

Why compare DNA or protein sequences?

Partial CTCF protein sequence in 8 organisms:

<i>H. sapiens</i>	-EDSSDS-ENAEPLDLDNEDEEEPAVEIEPEPE-----PQPVTTPA
<i>P. troglodytes</i>	-EDSSDS-ENAEPLDLDNEDEEEPAVEIEPEPE-----PQPVTTPA
<i>C. lupus</i>	-EDSSDS-ENAEPLDLDNEDEEEPAVEIEPEPE-----PQPVTTPA
<i>B. taurus</i>	-EDSSDS-ENAEPLDLDNEDEEEPAVEIEPEPE-----PQPVTTPA
<i>M. musculus</i>	-EDSSDSEENAEPLDLDNEEEEEPAVEIEPEPE--PQPQPPPPQPVPAPA
<i>R. norvegicus</i>	-EDSSDS-ENAEPLDLDNEEEEEPAVEIEPEPEPQPQPPPPQPVPAPA
<i>G. gallus</i>	-EDSSDSEENAEPLDLDNEDEEETAVEIEAPE-----VSAEAPA
<i>D. rerio</i>	DDDDDDSDDEHGEPLDLDIDEEDEDDL-LDEDQMGLLDQAPPSVPIP-APA

- Identify important sequences by finding conserved regions.
- Find genes similar to known genes.
- Understand evolutionary relationships and distances (*D. rerio* aka zebrafish is farther from humans than *G. gallus* aka chicken).
- Interface to databases of genetic sequences.
- As a step in genome assembly, and other sequence analysis tasks.
- Provide hints about protein structure and function (next slide).