

#-----#

# SIMULATED GENOME ASSEMBLY

#-----#

1. Here are a bunch of subsequences, sampled with error, from some text.
2. Each copy given to the separate groups are all from the same text, but they may have different read lengths and error rates.
3. If you have "word,word" output (from paired ends), this represent the ends of a larger fragment from the text. That is, you have the additional information that each pair of words is some specific distance apart.
4. " \_ " indicates a space (" ") in the text.
5. Please use pen/paper or a computer notepad to figure out what the source sentences are.
6. No, you may not use Google.

#-----#

Some facts that may help:

1. I generated these data here: <http://lyorn.idyll.org/~t/assembly-exercise/index.cgi>
2. Parameters were as follows: Read length=10, mutation rate = 20, coverage = 5, sorting = no, paired ends = yes, insert size = 25

Some questions to answer as you go:

1. How long did it take to complete your assembly? Extrapolate this to a 3 Gb genome assembly using similar parameters.
2. Are there any reads that are over-represented? What does this indicate?
3. How would this effort differ if you were assembling actual DNA data?

ts\_into\_ha, l\_and\_ever  
lorious\_ea, at\_themsel  
mselves\_op, \_custom\_al  
o\_the\_dust, \_before\_th  
e\_fresh\_de, \_death\_the  
\_down\_to\_t, \_beginning  
rabbits\_in, \_everythin  
ter\_out\_of, and\_ashes\_  
back\_to\_se, \_seed\_flee  
r\_beginnin, \_in\_wester  
esh\_dearh\_, seed\_death  
\_one\_in\_an, hinese\_box  
wn\_to\_thei, r\_beginnin  
btrning\_wo, stant\_beau  
manders\_th, s\_the\_gree  
ath\_the\_gr, the\_dust\_a  
\_sweeten\_t, ir\_turn\_ir  
ginnings\_s, estern\_ski  
g\_wnuld\_on, n\_the\_inst  
\_seed\_flee, flee\_iearth  
ments\_toot, out\_of\_cha

everything, gccupming\_  
d\_this\_bur, s\_burwing\_  
chinese\_bo, to\_hats\_al  
\_the\_old\_y, een\_years\_  
n\_glorious, eadts\_moon  
ing\_one\_in, e\_in\_anodh  
the\_green\_, leip;  
ibered\_the, he\_wording  
and\_set\_in, glorious\_e  
thhthe\_gre, reen\_death  
he\_old\_yea, \_green\_yea  
like\_golde, salamander  
glorious\_e, ons\_eat\_th  
94;\_all\_ev, l\_everythi  
ves\_opposi, he\_custom\_  
\_yeafs\_mig, eap;\_  
\_the\_wordi, advertisem  
elves\_oppo, e\_custom\_a  
burning\_wo, woyled\_on\_t  
o\_hqts\_all, erything\_r  
down\_to\_th, innings\_su

sqed\_death, death\_the\_  
ike\_golden, salamander  
ts\_wnto\_ha, all\_and\_ev  
themselvps, elvps\_oppo  
burning\_wo, he\_instant  
s\_easts\_mo, eat\_themse  
tant\_beaut, eautifully  
\_outkof\_ch, chars\_and\_  
nd\_ashes\_o, f\_dust\_and  
ut\_of\_dust, like\_golde  
nd\_and\_thi, ng\_would\_o  
\_the\_hand\_, ning\_would  
et\_in\_glor, orioas\_eas  
flee\_dgath, h\_dust\_and  
e\_itself\_e, els\_rememb  
\_it\_the\_me, \_merest\_to  
reverse\_it, elf\_eckels  
\_befoxe\_th, eginning\_a  
f\_chars\_an, es\_out\_of\_  
d\_tha\_word, ing\_in\_the  
ight\_do\_it, t\_the\_mere

mselves\_tz, the\_custom  
d\_the\_word, ording\_in\_  
sementslto, out\_of\_cha  
fly\_backft, seed\_flee\_  
gloriquus\_e, qus\_easts\_  
ars\_the\_gr, s\_might\_le  
ars\_migpt\_, 4;\_roses\_s  
nd\_mightzd, htzdo\_it\_t  
itself\_eck, nthe\_wordi  
the\_ond\_ye, \_yeari\_the  
other\_like, \_like\_chin  
in\_western, es\_and\_sec  
\_reverse\_i, f\_eckels\_r  
eten\_the\_p, \_turn\_iris  
ush\_down\_t, beginnings  
ght\_leap&#, 8u4;\_roses  
ly\_reverse, self\_eckel  
ther\_like\_, chinese\_bo  
and\_everyt, ing\_to\_the  
ke\_chinese, nese\_boxts

e\_air\_whit, rishblack\_  
d\_coalsfti, me\_before\_  
h\_down\_to\_, nings\_suns  
he\_green\_d, death\_to\_t  
o\_the\_lett, \_and\_ashes  
to\_hats\_al, \_andzevery  
innings\_su, \_suns\_rise  
t\_do\_it\_th, ch\_of\_a\_ha  
\_skies\_and, s\_easts\_mo  
tt\_r\_out\_o, of\_chars\_a  
e\_oldyyear, the\_green\_  
h\_of\_the\_h, hand\_and\_t  
\_boxes\_rab, \_all\_and\_e  
g\_in\_the\_a, \_to\_the\_le  
o\_the\_lett, rs\_and\_ash  
amd\_coals\_, als\_rush\_d  
t\_and\_coal, als\_like\_g  
g\_a\_touch\_, ight\_do\_it  
glorious\_e, us\_easts\_m  
osite\_to\_t, and\_everyt

fore\_the\_b, ewinning\_a  
k\_to\_seed\_, eed\_slee\_d  
wn\_to\_thyi, thyr\_begi  
e\_air\_whit, hite\_hair\_  
hblack\_wri, 94;\_all\_ev  
bucning\_wo, on\_the\_ins  
eeten\_the\_, air\_turn\_i  
e\_seed\_der, the\_green\_  
\_eat\_thems, selvey\_opp  
ng\_cupping, g\_one\_in\_a  
rush\_dowd\_, r\_beginnin  
om\_all\_and, upping\_one  
g\_fly\_back, back\_to\_se  
nd\_ashes\_o, akd\_coals\_  
would\_on\_t, tant\_beaut  
tisements\_, ut\_of\_char  
eap;\_, weeten\_the  
ies\_and\_se, ous\_easts\_  
eath\_the\_s, etth\_the\_g  
h\_the\_gree, \_green\_dea

d\_coals\_ru, ush\_down\_t  
rything\_re, g\_returnin  
ered\_the\_w, the\_wordin  
\_hand\_migh, \_it\_the\_we  
t\_and\_coal, ush\_dowq\_t  
n\_irishbla, nkles\_vani  
cupping\_on, g\_one\_in\_a  
osite\_to\_t, \_and\_every  
s\_itself\_, ered\_the\_w  
\_a\_touch\_o, \_hand\_migh  
\_the\_green, th\_to\_the\_  
erse\_itsel, elf\_eckels  
ons\_eat\_th, t\_themielv  
ed\_the\_wor, \_wording\_i  
eath\_dustn, \_rush\_down  
ginning\_a\_, \_might\_do\_  
the\_beginn, h\_of\_a\_han  
lden\_salam, \_old\_years  
agd\_set\_in, sts\_moons\_  
nd\_this\_bu, urning\_wou  
\_the\_green, \_to\_the\_du

lack\_wrink, h;\_al  
nt\_beautif, lly\_revers  
e\_custom\_a, tom\_all\_an  
rn\_mkiej\_s, ej\_snd\_set  
d\_the\_word, ng\_in\_thl\_  
to\_seed\_fl, ath\_dust\_a  
\_the\_hand\_, nd\_this\_bu  
d\_the\_word, dvertireme  
like\_gozde, s\_the\_old\_  
alamanders, s\_the\_old\_  
th\_dust\_an, ush\_down\_t  
\_in\_anothe, e\_boxes\_ra  
94;\_all\_ev, verythingd  
\_turv\_iris, \_irishblac  
\_rabbits\_i, ll\_and\_evn  
rishblack\_, s\_vanish&#  
ir\_beginni, \_rise\_in\_w  
d\_set\_in\_g, \_glorious\_  
wn\_to\_thei, ginnang\_\_s  
ts\_into\_hk, kts\_all\_an  
air\_turn\_i, \_irishblac

ack\_wrinkl, nkles\_vani  
\_the\_lust\_, lust\_and\_c  
ing\_return, ng\_to\_the\_  
ee\_death\_d, r\_coals\_ru  
nt\_beautif, itself\_eck  
might\_leap, yeten\_the\_  
fly\_back\_t, s\_seed\_fle  
h\_death\_th, \_the\_seed\_  
894;\_roses, eeten\_the\_  
urning\_wou, \_would\_on\_  
\_tust\_and\_, re\_the\_beg  
en\_years\_m, 94;\_roses\_  
itv\_to\_tqe, m\_all\_and\_  
turying\_to, the\_fresh\_  
nd\_everyth, ng\_returni  
g\_to\_the\_f, the\_peen\_d  
wording\_in, advertisem  
\_white\_hai, spblack\_wr  
inese\_boxe, \_into\_hats  
thing\_retu, ning\_to\_th  
erything\_f, o\_seed\_fle

\_coals\_lik, ike\_golden  
ld\_years\_t, e\_green\_ye  
mbered\_the, the\_wordin  
s\_the\_old\_, \_green\_yea  
fully\_reve, \_itself\_ec  
chinese\_bo, to\_hats\_al  
lly\_revers, ckels\_reme  
ats\_all\_av, everdthing  
the\_advert, tisements\_  
st\_and\_coa, oals\_like\_  
mbered\_tae, \_wording\_i  
green\_year, leap;  
unn\_rise\_i, es\_and\_set  
ch\_of\_a\_ha, ght\_do\_it\_  
all\_and\_ev, rning\_to\_t  
in\_anotser, e\_chinese\_  
\_custom\_al, evything\_c  
ach\_ofvthe, rand\_and\_t  
re\_the\_beg, ng\_a\_touch  
to\_the\_fre, eath\_the\_s

een\_yewrs\_, ap;\_r  
nother\_lik, \_chsnese\_b  
death\_the\_, \_the\_dust\_  
e\_green\_ye, rs\_miggt\_l  
white\_haib, irishblack  
thing\_fly\_, ed\_flee\_de  
in\_western, t\_in\_glori  
easts\_moon, ves\_opposi  
\_the\_meres, merest\_tou  
turning\_to, ath\_the\_se  
\_years\_the, years\_migh  
r\_beginnin, ns\_rise\_in  
the\_air\_wh, rn\_irishbl  
to\_the\_let, t\_of\_chars  
ard\_and\_as, \_dust\_and\_  
ten\_the\_ai, e\_air\_what  
er\_like\_ch, boxes\_rabb  
st\_and\_coa, time\_befor  
\_seed\_deat, \_green\_dea  
\_and\_every, g\_to\_the\_f

\_reverse\_i, se\_itself\_  
ll\_and\_eve, cupping\_on  
and\_everyt, \_to\_the\_fr  
\_coals\_lik, nders\_the\_  
894;\_roses, roses\_swee  
\_fresh\_dea, \_death\_the  
erything\_c, ing\_one\_in  
the\_beginn, ing\_a\_touc  
to\_seed\_fl, \_death\_dxs  
\_kt\_khe\_me, est\_aouch\_  
chinese\_bo, rabbitn\_in  
oals\_tihe\_, e\_before\_t  
\_offchars\_, ars\_and\_as  
fresh\_deat, ath\_the\_se  
and\_this\_b, rning\_woul  
iy\_western, t\_in\_glori  
tter\_out\_o, hes\_out\_of  
one\_in\_ano, other\_like  
\_death\_dus, sh\_downtto  
vust\_and\_c, \_golden\_sa

ng\_fly\_bac, o\_seed\_fle  
ir\_white\_h, \_turn\_iris  
\_rush\_down, their\_begi  
\_death\_the, he\_green\_d  
rinkles\_va, 894;\_all\_e  
he\_letter\_, nd\_ashes\_o  
les\_vanish, ;\_all\_ever  
ements\_to\_, out\_of\_cha  
ike\_golden, n\_salamand  
ust\_and\_co, golden\_sal  
eforc\_the\_, ning\_a\_tou  
the\_wordin, in\_the\_adv  
ten\_the\_ai, air\_white\_  
\_salamande, he\_old\_yea  
d\_coals\_ru, r\_beginnin  
l\_and\_ever, rything\_cu  
g\_fly\_back, back\_to\_se  
kles\_vanis, g894;\_afl\_  
upping\_one, ki\_chinese  
\_touch\_of\_, h\_of\_a\_han  
e\_advertis, sements\_to

h\_of\_a\_han, and\_might\_  
e\_green\_de, ss\_and\_coa  
\_beginning, touch\_of\_a  
ee\_death\_d, \_rush\_down  
te\_hair\_tu, shblack\_wr  
innings\_su, \_in\_wester  
ars\_and\_as, \_dust\_and\_  
st\_and\_coa, tkme\_befor  
ht\_leap, 94;\_roses\_  
golden\_sal, he\_old\_yea  
ng\_a\_touch, \_might\_do\_  
e\_the\_begi, \_touch\_of\_  
es\_sweeten, te\_hair\_tu  
rzing\_to\_t, esh\_death\_  
94;\_all\_ev, hing\_fly\_b  
\_before\_th, beginningx  
\_skies\_and, n\_glorious  
verything\_, ing\_one\_in  
death\_the\_, o\_the\_dust  
ee\_death\_d, rush\_down\_  
s\_the\_gree, rs\_might\_l

f\_the\_hand, and\_this\_b  
advertisem, o\_the\_lett  
eir\_beginn, \_in\_wester  
alamanders, ers\_the\_ol  
m\_all\_and\_, rything\_cu  
ginning\_a\_, and\_might\_  
all\_and\_ev, returning\_  
everything, pping\_one\_  
hemselves\_, site\_to\_th  
n\_salamand, ld\_years\_t  
\_advertise, to\_the\_let  
\_hats\_all\_, \_everythin  
ers\_the\_ol, \_the\_grxen  
ed\_death\_t, h\_the\_gree  
uzn\_irishb, \_wrinkles\_  
oons\_eat\_t, opposite\_t  
returning\_, he\_fresh\_d  
e\_gmlden\_s, rs\_the\_old  
n\_skies\_an, set\_in\_glo  
ats\_all\_an, l\_and\_ever  
ning\_would, d\_on\_the\_i

bered\_the\_, ing\_in\_the  
ng\_to\_the\_, dyath\_the\_  
e\_green\_de, st\_and\_coa  
;\_ros, exten\_the\_  
eckels\_rem, e\_wording\_  
sinning\_s, n\_western\_  
\_the\_adver, e\_letter\_o  
g\_revurnin, o\_the\_fres  
fully\_reve, self\_eckel  
s\_remember, rding\_in\_t  
nnsalamand, \_years\_the  
n\_another\_, ike\_chines  
bits\_into\_, \_and\_every  
rs\_and\_ash, ut\_of\_dust  
the\_green\_, een\_death\_  
the\_green\_, ears\_might  
ght\_leap&#, eeten\_the\_  
her\_like\_c, e\_chinese\_  
hats\_akl\_a, nd\_everyth  
\_the\_old\_y, en\_years\_m

ts\_all\_and, ing\_return  
st\_and\_coa, als\_time\_b  
;\_roses\_sw, \_sweeten\_t  
s\_rush\_dow, eginnings\_  
in\_another, r\_like\_chi  
like\_golde, \_the\_old\_y  
ears\_might, ghtileapd#  
ght\_lefp&#, p;\_ro  
t\_of\_dust\_, awd\_coals\_  
oses\_sweet, he\_air\_lhi  
\_and\_every, returning\_  
\_like\_gold, den\_salama  
ck\_to\_seed, eed\_flee\_d  
years\_migh, might\_loap  
ant\_beauti, itself\_eck  
s\_vanish&#, ish;\_  
hing\_retur, \_to\_the\_fr  
g\_one\_in\_a, \_chinese\_b  
\_death\_the, \_to\_the\_du  
\_fly\_back\_, ee\_death\_d

before\_the, inning\_a\_t  
hgloriojs\_, eat\_themse  
alaganders, rs\_the\_gre  
ath\_to\_the, and\_coals\_  
\_custom\_al, \_and\_every  
r\_turn\_iri, ishblackrw  
ern\_skies\_, \_in\_glorio  
rything\_re, \_to\_the\_fr  
he\_custam\_, ill\_and\_ev  
autifully\_, ully\_rever  
green\_dean, \_deanh\_to\_  
m\_all\_and\_, ng\_cupping  
\_the\_wordi, e\_advertis  
s\_burningm, ld\_on\_the\_  
ht\_leap, \_sweeten\_t  
oals\_tim\_, fore\_the\_b  
hing\_retur, esh\_death\_  
tself\_ecke, ed\_thj\_wor  
\_an\_coals, lamanders\_  
other\_lime, \_boxes\_rab

\_of\_chars\_, d\_ashea\_ou  
hand\_and\_t, is\_burning  
coals\_like, nders\_the\_  
\_of\_dust\_a, t\_anp\_coal  
stom\_all\_t, hingccuppi  
els\_rememb, rding\_in\_t  
t\_of\_chars, shes\_out\_o  
o\_the\_dust, st\_and\_coa

e\_boxes\_ra, \_hatszall\_  
e\_seed\_dea, the\_green\_  
the\_letter, s\_and\_ashe  
hars\_and\_a, ut\_of\_dust  
set\_in\_glo, iousbeasts  
\_death\_to\_, als\_time\_b  
\_a\_hand\_mi, t\_do\_it\_th  
se\_boxes\_r, nto\_hats\_a

oses\_sweet, air\_white\_  
ts\_into\_ha, \_hats\_all\_  
oals\_rush\_, sh\_down\_to  
ust\_and\_co, ike\_golden  
d\_years\_th, ars\_might\_  
seed\_death, h\_the\_gree  
death\_the\_, \_green\_dea  
94;\_roses\_, r\_white\_ha

#-----#

# SIMULATED GENOME ASSEMBLY

#-----#

1. Here are a bunch of subsequences, sampled with error, from some text.
2. Each copy given to the separate groups are all from the same text, but they may have different read lengths and error rates.
3. If you have "word,word" output (from paired ends), this represent the ends of a larger fragment from the text. That is, you have the additional information that each pair of words is some specific distance apart.
4. "\_" indicates a space (" ") in the text.
5. Please use pen/paper or a computer notepad to figure out what the source sentences are.
6. No, you may not use Google.

#-----#

Some facts that may help:

1. I generated these data here: <http://lyorn.idyll.org/~t/assembly-exercise/index.cgi>
2. Parameters were as follows: Read length=20, mutation rate = 20, coverage = 5, sorting = no, paired ends = yes, insert size = 25

Some questions to answer as you go:

1. How long did it take to complete your assembly? Extrapolate this to a 3 Gb genome assembly using similar parameters.
2. Are there any reads that are over-represented? What does this indicate?
3. How would this effort differ if you were assembling actual DNA data?

mselves\_opposite\_to\_, \_custom\_all\_and\_ever  
eed\_death\_the\_green\_, h\_the\_green\_death\_to  
mxelv, mxelv  
ginnings\_suns\_rise\_i, western\_hkies\_and\_se  
custom\_all\_and\_every, verything\_cupping\_on  
th\_dust\_and\_coals\_ru, \_and\_coals\_rush\_down  
ats\_all\_and\_everythi, returning\_to\_the\_fre  
\_all\_and\_everythk, \_all\_and\_everythk  
nese\_boxes\_rabbits\_i, nese\_boxes\_rabbits\_i  
e\_to\_the\_custom\_all\_, the\_custom\_all\_and\_e  
\_and\_coals\_like\_gols, ke\_golsen\_salamander  
s\_the\_old\_years\_the\_, e\_old\_years\_the\_gree  
ginning\_a\_touch\_of\_h, h\_hand\_might\_do\_it\_t  
thinw\_cupping\_o, thinw\_cupping\_o  
\_everything\_re, \_everything\_re  
ir\_white\_hair\_turj\_i, hite\_hair\_turj\_irish  
st\_an, st\_an  
ders\_the, ders\_the  
eath\_the\_green\_death, h\_to\_the\_dust\_and\_co  
\_hair\_turn\_irishblac,

en\_the\_air\_white\_, en\_the\_air\_white\_  
rs\_and\_ashes\_out\_ox\_, rs\_and\_ashes\_out\_ox\_  
\_advertisements\_to, \_advertisements\_to  
etkin\_glorious\_easts, tsemoons\_eat\_themsel  
elves\_opposite\_to\_th, e\_custom\_all\_axd\_eve  
returning\_to\_the\_fr, returning\_to\_the\_fr  
ir\_begpnnin, ir\_begpnnin  
the\_ndvertise, the\_ndvertise  
f\_a\_hand\_might\_do\_it, \_the\_merest\_touch\_of  
e\_instant\_beautifull, beautifullr\_reverse\_  
sweeten\_the\_aiu\_w, sweeten\_the\_aiu\_w  
he\_seed\_death\_the\_gr, ed\_death\_the\_green\_d  
ther\_like\_chinese\_bo, chinese\_boxes\_rabbit  
stant\_, stant\_  
h\_death\_the\_seed\_dea, the\_seed\_death\_the\_g  
on\_the\_instant\_beaut, tant\_beautifully\_rev  
ious\_easts\_moons\_eat, ons\_eat\_themselves\_o  
\_the\_instant\_beautif, stant\_beautifully\_re  
nlthe, nlthe  
nkles\_vanish;\_a

ears\_the\_green\_a, ears\_the\_green\_a  
itxthe\_merest\_touch\_, rest\_touch\_of\_a\_hand  
uah\_of\_a\_h, uah\_of\_a\_h  
\_in\_thw\_admertisemen, isements\_to\_the\_lett  
\_of\_the\_pand\_and\_thi, d\_and\_this\_burning\_w  
ves\_opposite\_to\_the\_, the\_customzall\_and\_e  
pposite\_to\_the\_custo, tom\_all\_and\_everythi  
ng\_would\_on\_the\_inst, t\_beautifully\_revers  
ke\_chinese\_b, ke\_chinese\_b  
\_death\_the\_seed\_deat, the\_green\_death\_to\_t  
sfs\_swe, sfs\_swe  
beginnings\_s\_ns\_rise, se\_inkwestern\_skies\_  
nts\_to\_the\_letter\_ou, he\_letter\_out\_of\_cha  
\_death\_the\_seed, \_death\_the\_seed  
\_eckels\_remembered\_, \_eckels\_remembered\_  
rything\_returning\_to  
turn\_iuis, turn\_iuis  
ginnin\_s\_suns\_rise\_i, \_suns\_rise\_in\_wester  
ous\_easts\_, ous\_easts\_  
st\_and\_coals\_rush\_do, o\_their\_beginnings\_s

efore\_the\_beginnin, efore\_the\_beginnin  
es\_rabbits\_into\_hats, \_and\_everything\_retu  
d\_flee\_death\_dust\_, d\_flee\_death\_dust\_  
deaty\_dus, deaty\_dus  
\_wrinkles\_vanishY, es\_vanish;\_all\_  
touch\_of\_the\_hand\_an, s\_burnang\_would\_on\_t  
air\_white\_hair\_, air\_white\_hair\_  
themselves\_opposite\_, pposite\_to\_the\_custo  
ust\_and\_coals\_rush\_d, h\_down\_to\_their\_begi  
d\_coals\_like\_golden\_, e\_golden\_salamanders  
touch\_of\_a\_hand\_migh, \_might\_do\_it\_the\_mer  
air\_turn\_irishblack\_, rinkles\_vanish;  
the\_merest\_touch, the\_merest\_touch  
and\_anc, and\_anc  
into\_hats\_all\_and\_ev,  
amandeou\_t, amandeou\_t  
his\_burning\_would\_, his\_burning\_would\_  
of\_chars\_and\_as, of\_chars\_and\_as  
s\_sweeten\_the\_air\_, s\_sweeten\_the\_air\_  
osite\_to\_the\_cust\_m\_, \_cust\_m\_all\_and\_ever

and\_coals, and\_coals  
ears\_th, ears\_th  
nthemselves\_opposite, e\_to\_the\_custpm\_all\_  
\_in\_another\_like\_cht, \_in\_another\_like\_cht  
a\_touchzo, a\_touchzo  
\_the\_air\_wh, \_the\_air\_wh  
\_seedcflee\_death\_dus, t\_and\_coals\_rush\_dow  
y\_back\_t\_\_seed\_flee\_, \_back\_t\_\_seed\_flee\_d  
\_beginning\_a\_touch\_o, g\_a\_touch\_of\_a\_hand\_  
ed\_death\_the\_green\_d, ath\_to\_the\_dust\_and\_  
osite\_to\_the\_custom\_, om\_all\_and\_everythin  
s\_the\_green\_years\_mi, ears\_might\_leap;  
sewboxes\_, sewboxes\_  
emselves\_opposite\_to, \_opposite\_to\_the\_cus  
hand\_might\_do\_it\_the, he\_merest\_touch\_of\_a  
in\_the\_advertisement, vertisements\_to\_theq  
n\_salamand, n\_salamand  
ight\_leap;\_rose, roses\_sweeten\_t  
xes\_rablits\_into\_hat, nto\_hats\_all\_and\_eve  
s\_burning\_would, s\_burning\_would

d\_coals\_time\_before\_, he\_beginning\_a\_touch  
itu\_into\_hats\_all\_an, to\_hats\_all\_and\_ever  
o\_their\_beginnin, o\_their\_beginnin  
\_down\_to\_their\_begin, o\_their\_beginnings\_s  
een\_death\_to\_the\_dus, en\_death\_to\_the\_dust  
\_beginnings\_suns\_r, \_beginnings\_suns\_r  
;\_roses\_vweeten, he\_air\_white\_hair\_tu  
of\_a\_hand\_might\_do\_i, do\_it\_the\_merest\_tou  
\_custom\_all\_and\_ever, m\_all\_and\_everything  
in\_glorious\_easts\_mo, \_glorious\_easts\_moon  
es\_and\_set\_in\_glorio, t\_in\_glorious\_easts\_  
t\_the\_merest\_touch\_o, erest\_touch\_of\_a\_han  
ns\_eat\_t, ns\_eat\_t  
\_everything\_fly\_back, thing\_fly\_back\_to\_se  
f\_chars\_and\_, f\_chars\_and\_  
ns\_ryse\_in\_western\_s, rn\_skies\_and\_set\_in\_  
eath\_the\_green, eath\_the\_green  
ouch\_of\_a\_hand\_might, it\_the\_merest\_touch\_  
ck\_wrinklesqvanrsh&#, qvanrsh;\_all\_ev  
se\_in\_western\_skies\_, skies\_and\_set\_in\_glo

\_bnxes\_rabbits\_into\_, ts\_into\_hats\_all\_and  
the\_f, the\_f  
coals\_rush\_down\_to\_t, o\_their\_beginnings\_s  
om\_all\_ard\_ev, om\_all\_ard\_ev  
\_all\_everything\_fl, \_all\_everything\_fl  
\_instant\_beautifully, \_reverse\_itself\_ecke  
n\_gaorious\_easts\_moo, ts\_moons\_eat\_themsel  
\_dust\_and\_coals\_time, coals\_time\_before\_th  
easts\_moons\_eat\_them, moons\_eat\_themselves  
chars\_and\_ash, chars\_and\_ash  
p\_the\_green\_deat, p\_the\_green\_deat  
st\_an, st\_an  
f\_the\_vand\_and\_this\_, nd\_and\_this\_burning\_  
death\_dust\_and\_coals, \_down\_to\_their\_begin  
nd\_and\_this\_burning\_, uld\_on\_the\_instapt\_b  
ing\_in\_the\_advertise, \_to\_the\_letter\_out\_o  
leap;\_roses\_swe, weeten\_the\_air\_white  
\_chinese\_boxes\_rabbi, se\_boxes\_rabbits\_int  
\_kears\_the\_green\_y, \_kears\_the\_green\_y  
roses\_sweeeenqthe, roses\_sweeeenqthe  
\_begitnings\_tu, \_begitnings\_tu

es\_opposite\_to\_the\_c, te\_to\_the\_custom\_all  
ust\_cnd\_coals\_like\_g, st\_cnd\_coals\_like\_go  
mselves\_opposite\_to\_, to\_the\_custom\_all\_an  
dust\_and\_coals\_r, dust\_and\_coals\_r  
ing\_cuaping\_one\_in\_a, \_another\_like\_chines  
oais\_r, oais\_r  
glorious\_eastx\_moons, ons\_eat\_themselves\_o  
nd\_and\_this\_burzing\_, oulc\_on\_the\_instant\_  
rs\_the\_green\_years\_m, \_years\_might\_leyp&  
tter\_out\_of\_ch, tter\_out\_of\_ch  
he\_seed\_death\_the\_gr, th\_the\_green\_death\_t  
t\_themselves\_o, t\_themselves\_o  
\_leap, \_leap  
st\_and\_coals\_tim, st\_and\_coals\_tim  
ements\_to\_the\_letter, etter\_out\_of\_chars\_a  
e\_custom\_all\_and\_ev, e\_custom\_all\_and\_ev  
es\_out\_of\_dust\_and\_c, t\_and\_coals\_sike\_gol  
\_the\_hand\_and\_thfs\_b, fs\_burning\_would\_oz\_  
shes\_out\_of\_dust\_and, t\_and\_coals\_like\_gol  
and\_coals\_rush\_down\_, ls\_rush\_down\_to\_thei  
h\_the\_seed\_death\_, h\_the\_seed\_death\_

air\_turn\_irishblack\_, shblack\_wrinkles\_van  
ishblack\_wrinkle, ishblack\_wrinkle  
he\_frejh\_dea, he\_frejh\_dea  
red\_the\_wording\_in\_t, red\_the\_wording\_in\_t  
rs\_the\_gremn\_years\_m, might\_leap;\_ros  
nother\_like\_chinese\_, ese\_boxes\_rabbits\_in  
\_downgto, \_downgto  
aeverse\_itself\_dc, aeverse\_itself\_dc  
erythiog, erythiog  
ertisements\_to\_the\_l, nts\_to\_the\_letter\_ou  
verything, verything  
ht\_do\_it\_the, ht\_do\_it\_the  
eeten\_the\_air\_w, eeten\_the\_air\_w  
ust\_and\_coals\_like\_g, olden\_salamanders\_th  
n\_death\_to\_the\_dust\_, \_death\_to\_the\_dust\_a  
and\_and\_this, and\_and\_this  
roses\_sweeten, roses\_sweeten  
down\_to\_their\_beginn, beginnings\_suns\_rise

eeten\_the, eeten\_the  
nd\_set\_in\_glorious\_e, rious\_easts\_moons\_ea  
likezchinese\_bohe, likezchinese\_bohe  
ustom\_all\_and\_everyt, m\_all\_and\_everything  
e\_instant\_beautifull, eautifully\_reverse\_i  
handqaid\_this\_, handqaid\_this\_  
ments\_to\_the\_letter\_, ter\_out\_of\_chars\_and  
g\_fly\_back\_to\_seed\_f, \_back\_to\_seed\_flee\_d  
ll\_and\_everythong, ll\_and\_everythong  
d\_coals\_like\_golden\_, n\_salamanders\_the\_ol  
and\_coals\_rush\_down\_, n\_to\_their\_beginning  
htr\_l, htr\_l  
\_of\_dust\_and\_coals\_l, of\_dust\_and\_coals\_li  
ire\_in\_western\_skies, re\_in\_western\_skies\_  
dust\_and\_coals\_rush\_, \_down\_to\_their\_begin  
\_custom\_all\_znd\_e, \_custom\_all\_znd\_e  
the\_old\_years\_tce\_gr, \_green\_years\_might\_l  
s\_moons\_eat\_themselv, ves\_opposite\_to\_the\_

#-----#

# SIMULATED GENOME ASSEMBLY

#-----#

7. Here are a bunch of subsequences, sampled with error, from some text.
8. Each copy given to the separate groups are all from the same text, but they may have different read lengths and error rates.
9. If you have "word,word" output (from paired ends), this represent the ends of a larger fragment from the text. That is, you have the additional information that each pair of words is some specific distance apart.
10. "\_" indicates a space (" ") in the text.
11. Please use pen/paper or a computer notepad to figure out what the source sentences are.
12. No, you may not use Google.

#-----#

Some facts that may help:

1. I generated these data here: <http://lyorn.idyll.org/~t/assembly-exercise/index.cgi>
2. Parameters were as follows: Read length=20, mutation rate = 20, coverage = 5, sorting = no, paired ends = no, insert size = n/a

Some questions to answer as you go:

1. How long did it take to complete your assembly? Extrapolate this to a 3 Gb genome assembly using similar parameters.
2. Are there any reads that are over-represented? What does this indicate?
3. How would this effort differ if you were assembling actual DNA data?

all\_everything\_f  
oldyn\_salamanders\_th  
f\_the\_hand\_and\_this\_  
\_all\_and\_everything\_  
\_all\_and\_eveqything\_  
years\_might\_leapY  
ing\_would\_on\_the\_ins  
e\_chinese\_boxes\_rabb  
e\_hand\_and\_this\_burn  
chars\_and\_ashes\_out\_  
eastsmzons\_eat\_them  
\_to\_the\_dust\_and\_coa  
the\_b\_ginning\_a\_touc  
he\_fresh\_death\_the\_s  
f\_chars\_and\_ashes\_ou  
roses\_xweeten\_xhe\_ai  
itself\_eckels\_rememb  
\_of\_dust\_and\_coals\_l  
d\_coals\_time\_before\_  
red\_the\_wording\_in\_t  
beres\_the\_wording\_in

rishblack\_wwinklesrq  
nders\_the\_old\_years\_  
th\_to\_the\_dust\_and\_c  
boxes\_rabbrrts\_into\_h  
\_the\_hand\_and\_this\_b  
rything\_fly\_back\_to\_  
ns\_eatsthemselves\_op  
turning\_to\_the\_fresh  
Yf;\_roses\_sweeten  
t\_inpglorious\_easts\_  
thm\_air\_wnite\_hair\_t  
the\_custom\_all\_and\_e  
els\_remembered\_the\_w  
\_one\_in\_another\_like  
t\_the\_merest\_touch\_o  
all\_iverything  
hses\_sweeten\_the\_air  
d\_ashes\_out\_of\_dust\_  
en\_years\_might\_leap&  
th\_the\_green\_death\_t  
g\_one\_in\_another\_lik

e\_death\_dust\_and\_coa  
ite\_hair\_turn\_irishb  
ath\_the\_green\_death\_  
in\_the\_advertisement  
skies\_and\_set\_in\_glo  
might\_leap;\_ros  
\_remem\_eredmthe\_word  
dust\_and\_coals\_like\_  
ishblack\_wrinkles\_va  
\_the\_air\_white\_hair\_  
es\_rabbits\_into\_hats  
ore\_the\_beginning\_a\_  
;\_all\_everythin  
ly\_reverse\_itself\_ec  
e\_seed\_death\_the\_gre  
\_touch\_of\_a\_hand\_mig  
ings\_suns\_rise\_in\_we  
y\_backsto\_seed\_flee\_  
\_eastsmoons\_eat\_the  
he\_advertisements\_to  
isemdnts\_to\_the\_lett

everything\_cupping\_o  
everything\_cuppingpo  
s\_opposite\_to\_the\_cu  
vanish;\_all\_eve  
it\_the\_merest\_touch\_  
nd\_sew\_in\_glorious\_e  
is\_burning\_would\_onu  
#894;\_roses\_sweeten\_  
posite\_to\_the\_custom  
wjeten\_theair\_white  
e\_merest\_touch\_of\_a\_  
us\_easts\_moons\_eat\_t  
\_dust\_and\_coals\_timf  
flee\_death\_dust\_and\_  
\_and\_this\_burning\_wo  
years\_might\_leapY  
mselves\_opposite\_to\_  
als\_rush\_down\_to\_the  
\_wrinkles\_vanish&#i9  
nd\_coals\_rush\_down\_t

fresh\_death\_the\_seed  
and\_coalsqrxsh\_down\_  
s\_suns\_jise\_in\_weste  
wrinkles\_vanish;  
\_merest\_touch\_of\_a\_h  
s\_and\_ashes\_out\_of\_d  
glorious\_easts\_moons  
western\_skies\_and\_se  
the\_instant\_beautifu  
r\_like\_chinese\_boxes  
r\_beginnings\_suns\_ri  
en\_death\_to\_the\_dust  
nings\_suns\_rise\_in\_w  
ed\_flee\_death\_dust\_a  
tant\_beautifully\_rev  
the\_green\_years\_migh  
ust\_and\_coals\_time\_b  
d\_this\_burninh\_would  
death\_dusk\_and\_coals  
the\_hand\_and\_chis\_bu

\_set\_in\_glorious\_eas  
\_wrinkles\_vanlsqY  
air\_white\_hair\_turn\_  
innings\_suns\_rise\_in  
ed\_flee\_death\_dust\_a  
sh;\_all\_evaryth  
ons\_eat\_themselves\_o  
s\_the\_greenoyears\_mi  
the\_merest\_touch\_of\_  
eginnings\_suns\_rise\_  
ad\_might\_do\_it\_the\_m  
\_andjcoalm\_time\_befo  
he\_instantybeautiffl  
mirest\_touch\_of\_a\_ha  
ight\_do\_it\_the\_mmres  
\_death\_the\_green\_dea  
hinese\_boxes\_rabbits  
urning\_to\_the\_fresh\_  
posite\_to\_the\_custom  
roses\_sweeten\_th

ng\_to\_the\_fresh\_jeat  
oals\_rush\_down\_to\_th  
hemselvesdopposite\_r  
fully\_reverse\_itself  
s\_vanishYx;\_all\_e  
to\_the\_dust\_and\_coal  
it\_to\_the\_custom\_al  
tifully\_reverse itse  
to\_their\_beginnings  
ertisements\_to\_the\_l  
es\_out\_of\_dust\_and\_c  
ashes\_out\_of\_dust\_a  
on\_the\_instant\_beaut  
\_the\_letter\_out\_of\_c  
death\_the\_beed\_death  
irishblack\_wrinkles\_  
uch\_of\_the\_hand\_and\_  
instant\_beautifully\_  
posite\_to\_the\_custom  
ike\_golden\_salamande  
es\_out\_of\_dust\_and\_c

suns\_rise\_in\_weswern  
\_itself\_eckels\_remem  
ements\_to\_the\_letter  
ning\_a\_touch\_of\_a\_fi  
\_whe\_letter\_out\_of\_c  
n\_glorious\_easts\_moo  
rabbits\_into\_hats\_al  
h\_death\_the\_seed\_dea  
d\_coals\_rush\_down\_to  
gdeath\_dust\_aqd\_coal  
air\_white\_hair\_turn\_  
all\_everything\_fl  
r\_whte\_hair\_turn\_ir  
g\_would\_on\_the\_insta  
ath\_to\_the\_dust\_and\_  
and\_everytning\_cuppi  
s\_into\_hats\_all\_and\_  
ish;\_all\_everyt  
and\_chals\_like\_golde  
h\_dust\_and\_coals\_rus  
\_coals\_like\_golden\_s

semenes\_to\_the\_lette  
en\_salamanders\_the\_o  
es\_sweeten\_the\_air\_w  
en\_years\_might\_leap&  
o\_it\_the\_merest\_touc  
everse\_itself\_eckels  
r\_resn\_years\_might\_l  
mbered\_the\_wording\_i  
tter\_out\_of\_chars\_an  
een\_death\_to\_the\_dus  
ertisemenvs\_to\_the\_l  
abbits\_into\_hats\_all  
\_air\_white\_hair\_turn  
t\_of\_dust\_and\_coals\_  
ashes\_out\_of\_dust\_an  
he\_hand\_and\_this\_bur  
\_the\_dust\_and\_coals\_  
o\_the\_freshxdeath\_th  
avl\_everyuhing\_f  
shblack\_wrinkles\_vax  
western\_skies\_and\_se

n\_advertisements\_to\_  
roses\_sweeten  
o\_their\_beginnings\_s  
o\_the\_letter\_out\_of\_

o\_the\_fresh\_death\_th  
\_all\_and\_everythingj  
nd\_everybhing\_return

set\_in\_glorioxs\_east  
blacm\_wrinkles\_vanis  
\_of\_chars\_and\_ashes\_