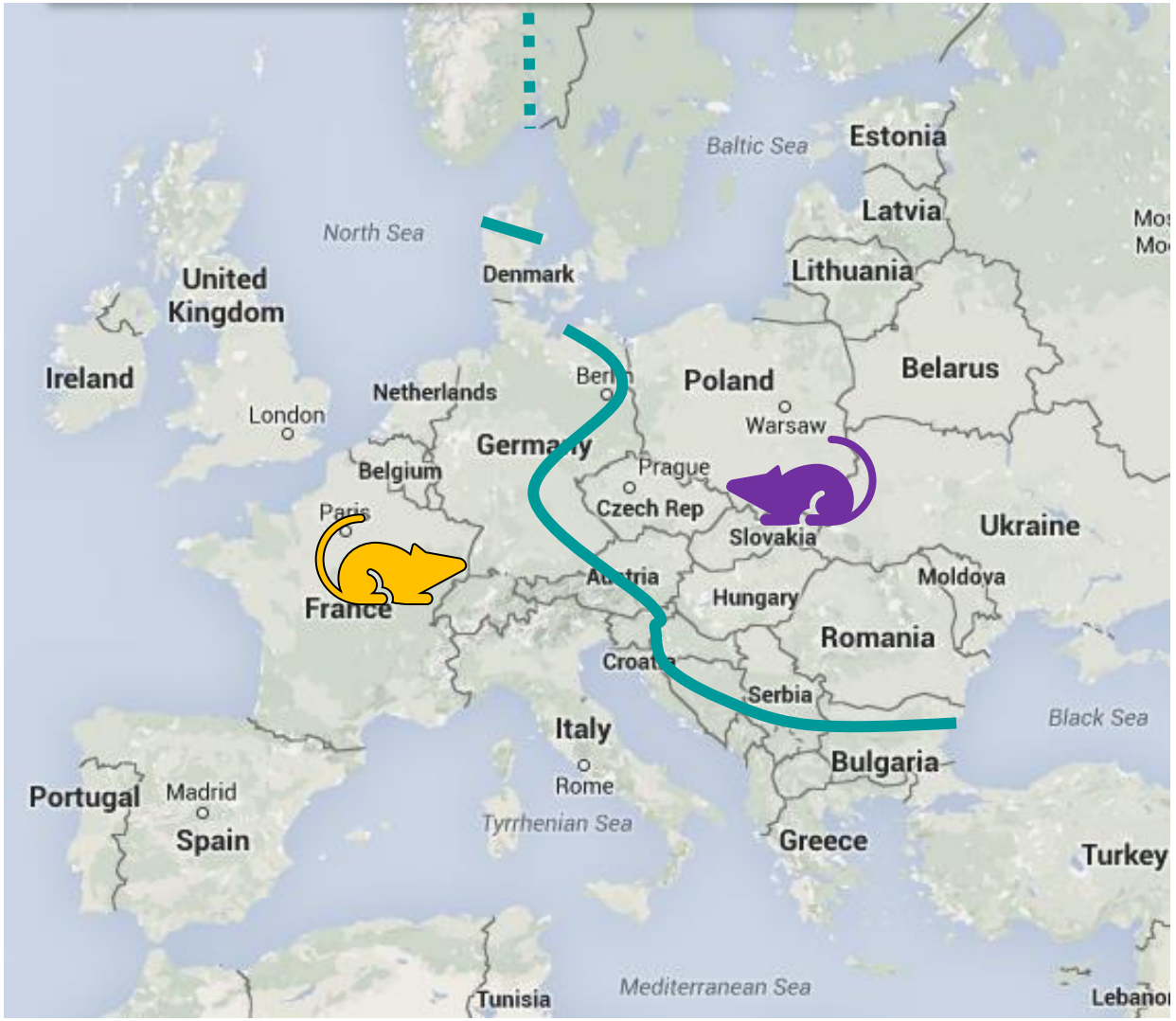




M. m. musculus

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

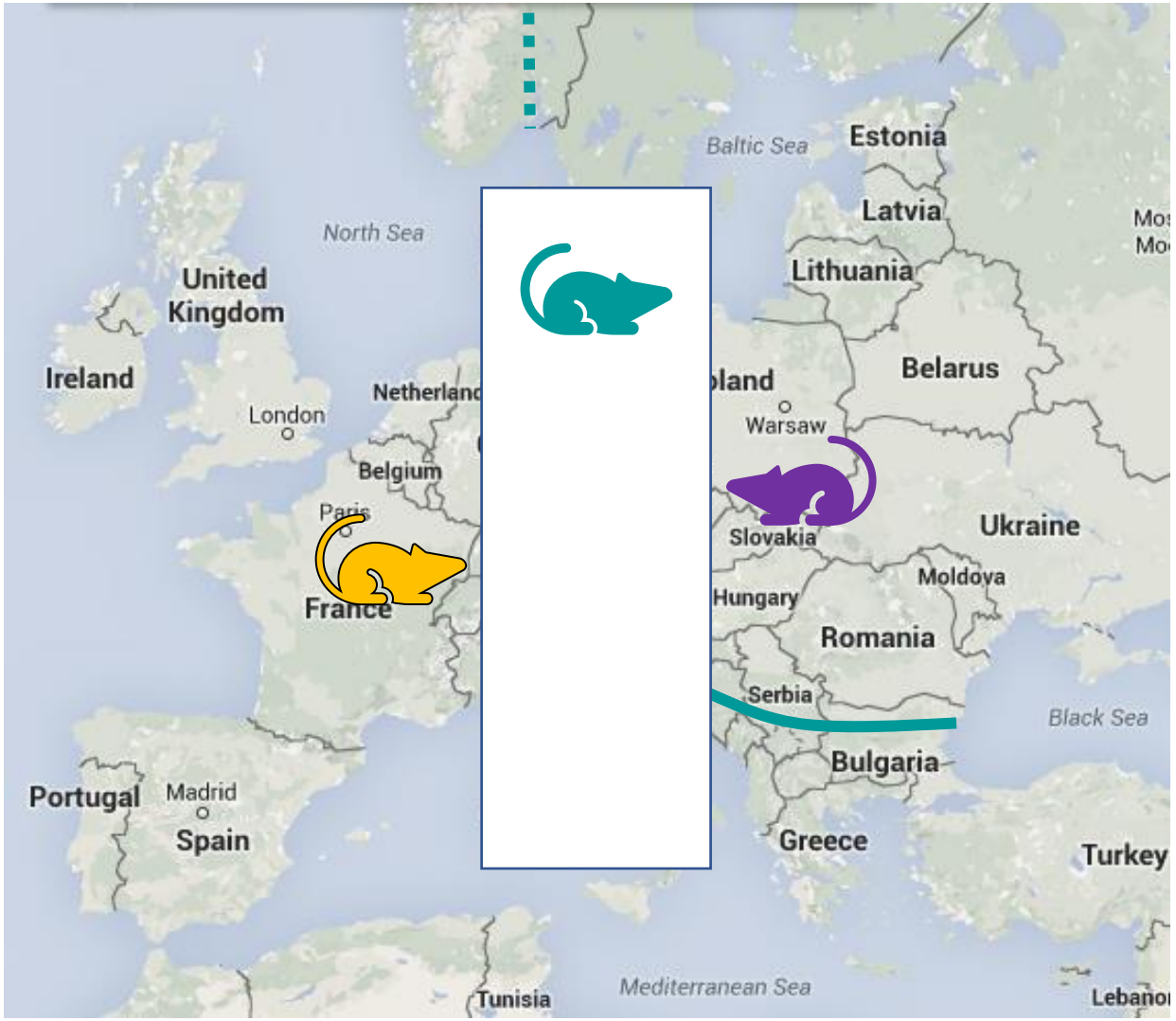
M. m. domesticus



M. m. musculus

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

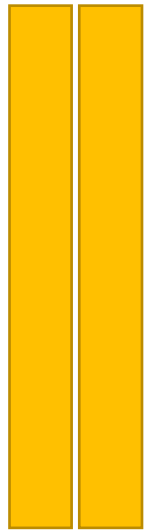
M. m. domesticus

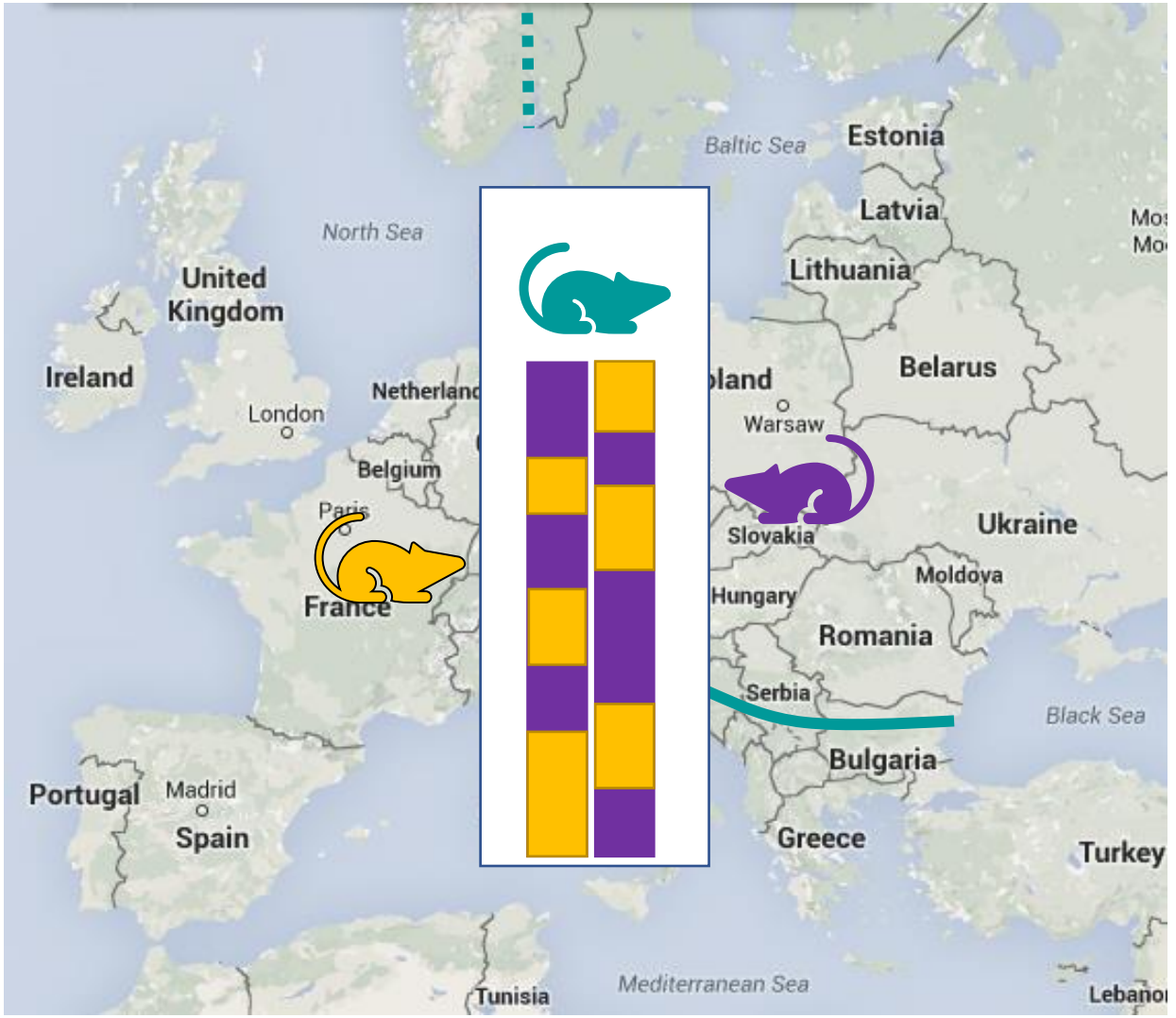


M. m. musculus

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

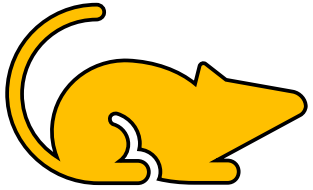
M. m. domesticus





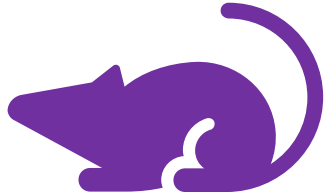
M. m. musculus

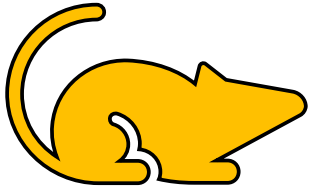
ATCGTCAGTCAGTCGATCGATACGTAGCATGCAGTACGATGCAGTACGATGATACG
TAGCAGTCAGACACGTAGCTATGCATCGTACGTCATGCTACGTCATGCTACTATGC



ATCGTCAGTCAGTCGATCGATACGTAGCATGCAGTACGATGCAGTACGATGATACG

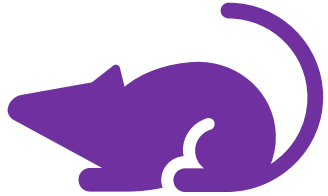
TAGCAGTCAGACACGTAGCTATGCATCGTACGTCATGCTACGTCATGCTACTATGC

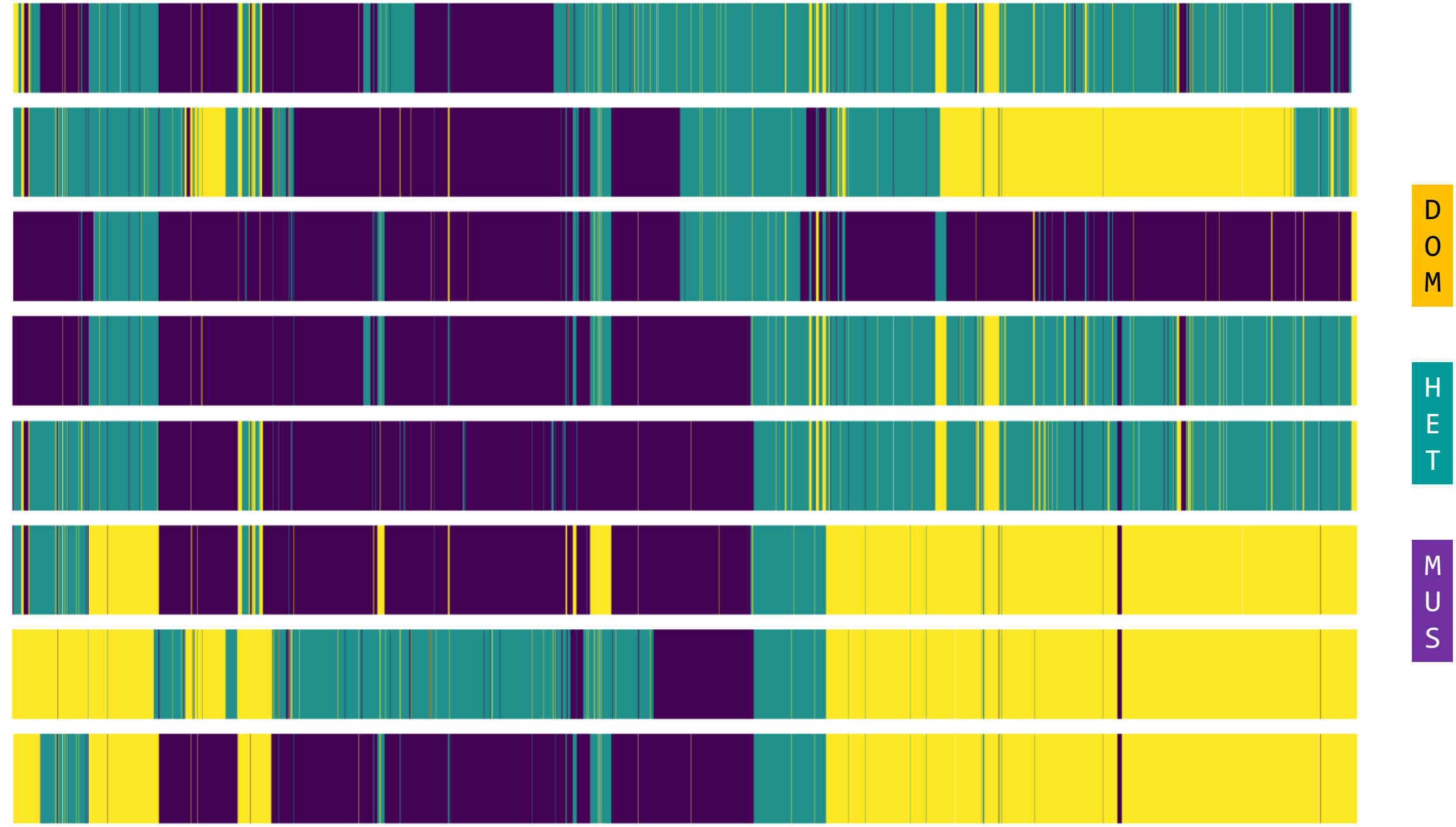




ATCGTCAGTCAGTCGATCGATACGTAGCATGCAGTACGATGCAGTACGATGATACG

TAGCAGTCAGACACGTAGCTATGCATCGTACGTCATGCTACGTCATGCTACTATGC




DOM
HET
MUS
← Centromere
Telomere →

Parameter grid search

Parameter grid search

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

Parameter grid search

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

Parameter grid search

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

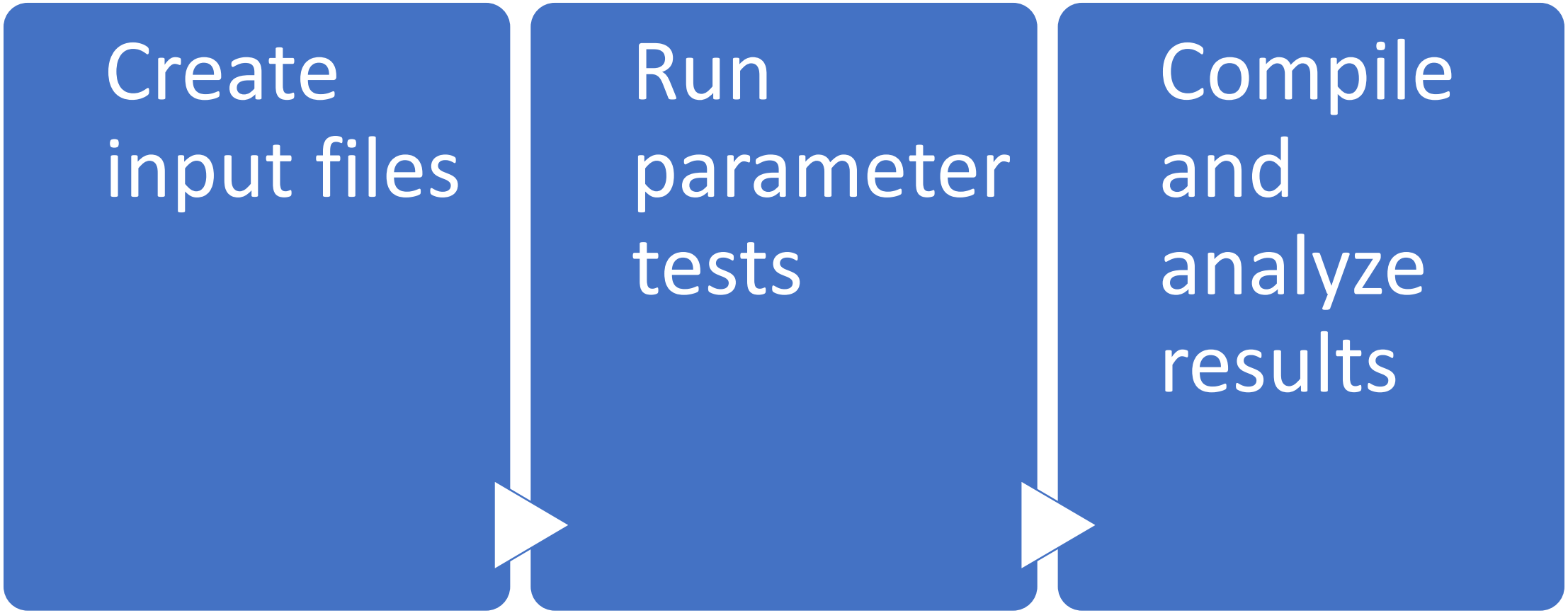
108,000 combinations of parameters to be tested

Parameter grid search

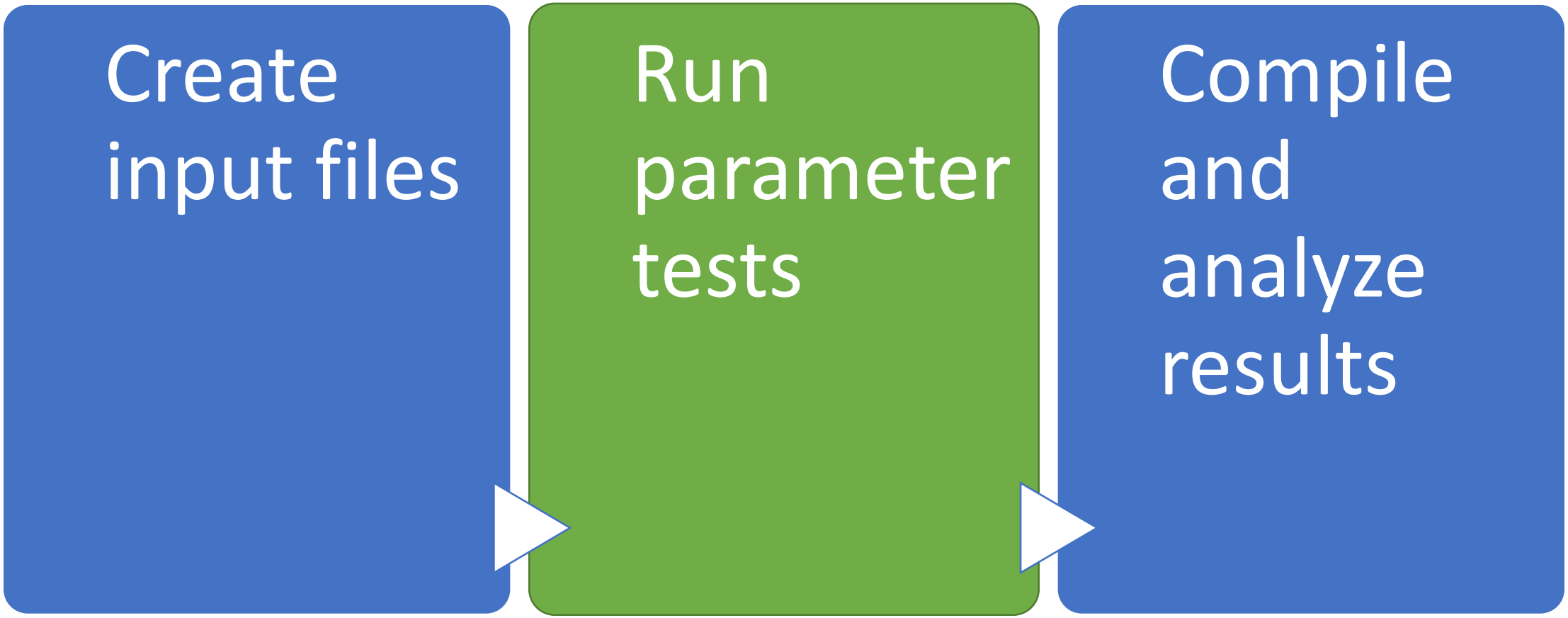
Parameter	Values to be tested									
defaultRate	0.8	0.86	0.99	1.15						
timeSince Admixture	1000	3750	6500	9250	12000	14750				
ancestryProp1	0.4	0.5	0.6							
ancestralRate1	41000	69250	97500							
ancestralRate2	14000	23650	33290	20815	35158	49500				
mutation1	1E-04	1E-05	1E-06	1E-07	1E-08	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

108,000 combinations of parameters to be tested

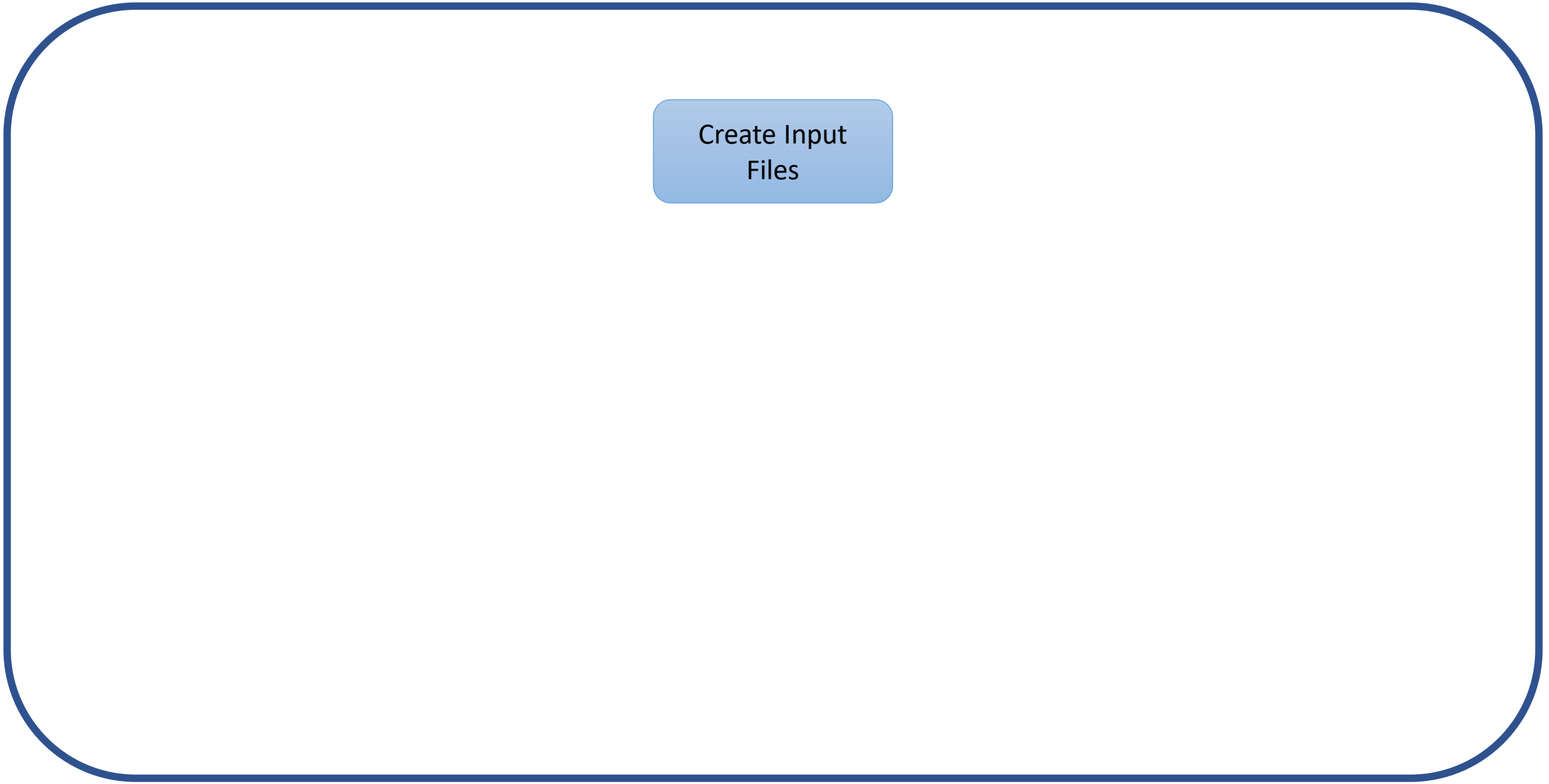
Parameter grid search



Parameter grid search



parameter_test.dag



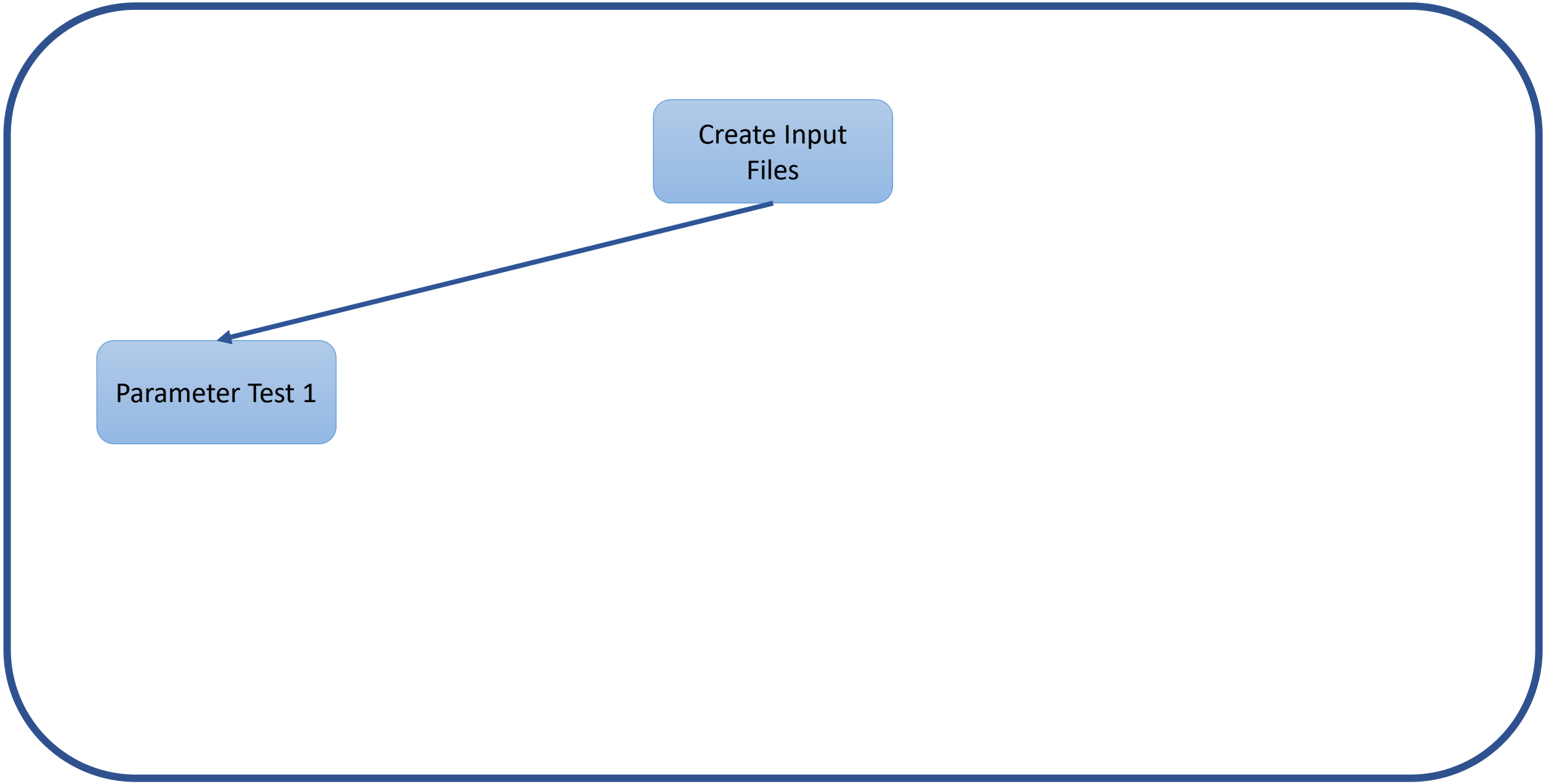
parameter_test.dag

Create Input Files

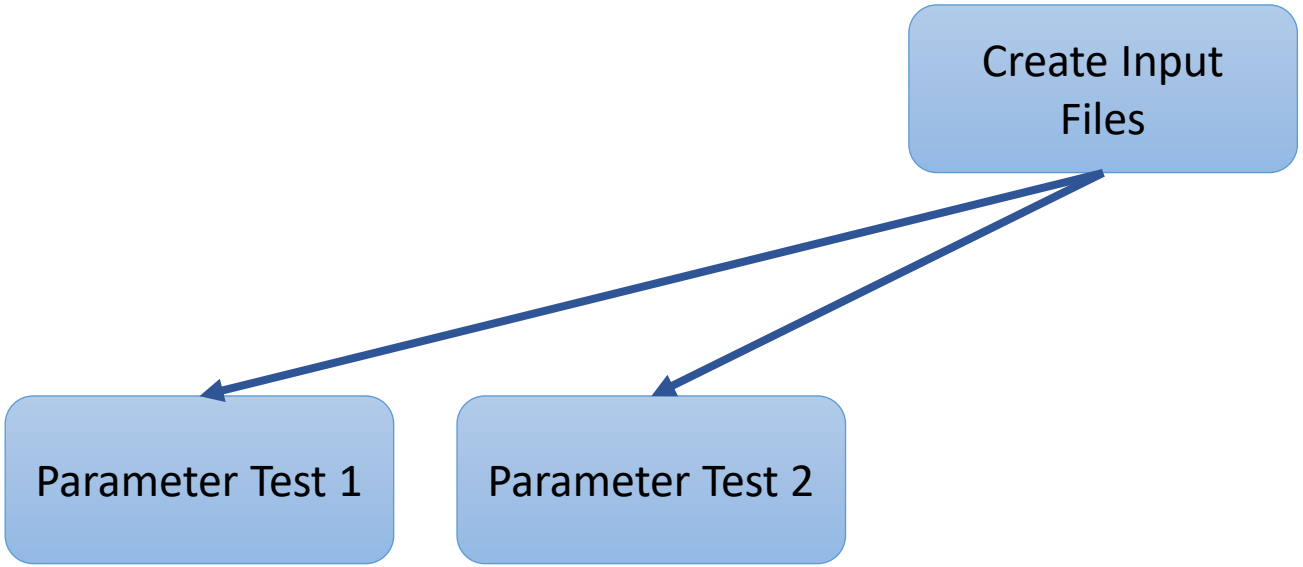
Examples of files to print:

Submit files
Executables
Input for programs being run
Scripts that will need to be run

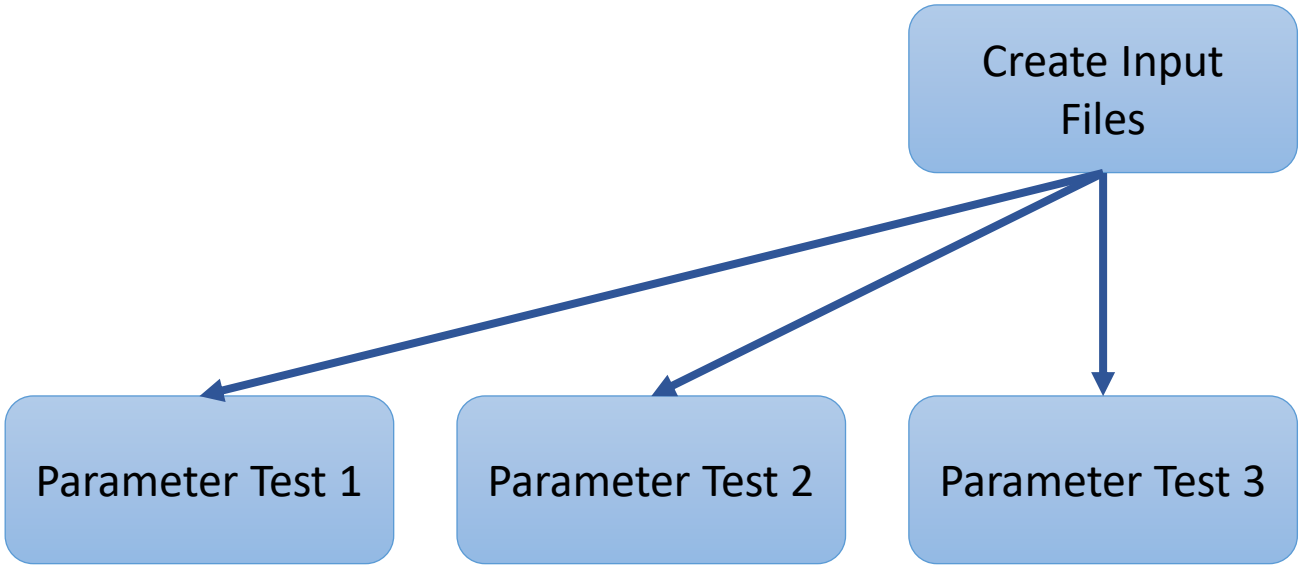
parameter_test.dag



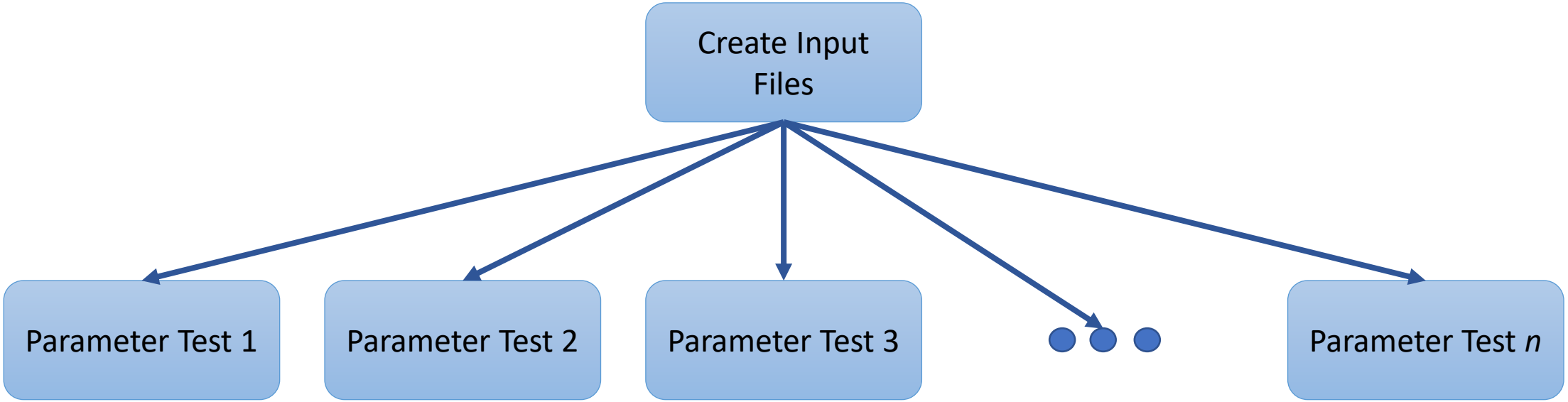
parameter_test.dag



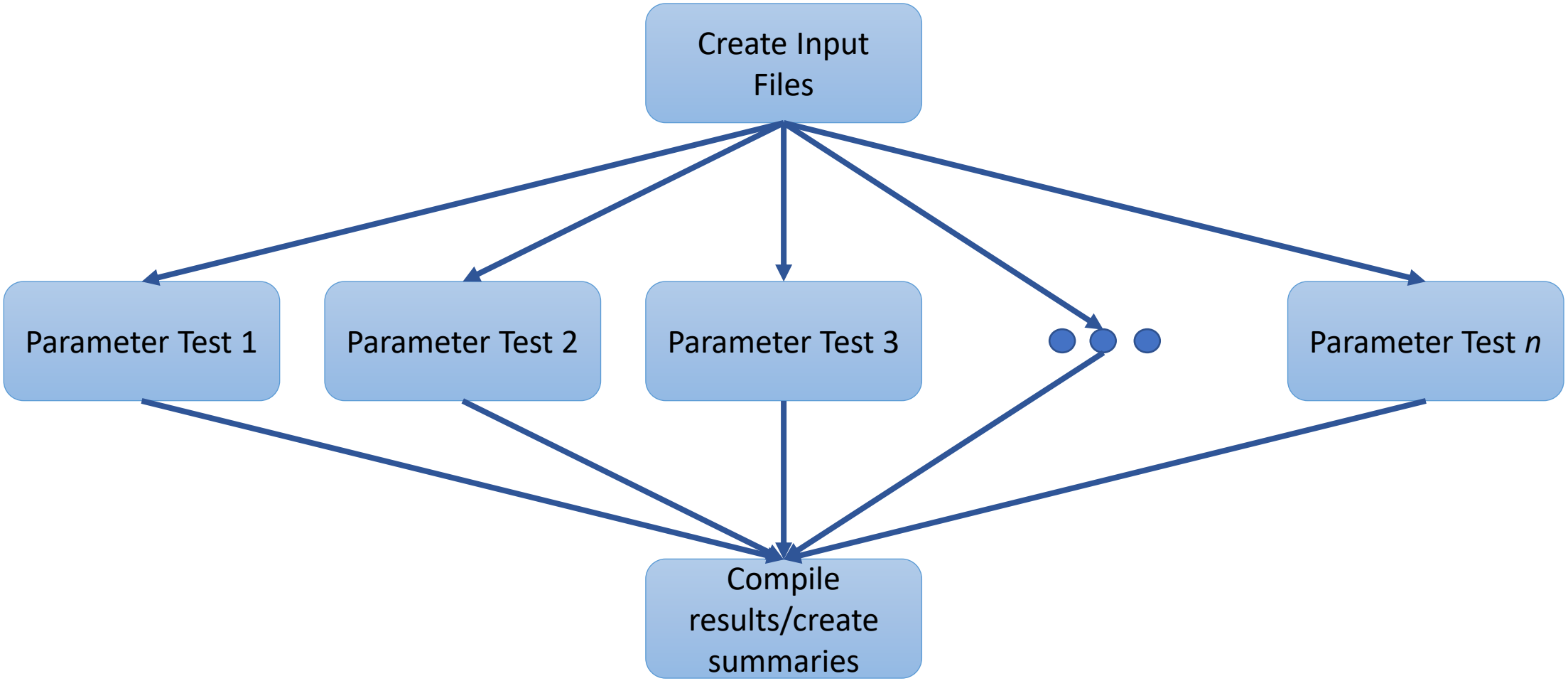
parameter_test.dag

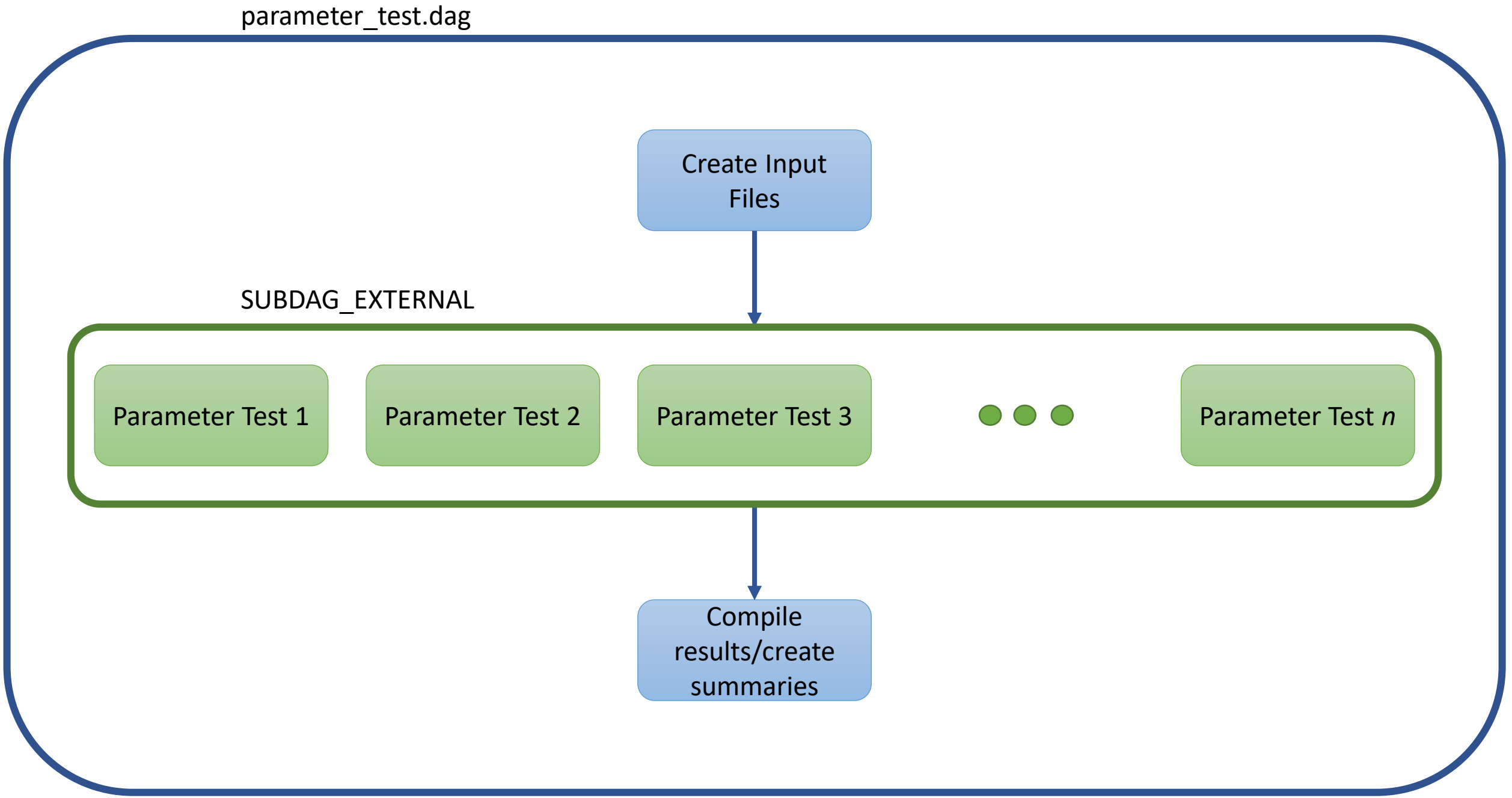


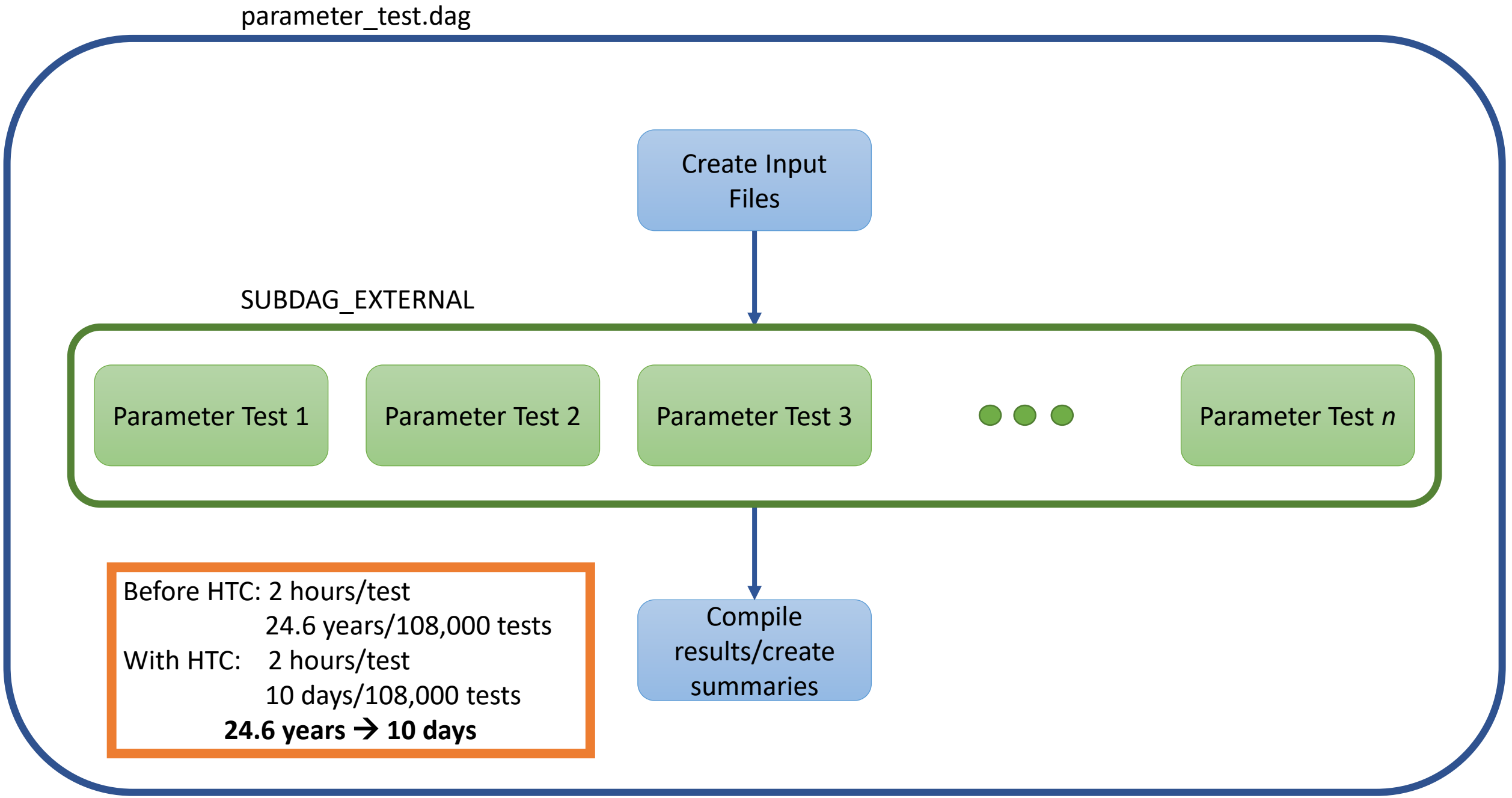
parameter_test.dag



parameter_test.dag




parameter_test.dag
Create Input Files
SUBDAG_EXTERNAL
Parameter Test 1
Parameter Test 2
Parameter Test 3
Parameter Test *n*
Compile results/create summaries

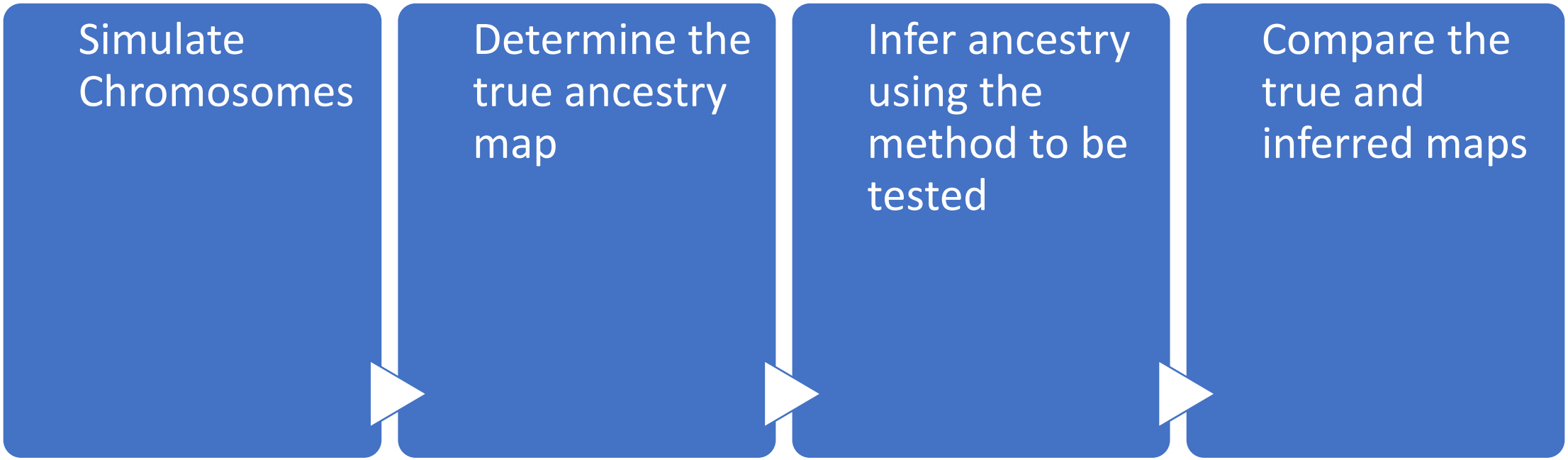

parameter_test.dag
Create Input Files
SUBDAG_EXTERNAL
Parameter Test 1
Parameter Test 2
Parameter Test 3
Parameter Test n
Compile results/create summaries
Before HTC: 2 hours/test
24.6 years/108,000 tests
With HTC: 2 hours/test
10 days/108,000 tests
24.6 years → 10 days

Testing with Simulated Chromosomes

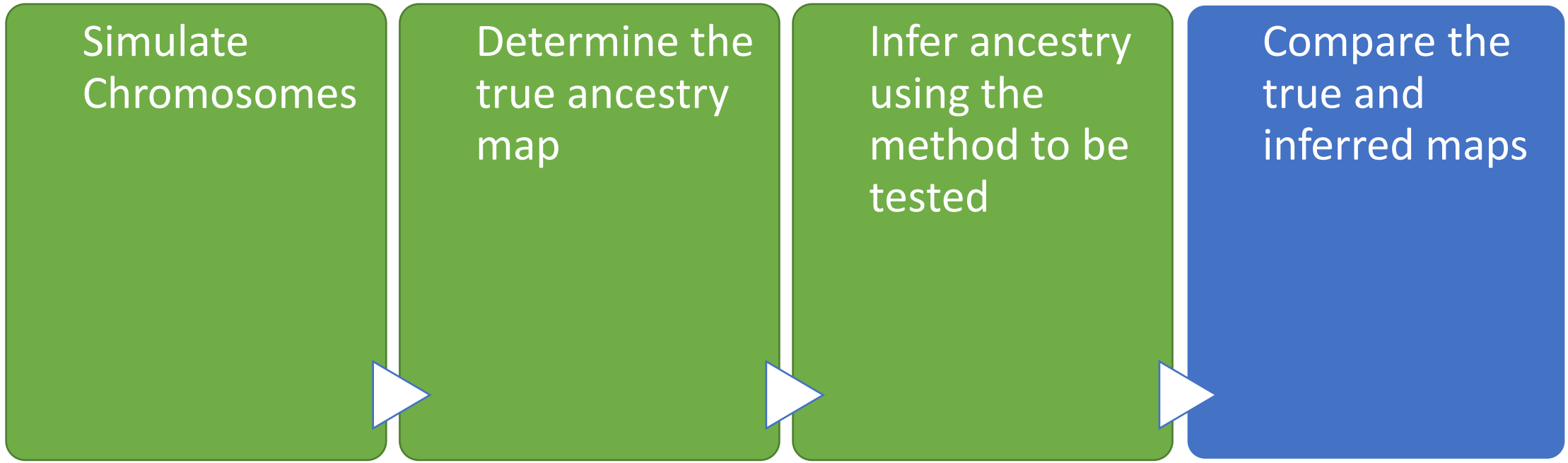
Testing with Simulated Chromosomes

- How well is the program performing?

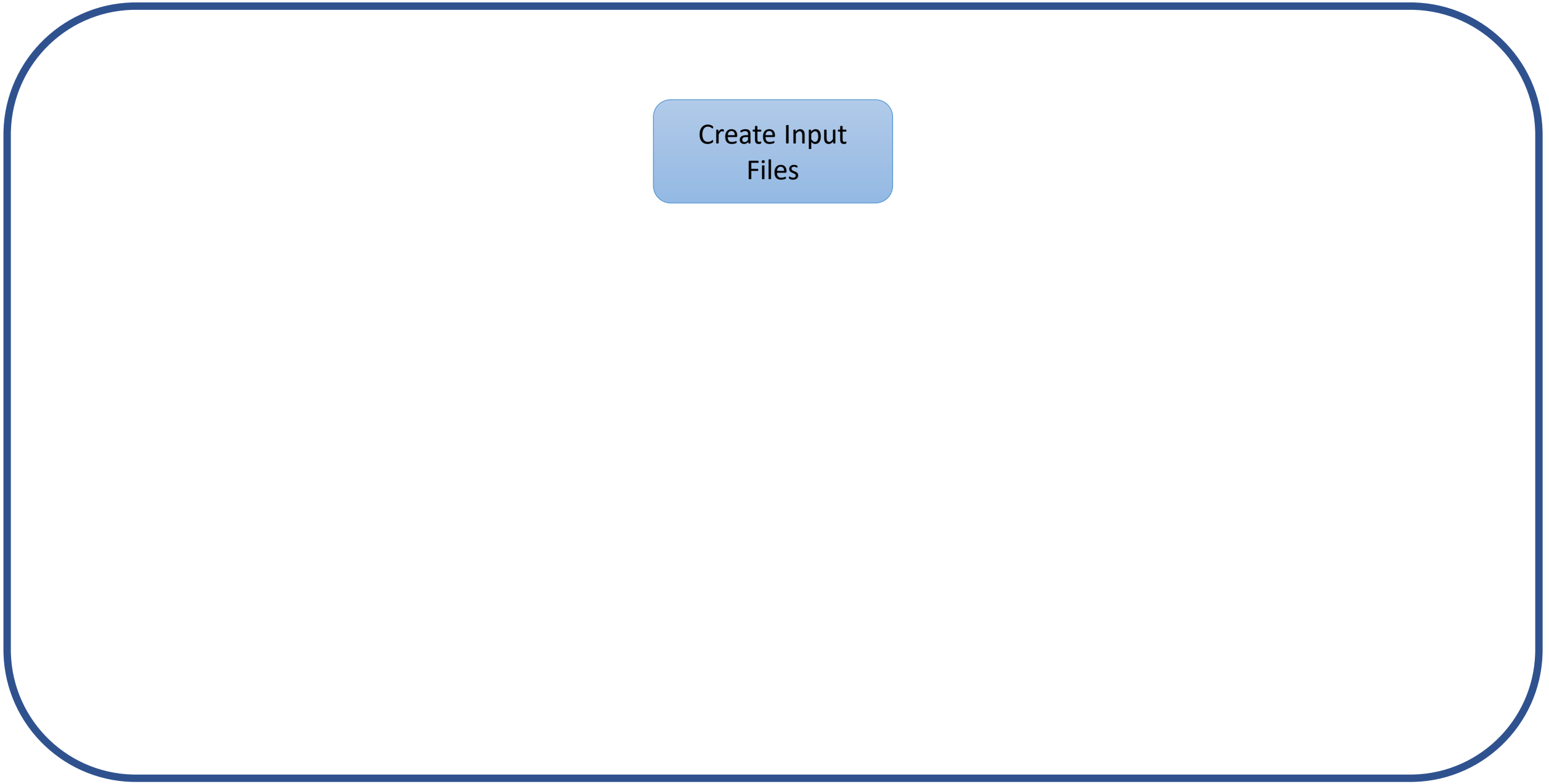
Testing with Simulated Chromosomes



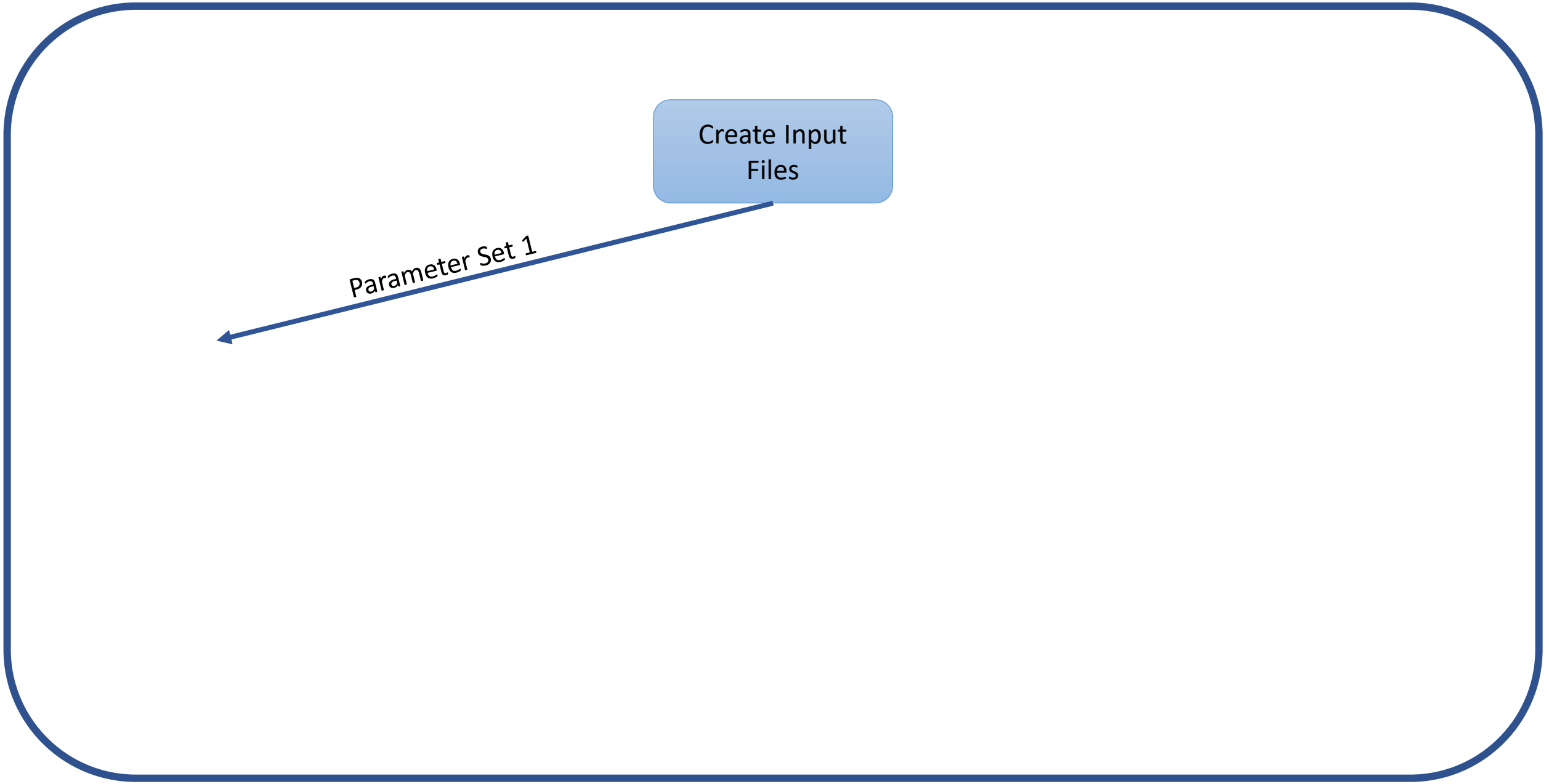
Testing with Simulated Chromosomes

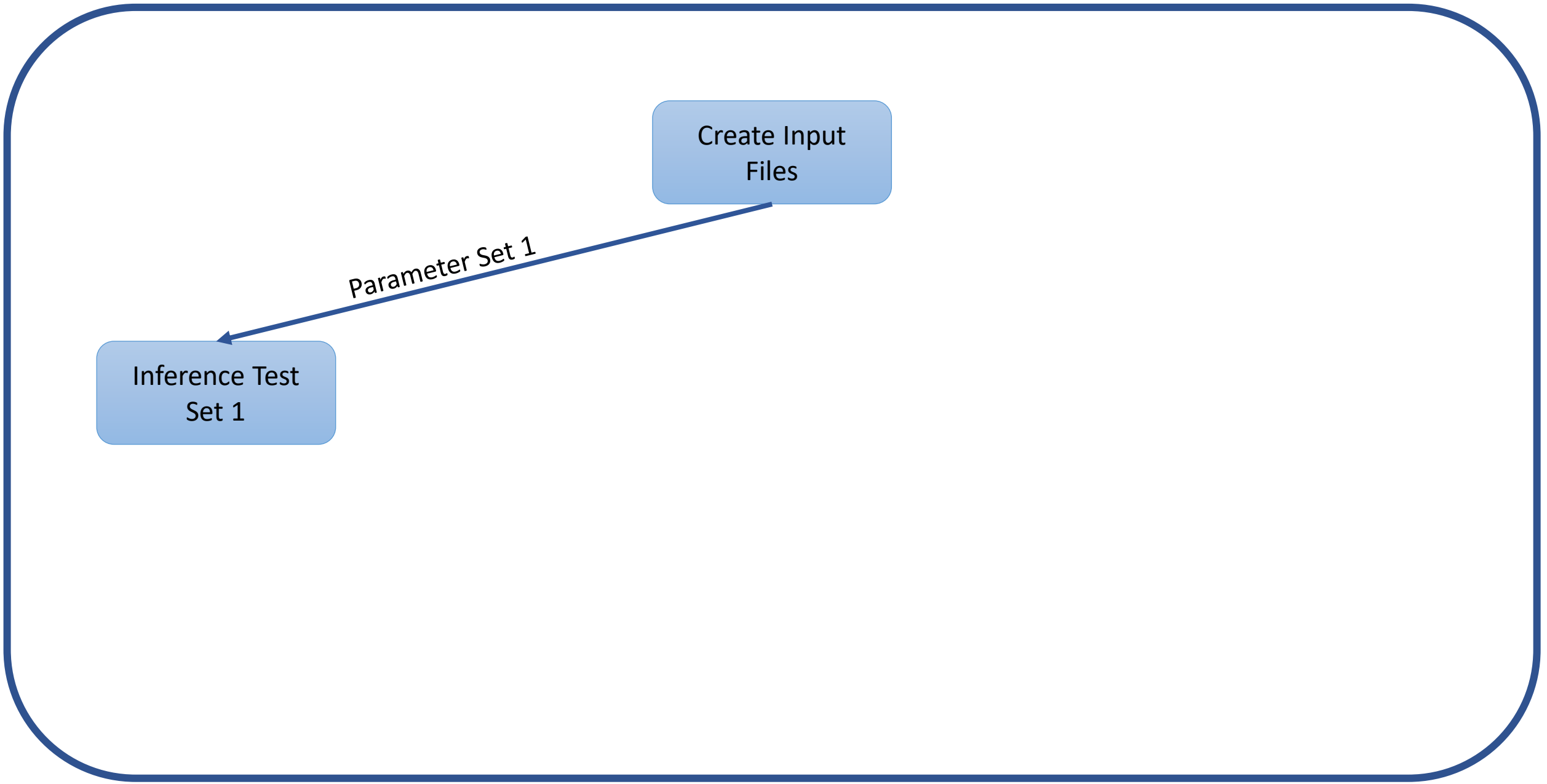


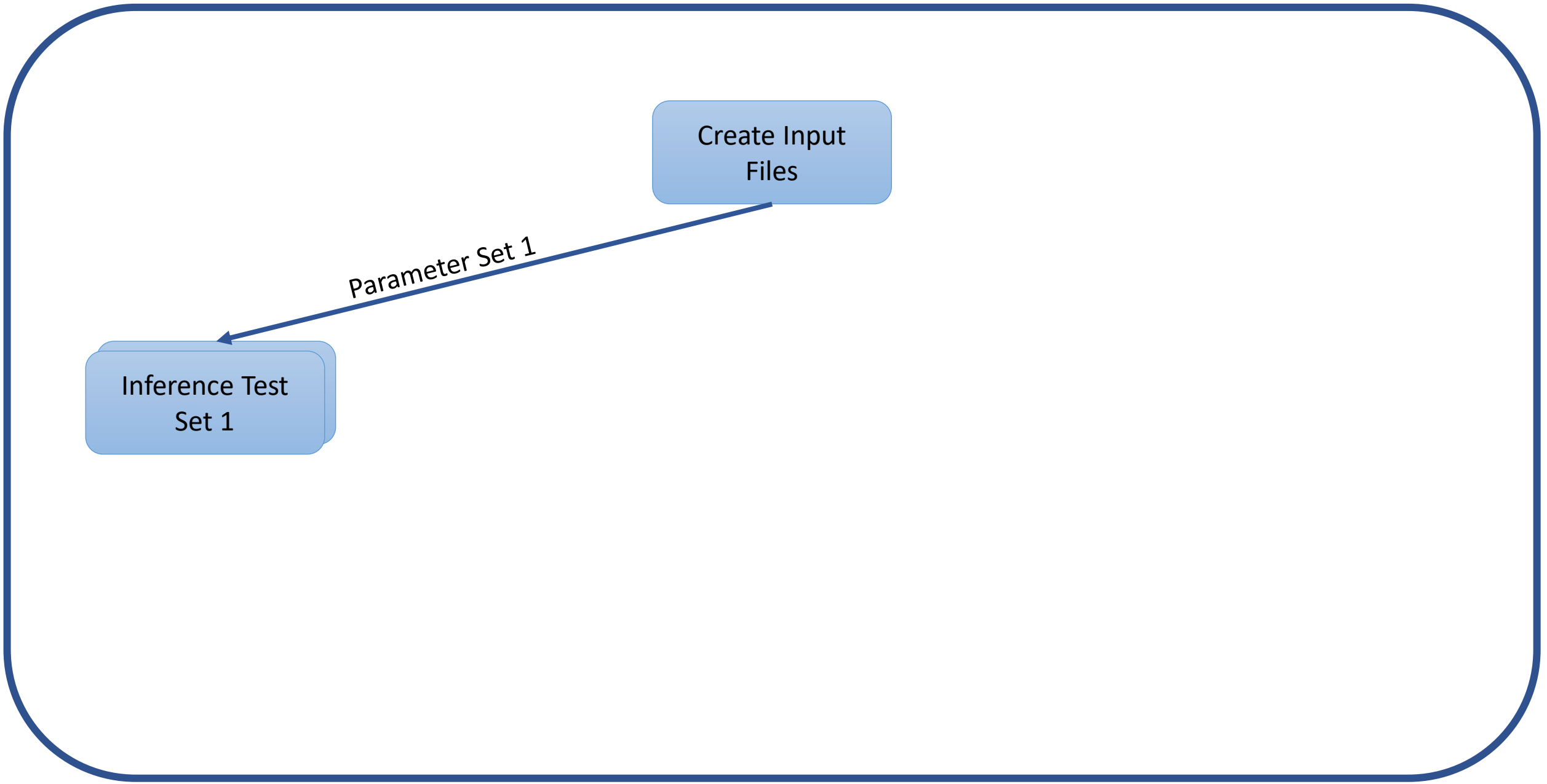
inference_testing.dag

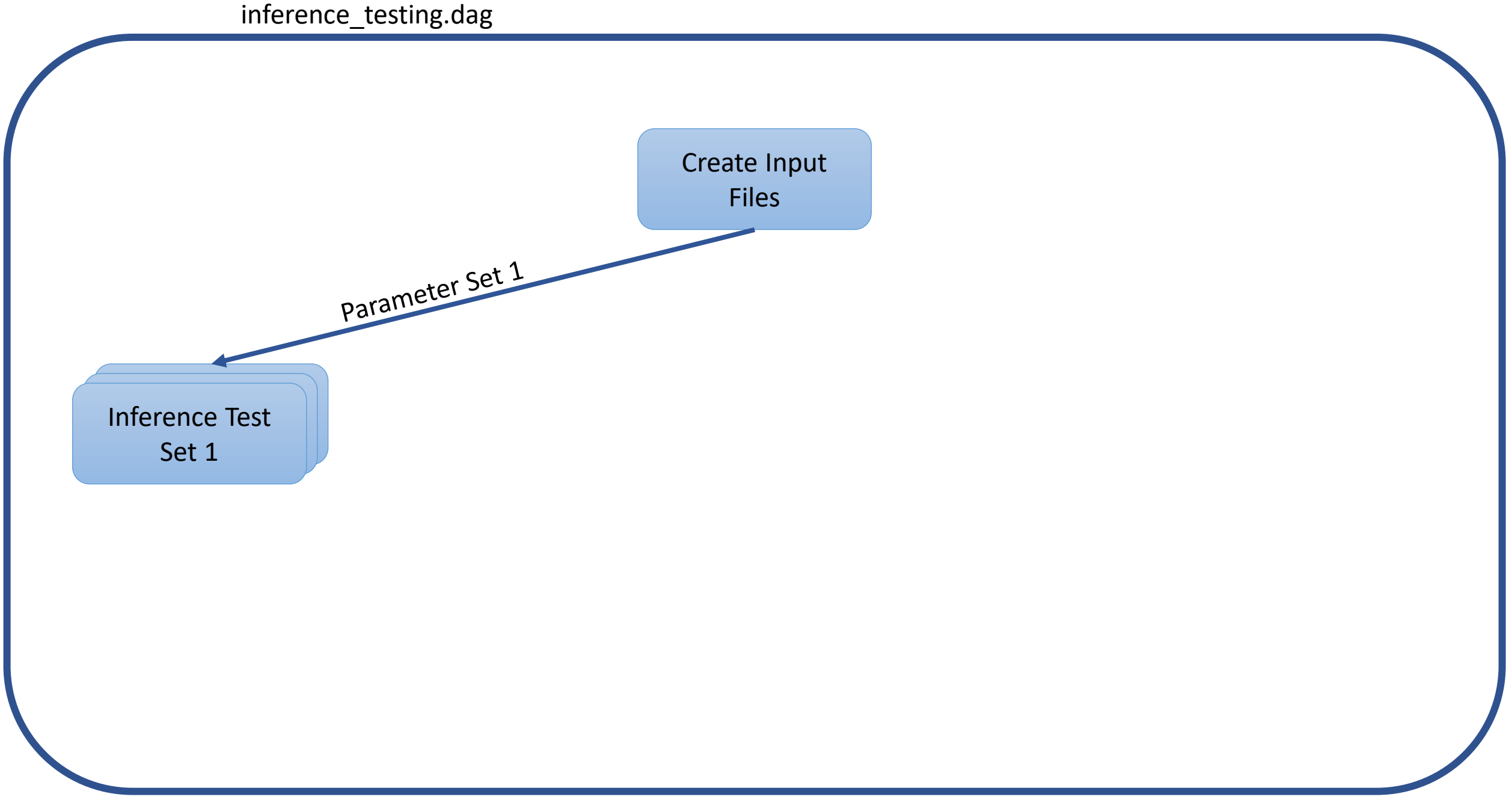


inference_testing.dag

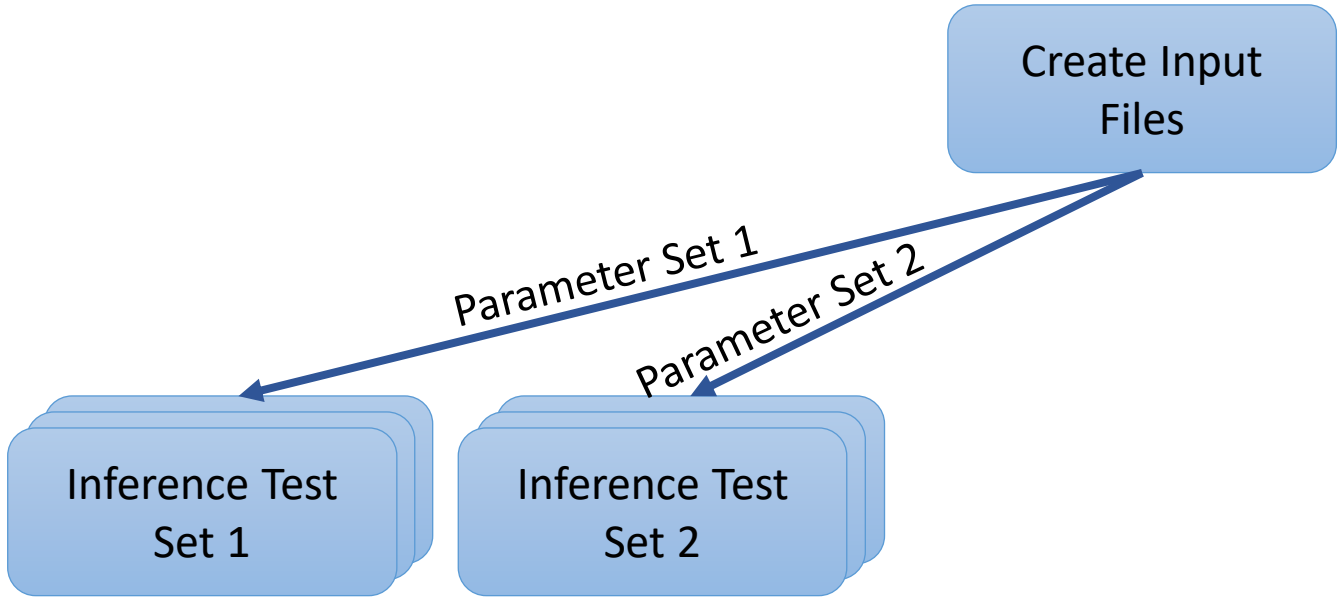



inference_testing.dag
Create Input Files
Parameter Set 1
Inference Test Set 1

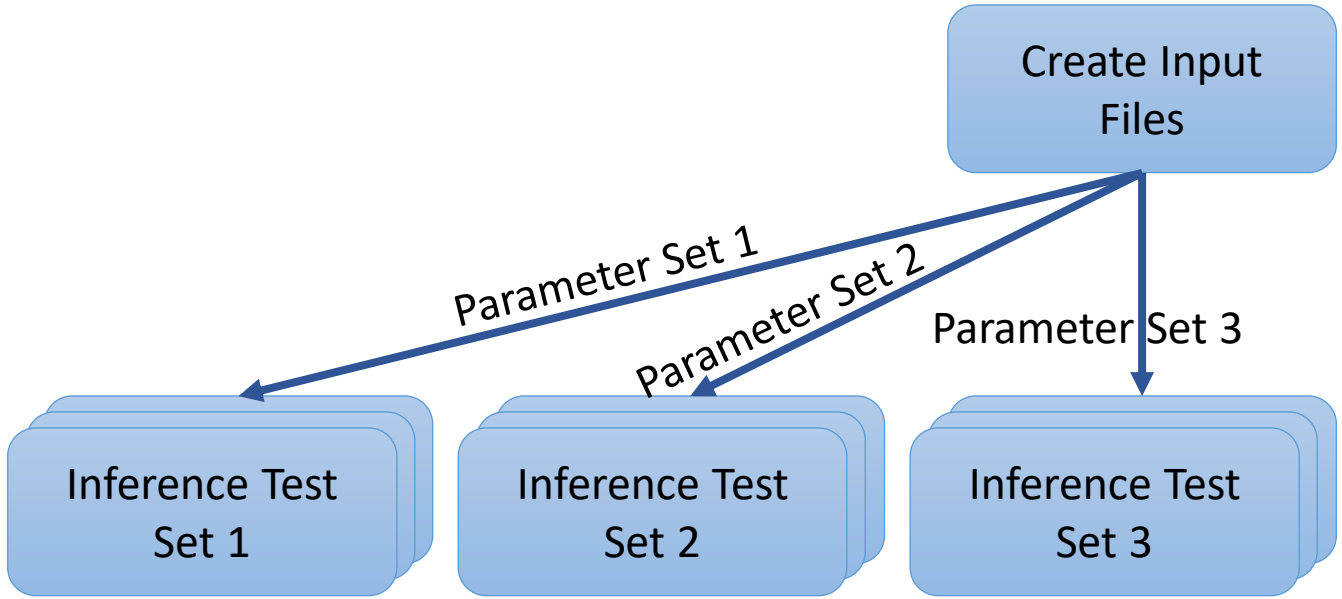

inference_testing.dag
Create Input Files
Parameter Set 1
Inference Test Set 1


inference_testing.dag
Create Input Files
Parameter Set 1
Inference Test Set 1

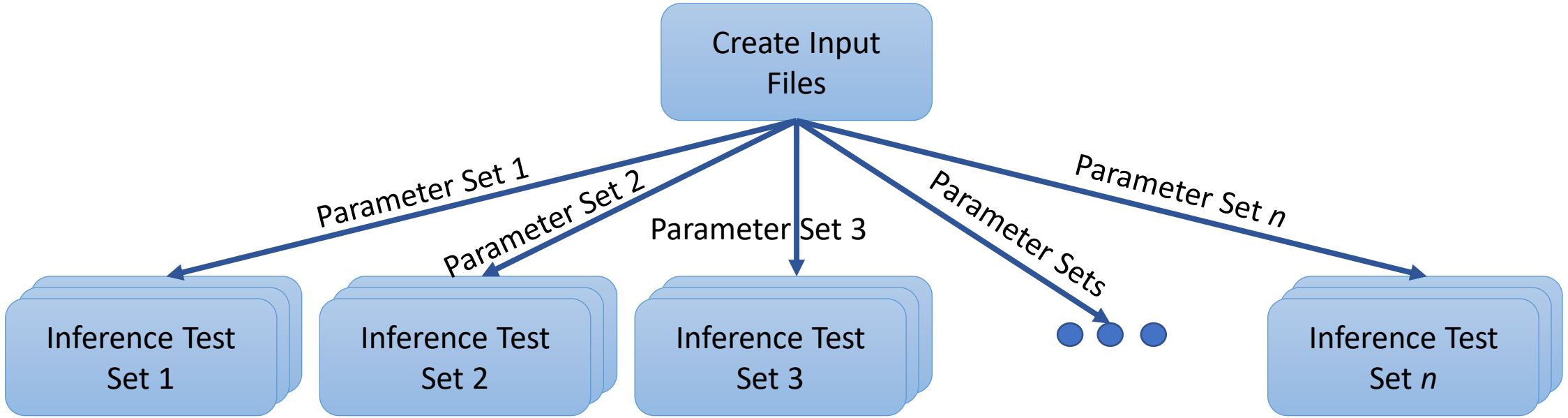
inference_testing.dag



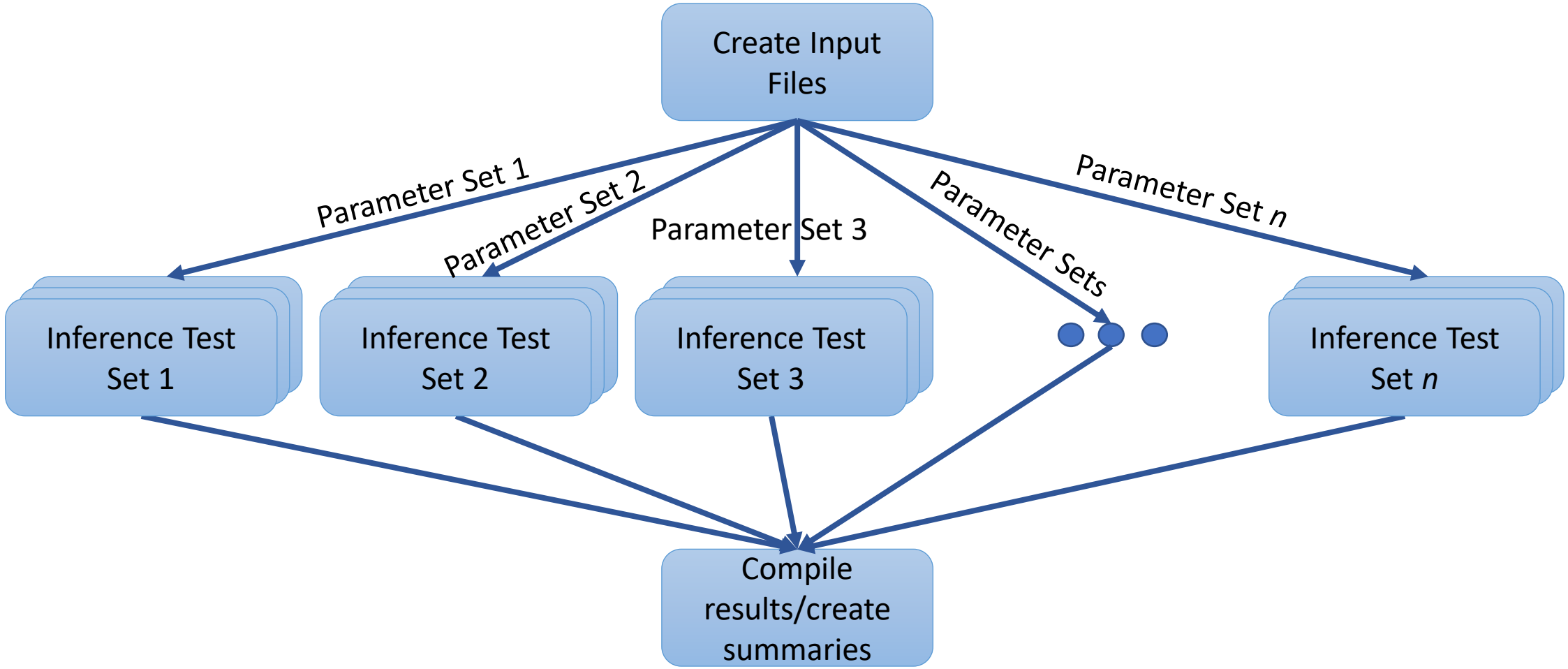
inference_testing.dag



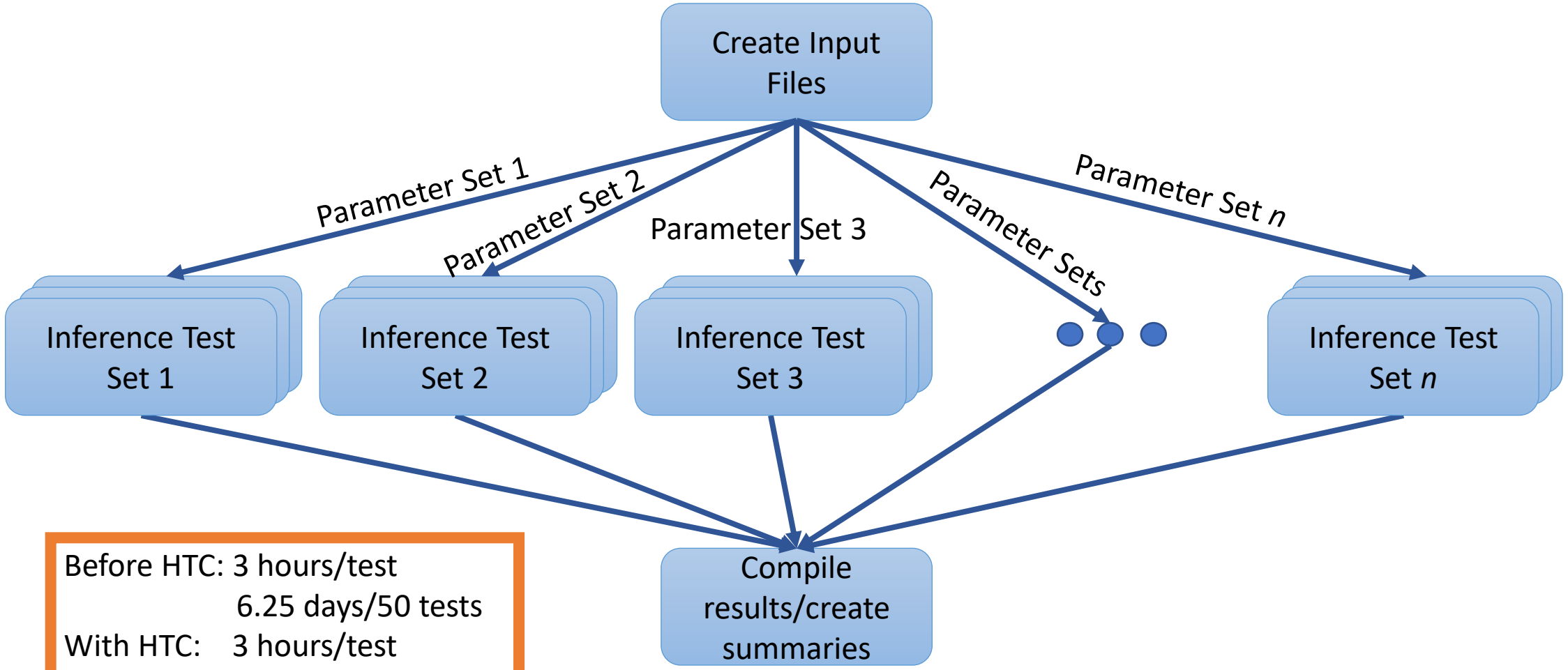
inference_testing.dag



inference_testing.dag



inference_testing.dag



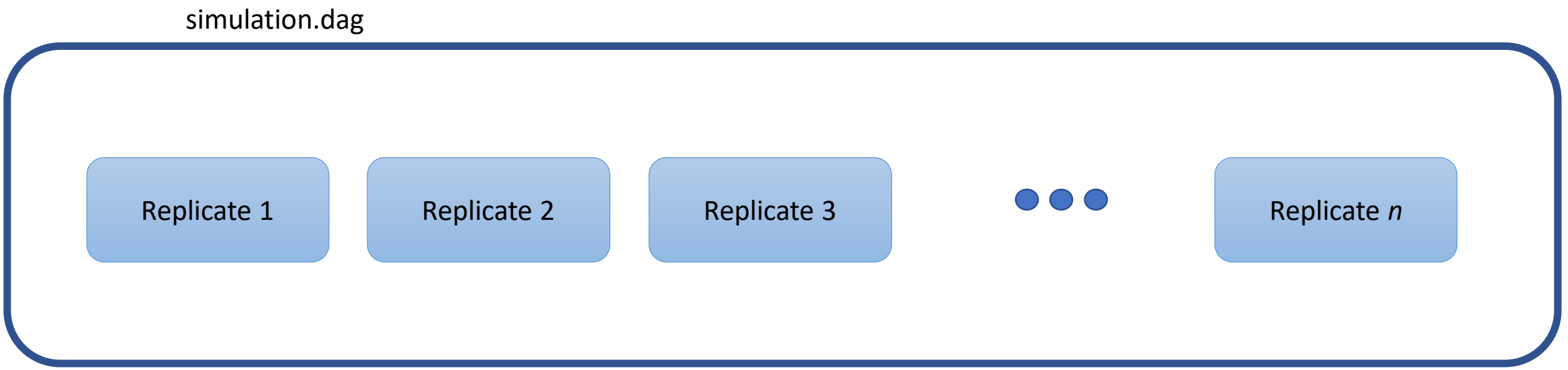
Before HTC: 3 hours/test
6.25 days/50 tests
With HTC: 3 hours/test
10 hours/50 tests
6.25 days → 10 hours

Simulations

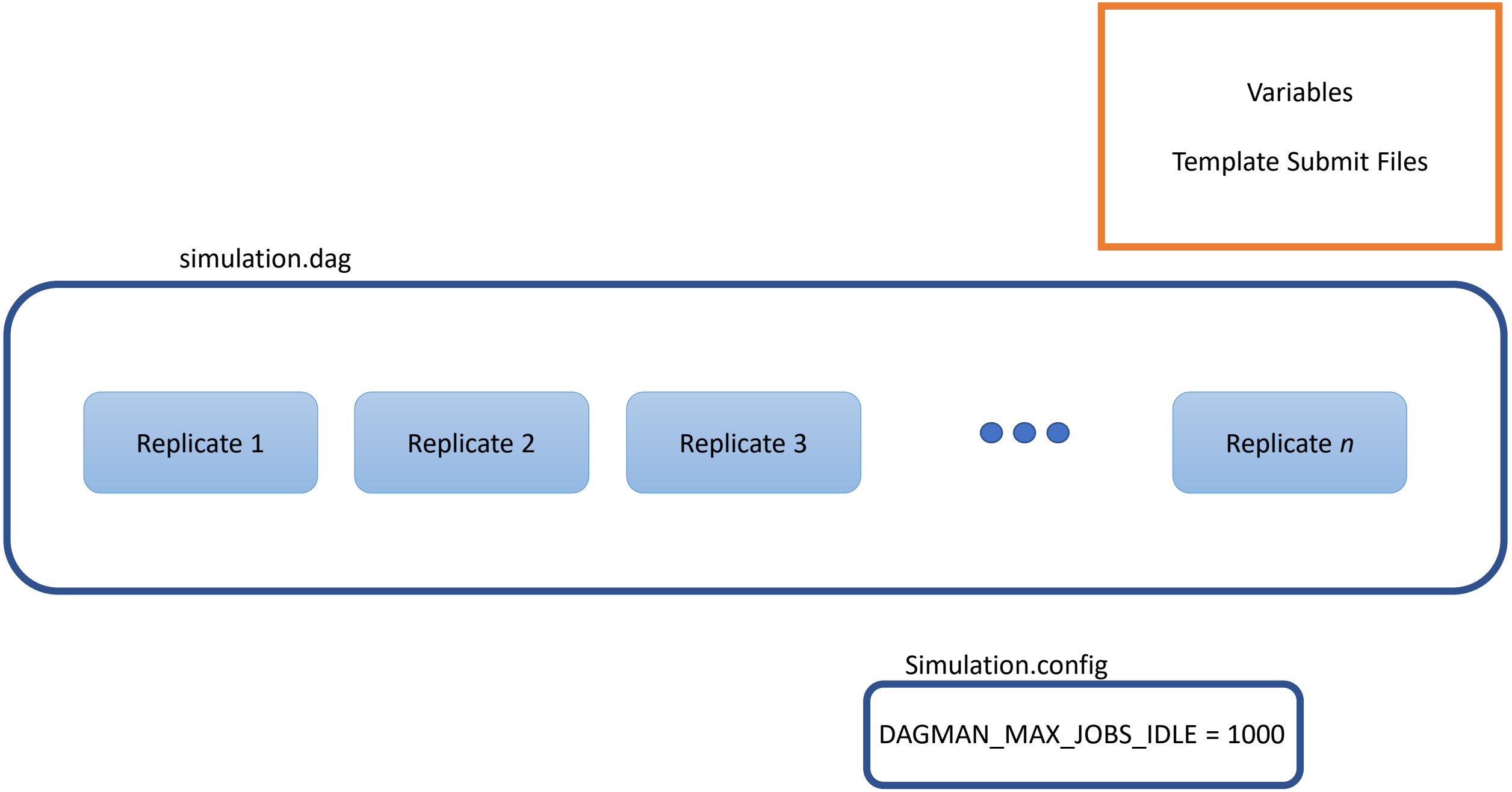
Simulations

Simulate data and run a script to make a summary

simulation.dag


simulation.dag
Replicate 1
Replicate 2
Replicate 3
Replicate *n*

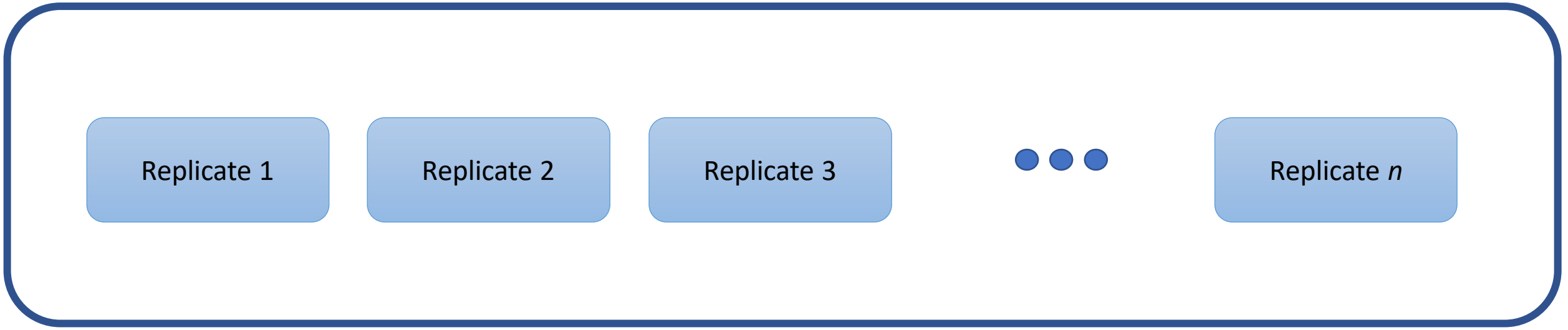
Variables
Template Submit Files
simulation.dag
Replicate 1
Replicate 2
Replicate 3
Replicate *n*


Variables
Template Submit Files
simulation.dag
Replicate 1
Replicate 2
Replicate 3
Replicate *n*
Simulation.config
DAGMAN_MAX_JOBS_IDLE = 1000

Variables

Template Submit Files

simulation.dag



Before HTC: 2 hours/test
2.7 years/12,000 tests
With HTC: 2 hours/test
30 hours/ 12,000 tests
2.7 years → 30 hours

Simulation.config

DAGMAN_MAX_JOBS_IDLE = 1000

Variables

Template Submit Files

simulation.dag

Replicate 1

Replicate 2

Replicate 3

Replicate *n*

Before HTC: 2 hours/test
2.7 years/12,000 tests
With HTC: 2 hours/test
30 hours/ 12,000 tests
2.7 years → 30 hours

Simulation.config

DAGMAN_MAX_JOBS_IDLE = 1000

Conclusion

Conclusion

- HTC can improve research in biological sciences

Conclusion

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

Conclusion

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

Conclusion

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

In the last year, I have used 8.5 million HTC 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 almost 1,000 years.