

Genomics in Action - DNA Sequences Compared

Genomics allows us to search genomes and align sequences to identify and compare closely-related genes within one organism, or between organisms.

The figure below shows the sequence alignment of the human *mlh1* gene with genes that encode proteins with the same function in humans and other organisms. This is the first half of the sequence in the bead curtain.

Human <i>mlh1</i>	GTTATCCAGC	GGCCAGCTAA	TGCTATCAAA	GAGATGATTG	AGAACTGTTT	AGATGC AAAA	TCCACAAGTA	TTCAA
Mouse <i>mlh1</i>	GTCATT CAGC	GGCCGGCCAA	TGCTATCAAA	GAGATGATAG	AAAAC T GTTT	AGATGC AAAA	ICTACAAATA	TTCAA
Fruit fly <i>mlh1</i>	ATTATTCAGC	GACCCGCCAA	CGCCCTAAAA	GAGCTGCTCG	AGAACAGTTT	GGACGCCCAA	TCGACCCACA	TCCAG
Nematode <i>mlh1</i>	GTA CTAGCCC	GGCCGTGTAA	TGCCAT TAAA	GAGCTCGTTG	AGAACTCACT	GGACGCCGGC	GCCACCGAAA	TTATG
Mustard <i>mlh1</i>	GTAATCCAGC	GTCCAGTTTC	AGCTGTGAAA	GAGCTCGTTG	AGAACAGCCT	CGACGCCGAT	TCAAGTTCCA	TAA GC
Yeast <i>mlh1</i>	ATCATAATAT	CCCCCGTAAA	TGCTCTCAAA	GAAATGATGG	AGAATTCCAT	CGATGCGAAT	GCTACAATGA	TTGAT
<i>E. coli mutL</i>	GTGGTCGAGC	GACCTGCGTC	GGTAGTCAAA	GAAGTAGTGG	AAAACAGCCT	CGATGCAGGT	GCGACGCGTA	TCGAT
Human <i>pms1</i>	ATCATCACTT	CGGTGGTCAG	TGTTGTAAAA	GAGCTTATTG	AAAAC T C C T T	GGATGCTGGT	GCCACAAGCG	TAGAT
Mouse <i>pms1</i>	ACAATCACGT	CAGTGGTCAG	CGTTGTGAAA	GAGCTCATTG	AAAAC T C C T T	GGATGCCGGA	GCCACCAGCA	TTGAA
Mustard <i>pms1</i>	GTCATCTTAG	ACCTCTCTTC	GGCCGTCAA	GAGCTTGTTC	AGAATAGTCT	CGACGCCGGC	GCCACAGTA	TAGAG
Yeast <i>pms1</i>	GTTATCACCG	ACTTAACAAC	TGCAGT GAAA	GAAGTCTGTTG	ATAATAGTAT	AGATGCCGAA	GCAAACCAGA	TCGAA
Human <i>pms2</i>	GTGGTACTGA	GTCTAAGCAC	TGCGGTAAA	GAGTTAGTAG	AAAACAGTCT	GGATGCTGGT	GCCACTAATA	TTGAT
Mouse <i>pms2</i>	GTGATACTCA	GTTTAAGCAC	CGCTGTGAA	GAGTTGATAG	AAAATAGTGT	AGATGCTGGT	GCTACTACTA	TTGAT
Fruit fly <i>pms2</i>	GTGGTTCTCA	GCTTGGCTGT	GGCTGTCAA	GAGCTAGTGG	AGAACTCAAT	AGATGCCGGC	GCAACGCTGG	TGGAG
Nematode <i>pms2</i>	GTTGTTGTCA	GCTTGTCAA	TGCAATTCGG	CAGTTAATCG	ACAATTC AAT	CGATGCTGGT	TCGACAATTA	TCGAT
Human <i>mlh3</i>	GCCATAAGCT	CCTTGGGCCA	ATGTTGTGAG	GAACTTGCCC	TCAACAGTAT	TGATGCTGAA	GC AAAATGTG	TGGCT

The human genome contains 1 copy of the *mlh1* gene. In addition, it has 3 closely-related genes (*pms1*, *pms2* and *mlh3*) that encode proteins that carry out the same function.

Forms of *mlh1* exist in all forms of life on earth, from bacteria and archaea to plants and animals. Note how similar the DNA sequences are from very different organisms. This suggests that the proteins encoded by these genes carry out a function critical to life.

Genomics in Action - Protein Sequences Compared

Comparing AA sequences of different proteins to find parts that are conserved allows scientists to predict which parts of the protein are critical to its function.

The figure below shows the AA sequence alignment of the proteins encoded by the *mlh1* gene in humans and other organisms. Each letter represents a specific AA.

	A		1	2	B
Human Mlh1	VIQR PANAIK	EMIENCLDAK	STSIQVIVKE	GGLKLIQIQD	NGTGIRKEDL
Mouse Mlh1	VIQR PANAIK	EMIENCLDAK	STNIQVVVKE	GGLKLIQIQD	NGTGIRKEDL
Fruit fly Mlh1	IIQR PANALK	ELLENSLDAQ	STHIQVQVKA	GGLKLLQIQD	NGTGIRREDL
Nematode Mlh1	VLARPCNAIK	ELVENSLDAG	ATEIMVNMQN	GGLKLLQVSD	NGKGIEREDF
Mustard Mlh1	VIQR PVS AVK	ELVENSLDAD	SSSISVVVKD	GGLKLIQVSD	DGHGIRRELD
Yeast Mlh1	IIIS PVNALK	EMMENSIDAN	ATMIDILVKE	GGIKVLQITD	NGSGINKADL
<i>E. coli MutL</i>	VVERP ASVVK	ELVENSLDAG	ATRIDIDIER	GGAKLIRIRD	NGCGIKKDEL

▲ ▲ ▲ ▲ ▲

- The marked AA (▲) are critical to ATP binding. Note that they are identical in the proteins from all organisms from bacteria to humans.
- In addition, many other AA are conserved. Even if the specific AA is different, the AA is often replaced by an AA with similar characteristics (vertically-shaded boxes indicate conserved positions; their color indicates similar characteristics). For example, for the first AA in the protein, "V" and "I" are both hydrophobic and are thus functionally similar.
- The fact that these AAs have been conserved from bacteria to humans indicates they are critical to the function of Mlh1.

Red and green boxes above the sequences correspond to segments in the MutL 3-dimensional structure ("Mlh1 Structure and Function" panel and protein model).

- A - conserved hydrophobic AA
- K - conserved positively-charged AA
- E - conserved negatively-charged AA
- N - conserved hydrophilic AA