

## Using Protein Sequences to Test Gene Transfer

Carotenoids are colored molecules that have diverse functions in many organisms, such as photosynthesis in plants and light detection in animals. Plants and many microorganisms can synthesize carotenoids from scratch, but animals generally cannot (they must obtain carotenoids from their diet). One exception is the pea aphid *Acyrtosiphon pisum*, a small, plant-dwelling insect whose genome includes a full set of genes for the enzymes needed to make carotenoids. Because other animals lack these genes, it is unlikely that aphids inherited them from a single-celled common ancestor shared with microorganisms and plants. So where did they come from? Evolutionary biologists hypothesized that an aphid ancestor acquired these genes by horizontal gene transfer from distantly related organisms.

Scientists obtained the DNA sequences for the carotenoid-biosynthesis genes from several species, including aphids, fungi, bacteria, and plants. Computer software was then used to convert the DNA sequences into amino acid sequences. This allowed the researchers to compare the corresponding polypeptides in the different organisms.

The sequences in Table 1 show the first 60 amino acids of one polypeptide of the carotenoid-biosynthesis enzymes in the plant *Arabidopsis thaliana* and the corresponding amino acids in five non-plant species, using the one-letter abbreviations for the amino acids. A dash (–) indicates a gap inserted in a sequence to optimize its alignment with the corresponding sequence in *Arabidopsis*.

**Table 1 Amino acid sequences of carotenoid-biosynthesis enzymes from several species**

Organism	Alignment of Amino Acid Sequences
<i>Acyrtosiphon</i> (aphid)	IKIIIIIGSGV GGTAAAARLS KKGFOVEVYE KNSYNGGRCS IIR-HNGHRF DQGPSL--YL
<i>Ustilago</i> (fungus)	KKVVIIGAGA GGTALAAARLG RRGYSVTVLE KNSFGGGRCS LIH-HDGHRW DQGPSL--YL
<i>Gibberella</i> (fungus)	KSVIVIGAGV GGVSTAARLA KAGFKVTILE KNDFTGGRCS LIH-NDGHRF DQGPSL--LL
<i>Staphylococcus</i> (bacterium)	MKIAVIGAGV TGLAAAARIA SQGHEVTIFE KNNNVGGRMN QLK-KDGFTF DMGPTI--VM
<i>Pantoea</i> (bacterium)	KRTFVIGAGF GGLALAIRLQ AAGIATTVLE QHDKPGRAY VWQ-DQGFTF DAGPTV--IT
<i>Arabidopsis</i> (plant)	WDAVVIGGGH NGLTAAAYLA RGGLSVAVLE RRHVIGAAV TEEIVPGFKF SRCSYLQGLL

1. In the rows of data for the organisms being compared with the aphid, highlight the amino acids that are identical to the corresponding amino acids in the aphid.
2. Identify the organism that has the most amino acids in common with the aphid. Rank the partial polypeptides from the other four organisms in degree of similarity to that of the aphid.
3. a) Do these data support the hypothesis that aphids acquired the gene for this polypeptide by horizontal gene transfer? Provide support for your answer.  
b) If horizontal gene transfer did occur, what type of organism is likely to have been the source?
4. Identify additional evidence that would support your hypothesis.
5. Propose an explanation for the similarities between the aphid sequence and the sequences for the bacteria and plant.