

Picturing the genetic code

A.W.F.Edwards
Gonville and Caius College
Cambridge, CB2 1TA, UK
awfe@cam.ac.uk

An important characteristic of the genetic code is that changing one of the bases of a triplet codon often leads to no change in the corresponding amino-acid. Analogously, a Venn diagram possesses the property that each of the areas into which the total space (or ‘universe’) is partitioned is surrounded by adjoining areas which differ from it in respect of only a single set character¹.

It is therefore natural to enquire whether placing the 64 triplets of the genetic code in a six-set Venn diagram ($64 = 2^6$) might not be possible in such a way that triplets differing by a single base but coding for the same amino-acid appear in adjoining areas. If so, blocks of triplets corresponding to each amino acid might be generated.

The following solution shows that this is indeed possible. Starting with a six-set Edwards–Venn diagram¹ (Figure 1) separate it into three diagrams by taking the sets in pairs and labelling their areas with the four bases as shown in Figure 2. The first diagram refers to the first position, the second to the second and the third to the third. Now reassemble the original diagram by overlaying the three separate ones (Figure 3). Each of the 64 areas of this complete diagram corresponds to a triplet, and each triplet is surrounded only by triplets that differ from it by exactly one base.

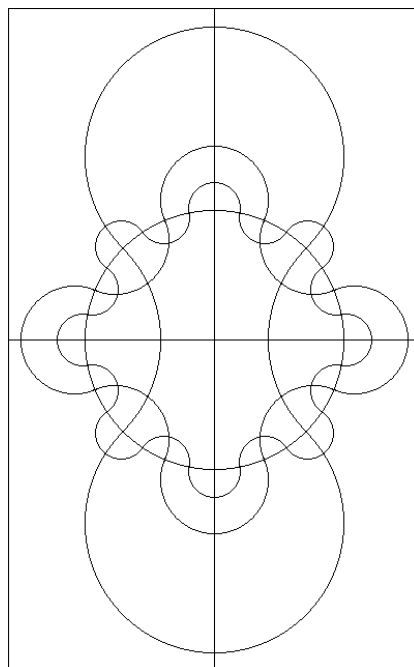


Figure 1. The six-set Edwards–Venn diagram

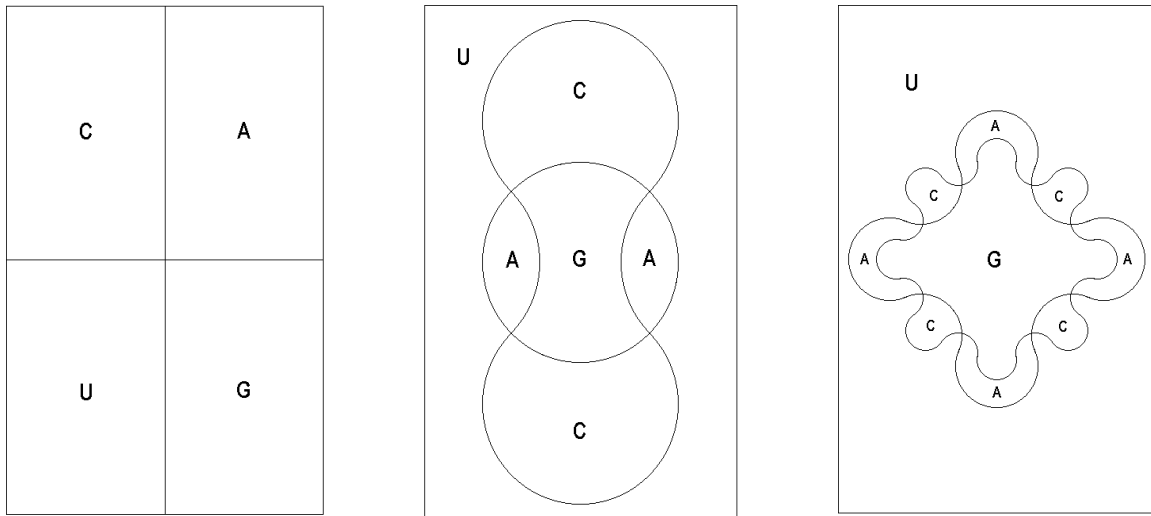


Figure 2. The three separate diagrams labelled with the four bases at the first, second and third codon positions respectively

Finally, colour each triplet's area according to its corresponding amino-acid. The result is a drawing in which nearly all the triplets coding for a given amino-acid form blocks of contiguous areas (*Nature Genetics* cover, September 2007). The exceptions are serine and the stop codon. Two of the serine codons cannot be obtained from the other four by changing just one base, so their separation is correct. As for the stop codon, not all the triplets differing from any particular one by a single base can be made adjacent to it, and the separation of the stop codons is an example of this. It is impossible to achieve exhaustive adjacency in a representation in only two dimensions, but a Venn diagram is known to provide the best solution possible.

It is important not to overlook the fact that not only are triplets coding for the same amino-acid clustered, but that adjacent triplets always differ by precisely one base, and are thus just one mutation apart. This property would be destroyed by, for example, an overenthusiastic rotation of the section of the diagram within the circle by 180° to make the two serine groups adjacent.

One of the characteristics of an ordinary binary Edwards–Venn diagram is that it reveals, and indeed formally corresponds to, a Gray code ordering of the numbers¹. Thus with six sets the binary numbers 1 to 64 are ordered in a cycle in which adjacent numbers always differ in respect of a single binary digit. The same property holds for the 64 codons in the present diagram – they are ordered in a cycle such that each differs from its predecessor by exactly one base. There is more than one Gray code order possible, but the one implicit in the present arrangement (Figure 3) seems best for displaying the genetic code pictorially.

The natural universe for an Edwards–Venn diagram is the surface of a sphere. The representations used here are stereographic projections from the ‘north’ pole; the circle is the equator¹.

1. Edwards, A.W.F. (2004) *Cogwheels of the Mind: The Story of Venn Diagrams* (The Johns Hopkins University Press, Baltimore, 2004).