

De Novo Genome Assembly is the process of taking lots of short (error filled) DNA sequence reads and trying to construct the complete underlying genome. To get a better sense for this process, and what makes it difficult, let's try out a small example (without any errors).

1. Suppose you have the following list of sentence fragments. Can you construct the original statement? How did you do it?

on this contine
hers brought for
r score and sev
ears ago our father
ew nation, conc
ought forth, upon th
four sco
tinent, a new na
and seven years

2. Suppose you have the following list of sentence fragments. Can you construct the original sequence? How did you do it? What are pros and cons of the approach you tried?

u have to kno
ake! i ca
the toy ship a
p and a littl
k and the cak
i can hold t
nd a little toy m
e toy man!
e cup and t
how. i can ho
h on a rak
n hold up the cu
ks! and the fi
to know ho
it is fun to ha
fun but you ha
he cup and the mi
an hold up t
p these boo
un to have fu
e books! an
he fish on a
d the ca
the cake! i ca

3. Was question 1 or 2 more difficult? Why?

Source: Layla Oesper, Carleton College.