

# How do we store data?

Steven Salzberg

**Data structures allow us to store  
large amounts of data efficiently**

```

Q5E940_BOVIN -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_HUMAN -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_MOUSE -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_RAT -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_CHICK -----MPREDRATWKSNYFMKTIQLDDYPKCFVVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_RANSY -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQQIRMSLRGK-AVVLGMGKNTMMRKAIRGHLENN--SALE 76
Q7ZUG3_BRARE -----MPREDRATWKSNYFLKTIQLDDYPKCFIVGADNVGSKMQOTIRLSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0 ICTPU -----MPREDRATWKSNYFLKTIQLNDYPKCFIVGADNVGSKMQOTIRLSLRGK-AVVLGMGKNTMMRKAIRGHLENN--PALE 76
RLA0_DROME -----MVRENKAARKAQYFIKVVLEFDEFKCFIVGADNVGSKMQONIRTSLRGL-AVVLGMGKNTMMRKAIRGHLENN--POLE 76
RLA0_DICDI -----MSGAG-SKRKKLFIEKATKLFTTYDKMIVAADDFVGSQLOKIRKSIIRGI-CAVLMGKKIMIRKVIRDLADSK--PELD 75
Q54LP0_DICDI -----MSGAG-SKRKNVFIEKATKLFTTYDKMIVAADDFVGSQLOKIRKSIIRGI-CAVLMGKKIMIRKVIRDLADSK--PELD 75
RLA0_PLAF8 -----MAKLSKQQKKQMYIEKLSLLIQQYSKILIVHVDNVGSKMQASVRKSLRGK-ATILMGKNTIRIRTALEKKNLQAV--POIE 76
RLA0_SULAC -----MIGLAVITTTKKIAKWKVDEVAELTKLKHKTITIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNLFNIALKNAG--YDIK 79
RLA0_SULTO -----MRIMAVITQERKIAKWKIEVKELEQLREYHTIIITIANIEGFPADKLHDIRKKMRGM-AEIKVTKNTLFGIAAKNAG--LDVS 80
RLA0_SULSO -----MKRLALALKQRKVASWKLLEVKELELIKNSNTIILGNLEGFPADKLHEIRKKLRGK-ATIKVTKNTLFFKIAAKNAG--IDIE 80
RLA0_AERPE MSVVSIVGQMYKREKPIPEWKTLMLELEELFSKRRVVFADLTGTPFVVVRVKKLWKK-YDMMVAKRIILRAMKAAGLE--LDDN 86
RLA0_PYRAE MMLAIGKRRYVRTQYPARKVKIVSEATELLQKYPYVFLFDLHGLSRILHEYYRRLERY-GVIKIIPKTLFKIAFTKVYGG--IPAE 85
RLA0_METAC -----MAEERHHTEHIPQWKDEIENIKELIQSHKVFGMVGIEGILATKMKIRRDLDKV-AVLKVSNTLIERALNQLG---ETIP 78
RLA0_METMA -----MAEERHHTEHIPQWKDEIENIKELIQSHKVFGMVRIEGILATKIKIRRDLDKV-AVLKVSNTLIERALNQLG---ESIP 78
RLA0_ARCFU -----MAAVRGS--PPEYKVRVEEIKRMISSKPVVAIVSFRNVPAQOMOKIRREFRGK-AEIKVYKNTLIERALDALG--GDYL 75
RLA0_METKA MAVKAKGPPSGEYEPKVAEWKRREVEKLELMDEYENVGLVDLEGIPAPQLQEIIRAKLREEDTIIRMSRNTLMRIALEEKLDER--PELE 88
RLA0_METTH -----MAHVAEWKKKEVQQLHDLIKGEYEVVGIANLADIPAROLQKMRQTLRDS-ALIRMSKKILLISLALEKAGREL--ENVD 74
RLA0_METTL -----MITAESEHKIAPWKIEEYVNLKLELLKNGQIVALVDMMEVPAROLQEIIRDKIR-ETMTLKMRSNTLIERAIEVAETGNEPEFA 82
RLA0_METVA -----MIDAKSEHKIAPWKIEEYVNLKLELLKNSANVIALIDMMEVPAVOLQEIIRDKIR-DMTLMKMSRNTLIERAIEVAETGNEPEFA 82
RLA0_METJA -----METKVKAHVADWKIEEVKTLKGLIKSKPVVAIVDMMDVPAPQLQEIIRDKIR-DKVKLRMSRNTLIERALKEAAEELNPEKLA 81
RLA0_PYRAB -----MAHVAEWKKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSNTLIERALKEAAEELGKPELE 77
RLA0_PYRHO -----MAHVAEWKKKEVEELAKLIKSPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSNTLIERALKEAAEELGKPELE 77
RLA0_PYRFU -----MAHVAEWKKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRLIRENGLLRVSNTLIERALKEAAEELGKPELE 77
RLA0_PYRKO -----MAHVAEWKKKEVEELANLIKSPVIALVDVAVPAPPLSKMRDKLR-GKALLRVSNTLIERALKEAAEELGQPELE 76
RLA0_HALMA -----MSAESEKKTETPEWKQEEVDVAIVMIESYESVGVVNIAGIPSRLODMRRDLHGT-AELRVSNTLIERALDDVD--DGLE 79
RLA0_HALVO -----MSESEVRQTETVPQWKREEVDLVDLIESYESVGVVGVAGIPSRLODMRRDLHGS-AAVRMSRNTLVNRALDEVN--DGFE 79
RLA0_HALSA -----MSAEEQRTTEEPPEWKQEEVAELVDLLETYDSGVVNVNVTGIPSKLODMRRGLHGO-AALRMSRNTLLVRALEEAG--DGLD 79
RLA0_THEAC -----MKESVQOKKELYNETTRIKASRSVAIVDAGIRIRROIODIRGNRGRK-INLKVIKKLLFFKALENLGD---EKLS 72
RLA0_THEVO -----MRKINPKKEIYSELAQDITSKAIVADIKGVRIROMODIRAKNRDK-VKIKVVKLLFFKALDSIND---EKLT 72
RLA0_PICTO -----MTEPAQWKIDFYKNLENIINSRKVAAIVSTKGLRNNFQKIRNSIRDK-ARIKVSARLLRLALIENTGK---NNIV 72
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90

```

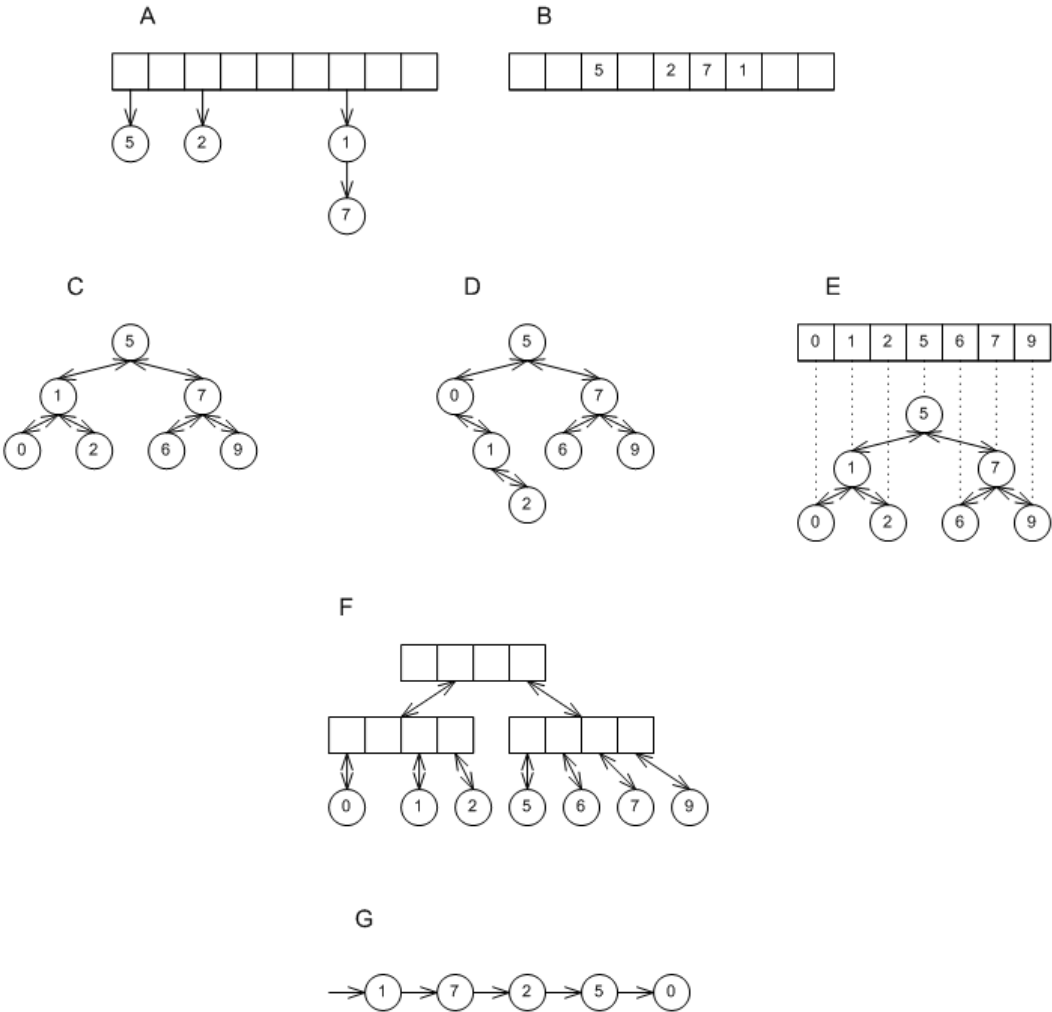


# We need to be able to find stuff



Don't  
panic

it's only a  
tree



CGTAGCTAGCATGCATCAGTCACA

1 binary digit = 1 bit = 0 or 1

1 byte = 8 bits = 0000 0000

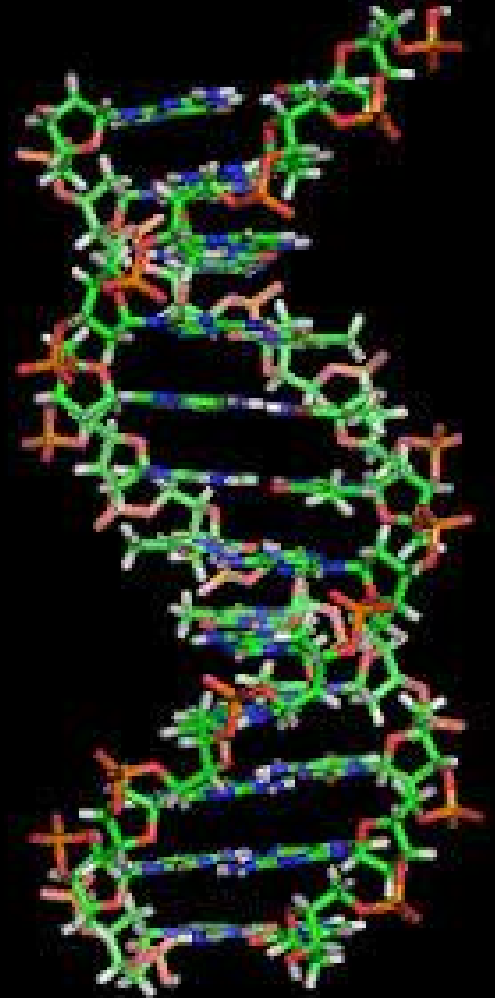
$\text{DNA} = \{A, C, G, T\}^*$

A = 00

C = 01

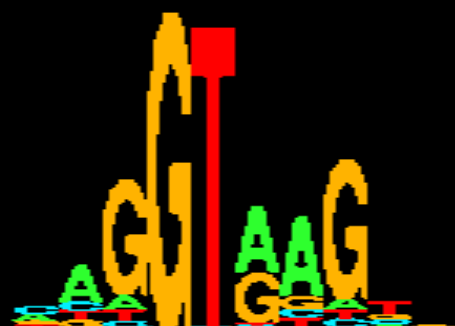
G = 10

T = 11



Donor

Acceptor



caggttaa

a\_ggttaag

c\_ggtta\_g

aaggtgag

a\_ggttaagt



ccccc\_bccccc\_tatat\_tccct\_cagg

\_ccccc\_gaggctttttttttt\_agg

bccccc\_tatat\_tatat\_tcc\_cagg

ccccc\_gaggctttttttttt\_cagg

ccccc\_gaggctttttttttt\_cagg