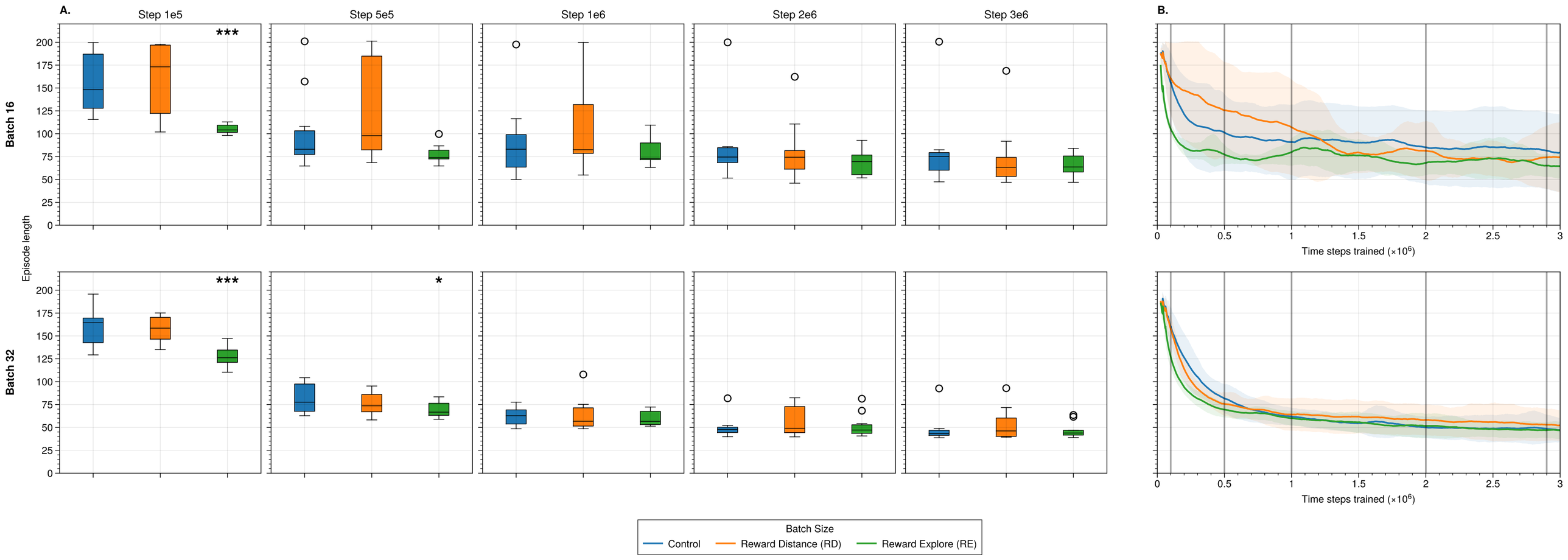
A.
Step 1e5
Step 5e5
Step 1e6
Step 2e6
Step 3e6
Batch 16
Batch 32
Episode length
***
*
B.
Time steps trained (×10^6)
Batch Size
Control
Reward Distance (RD)
Reward Explore (RE)