Swansea talk, July 23, 2006
8th International Symposium on Cytochrome P450 Biodiversity and Biotechnology
University of Swansea, Wales, UK
[slide 1]

I am delighted to be here at the 8th International Conference on P450 Biodiversity. This is my 11th year of attending these meetings, starting with the third one in Woods Hole Massachusetts. Unfortunately I missed the first two in Berlin and Tokyo. These were followed by venues in Strasbourg, Elsinore Denmark, Los Angeles, Awaji Island Japan and now Swansea. Many thanks to the committee and especially to Steve and Diane Kelly and their co-workers for organizing this meeting.

Many of you know that I am a bit obsessed with the P450 superfamily and with understanding its history and evolution. I view my task as similar to assembling a huge jigsaw puzzle, with no boundaries and pieces that do not come with the box. [slide 2] Nevertheless, we now have over 6000 pieces to work with as summarized on this Tree of Life drawing borrowed from the Joint Genome Institute website. These are only the named P450s. There are obviously more that are in the databases waiting to be named. 90% of these sequences are from eukaryotes. The numbers in parentheses are the number of CYP families in each group. The seed plants have 1822 sequences in 64 families for a ratio of 28 genes/family. Animals average 23 genes per family, but the Fungi, protists and bacteria average only 3-4 genes per family with many families having only one sequence. We clearly need more pieces for the puzzle from these groups.

I will divide my talk into three parts. The largest part will focus on chordates and the use of synteny to trace the line of P450 evolution back from mammals to fish and maybe a little beyond. The second part will briefly cover what the moss and Chlamydomonas genomes have revealed about early plant P450 evolution. Finally, I will mention a curious observation seen in some Aspergillus P450s that may open possibilities for P450 engineering not previously considered.

It is my desire, as mentioned in my abstract, to follow present day P450s back to their origin at the beginning of each clan. However , as the sequences diverge over time, it may not be possible to do this by comparing the P450 sequences alone. For example, of the CYP2 subfamilies in fish, only two, CYP2R and CYP2U are clearly recognizable in mammals. Synteny or gene order in the neigborhood around the P450s can identify the orthologous subfamilies and in one case perhaps the origin of a family.

[slide 3] Synteny is highly conserved in mammals as shown here at the synteny viewer at the Phylogenetically Inferred Groups web site. This is the first part of human chromosome 1 compared to mouse. The bars indicate orthologs. Conservation of synteny drops off as we move to greater distances between organisms. [slide 4] This view adds chicken to the mouse-human comparison. You can see that the number of bars is decreasing. If we now compare human with zebrafish [slide 5] you can see a dramatic decrease in the shared synteny. Each color coded block from zebrafish represents a different chromosome or scaffold. The largest syntenic group here is about 3-4 genes,

and they are not staying adjacent to one another. They are beginning to space out. [slide 6] This shows where synteny becomes almost absent in a comparison of human to Ciona or sea squirt. There are only two genes adjacent in both species and you have to look over much of the genome to find other examples like this.

Now let's apply synteny to P450 origins. [slide 7] This first came up for me when I was assembling a group of frog P450s. This was a very large cluster of about 35 genes and it was bordered on one side by the rhag gene for rh blood group antigen glycoprotein. These genes were CYP2s, but it was not clear which subfamily they matched. I did not think much about this until I accidentally saw the rhag gene again while working on rhesus monkey P450s. It was next to a P450 pseudogene called CYP2AC1P. The rhag gene was next to 2AC1P in humans and next to two functional 2AC genes in chicken. On the other side of the 2AC genes is a gene called MUT for methy malonyl CoA mutase. The frog genes are clearly orthologous to CYP2AC1P. The rhag gene is an ammonium transporter that is conserved back to yeast, but no P450s can be found near it in fish or Ciona. The CYP2AC genes do not have good matches in fish. The ancestor of this gene probably arose in the lineage to tetrapods.

The next slide [slide 8] shows the human and chicken 2J genes in red and their neighbor HOOK1 in blue. The fish cluster has 11 genes that include four subfamilies CYP2P, 2V, 2AD and 2N. These are descended from a common ancestor with the mammalian CYP2Js. Since humans have only the one 2J2 gene, the four fish subfamilies probably arose from a single gene after tetrapods separated from fish.

Synteny in Fugu identifies the origin of the CYP2ABFGST cluster. [slide 9] The end gene CYP2S1 in humans is 10 kb from the receptor tyrosine kinase AXL which is next to TGFB1. In frogs, the CYP2Q cluster shares these neighbors. In Fugu, AXL is 9 kb from CYP2Y2. In zebrafish the CYP2Y3 gene is 30 kb from TGFB1. Therefore, the fish CYP2Y genes share an ancestor with the CYP2ABFGST cluster, once again probably a single gene.

Occasionally genes can move and this causes some problems with synteny analysis. [slide 10] In this example the CYP2AB1 gene in frogs is bracketed by FOXD4L2 and CBWD3 however, there is no P450 in humans or frogs at that location. The CYP2AB1 gene does exist in humans as a pseudogene [slide 11]. It is at this location on chr 3, also seen at the orthologous position in chickens. There is no related gene in fish, and all five neighboring genes do not have P450 neighbors in fish. The 2AB1 genes are probably tetrapod specific genes.

Conservation of synteny can be lost in some species while it is retained in others. It is always good to use all available information to make an assignment. This example shows [slide 12] synteny is lost on the left side of the CYP2K genes in zebrafish, but not in fugu or tetraodon. If one only looked in zebrafish, the relationship between the 2K gene cluster in fish and the 2W genes in tetrapods would go undetected.

In addition to gene migration and sequence divergence, genes may be lost and this is a factor in P450 evolution too. [slide 13] This slide shows the location of the CYP2X genes in zebrafish. Notice that they are on two different chromosomes. Chr 25 retains the left side syntenic genes SETD6 and CNOT1. Chr 7 has GOT2 and an numbered neighbor. In tetrapods, these two regions are together on one chromosome. The original CYP2X gene cluster got split, creating two clusters. The CYP2X ancestor was apparently lost on the way to tetrapods since it is not found in frogs, birds or mammals.

These examples all involve subfamilies, but what about families. CYP families have emerged over time. Can synteny address the origin of a family? Steroids and steroid receptors do not exist in Ciona, but they are found in vertebrates. The CYP19 gene is seen in Amphioxus but not in Ciona. [slide 14] In humans and fish CYP19 is neighbor to gliomedin. These genes face away from each other about 13 kb apart in fish. BLAST searches of Ciona with human gliomedin find no good matches. Searches with the fish ortholog show a best hit to a sequence that is 10 kb from the Ciona CYP20 gene. These genes also face away from each other. This suggests that CYP19 arose by a gene duplication of CYP20. In this case the orthologous position became CYP19 and the new location of the CYP20 gene retained the original function. CYP20 may have a substrate that is similar to testosterone.

We have been talking about local gene neighborhoods on a scale called microsynteny. Another driving force behind vertebrate evolution is whole genome duplication. The 2R hypothesis states that the ancestor of all vertebrates underwent two rounds of whole genome duplication. This has recently been confirmed by statistical analysis of paralogous segments in four chordate genomes by Dehal and Boore. This opened the possibility to detect P450 genes duplicated during these whole genome duplication events. We have to look at macrosynteny over millions of bp to see this. Surprisingly, only four vertebrate P450s seem to be strong candidates. [slide 15] CYP7A1 and CYP8A1 are on paralogous segments identified by Dehal and Boore. There is no CYP8 found in Amphioxus, a species that split from vertebrates before the 2R duplications. Here we see CYP8A1 and the two CYP7 sequences in humans. The regions are large, 3 Mb and 16 Mb. Note the five paralogs indicated by the arrows. This supports the idea that CYP8A is paralogous to CYP7A via whole genome duplication.

The other pair of CYPs that reside on paralogous segments are CYP26B and the CYP26A/CYP26C pair. The CYP26A/CYP26C genes are only 3 kb apart. Amphioxus has two CYP26 related genes. They are 65% identical to each other and so they are both from the same subfamily. These genes are most like CYP26C. Therefore, CYP26B may have been created in a whole genome duplication from a 26C precursor and 26A came last.

This gives you a taste for the usefulness of synteny in constructing P450 genealogies. It also should warn you that it is very hard to move backward from fish to Ciona and beyond. The same approach should work in insects and plants. Now I want to change topics to plants and discuss the progression of P450s in the transition from green algae to moss and then on the vascular plants. The genome sequences are now available from

Chlamydomonas reinhardtii (a green alga) and Physcomitrella patens (a moss). Though Arabidopsis and rice have hundreds of P450 genes, Chlamydomonas has only 39 and moss has 71. [slide 16] The moss sequences are shown in green on this tree, which I will refer to to make some general points, then we will look at subsections of the tree in more detail. You do not need to read the lables on this figure.

The first items to notice are the colored ovals. These mark the 11 plant P450 clans. All named plant P450s fit in these 11 clans. The blue ovals are for ancient P450 families that exist in algae and land plants and so they predate emergence onto the land. These are CYP51, CYP710, CYP97 and CYP746. Dr. Ohta will talk about the function of CYP710 later this meeting. The next group of yellow ovals are the clans that have members in algae, but with no CYP families shared between algae and the land plants. These clans have genes important for early adaptation to life on land. They include the CYP72 clan, the CYP85 clan and the CYP711 clan. The green ovals are clans that are only found in land plants with no obvious members in algae. They are more recent in origin than the genes in the yellow or blue clans. These include the CYP71 clan, the CYP74 clan, the CYP86 clan, and the CYP727 clan.

[slide 17] This is the top portion of the tree. CYP711A1 has been identified as MAX1 in Arabidopsis. It makes a carotenoid-derived branch inhibiting hormone. It is not seen in moss, but there are clan members in algae, so perhaps its ancestor was lost in this moss. The unusual property of the CYP711 family is that it is more animal like than plant like. After matches to other CYP711 sequences, the best BLAST hits are to CYP5 and CYP3 members. Two families in Chlamydomonas that are in the CYP711 clan also have this property. CYP743 and CYP744 show the top 100 BLAST hits are all animal sequences in the CYP5 and CYP3 families. If CYP4 sequences are included in a tree like this one , the CYP4 sequences cluster with the CYP711 clan sequences. Therefore, the CYP711 clan may be the plant equivalent of the animal CYP3/CYP4 clans and they may share a common ancestor that predated the plant-animal divergence.

The CYP74 family includes allene oxide synthase, hydroperoxide lyase and divinyl ether synthases. Though this family has the longest branch on the tree it is not the oldest family, since it is not found in algae. Richard Hughes will talk about CYP74s in the protein structure section. The CYP85 clan includes the brassinosteroid P450s CYP85 and CYP90. These specific families are not in moss but a related family CYP763 is present. A clear CYP90 fragment can be found in ferns, so the brassinosteroid pathways probably developed after moss, on the way to vascular plants. Gerard Bishop will talk about these pathways.

The CYP710 family is another ancient P450 with function identified as a C-22 sterol desaturase. This is the same function as the CYP61 P450 in fungi. These are pretty clearly in the same clan that predated the plant-fungi divergence. A CYP61/CYP710 clan member is not known from animals, so it may have been lost in this kingdom. This would change the character of animal sterols compared to plant and fungal sterols.

[slide 18] The middle of the tree is exclusively the CYP71 clan. This is a land plant clan, since there are no members in algae. These gene products are likely involved in adaptation to life on land and in secondary metabolite synthesis and degradation. There are five families shared between moss and higher plants (arrows). These would represent members of early biochemical pathways emerging in land plants. CYP73 and CYP98 are members of the phenylpropanoid pathway, leading to the formation of G-lignols, but CYP84 is missing in moss so S-lignols cannot be made. This is an example of an intermediate stage in pathway evolution, in this case for lignin biosynthesis. A similar partial pathway is present for gibberellin. CYP701 the ent-kaurene oxidase is present in moss, but CYP88 the kaurenoic acid oxidase is missing. This means that moss has evolved part way to forming GA_{12} but not the whole way.

[slide 19] This is the bottom part of the tree. It contains the CYP86 clan important for cuticle formation. The blue ovals are the CYP97 family, which includes multiple carotenoid hydroxylases and CYP746, a family only observed in moss, Chlamydomonas and Volvox. CYP746 and CYP727 are the only clans without any known function. The blue clans tell us what was critical and fundamental to the green algae before land colonization. This is sterol synthesis with CYP51 and CYP710 in the pathway to the basic plant sterols, and carotenoid metabolism. All other plant P450s presumably derived from these four blue clans.

CYP51 and CYP61/710 predate the plant-fungi divergence because fungi have both of these clans. That also means they predate the plant-animal divergence. CYP97 may have its origins in cyanobacteria, since plants acquired cyanobacteria to conduct photosynthesis and carotenoids are intimately involved in this process. Cyanobacteria contain several P450s that are eukaryote-like (CYP110 and CYP120). Drs. Sligar and Schuler have shown that CYP120 has the ability to bind retinoic acid with high affinity. BLAST searches with CYP97 sequences against Genbank show all the best hits are clear CYP97 sequences followed by bacterial sequences. No other eukaryote P450s match CYP97 better than other CYP97s and some bacterial P450s. The best bacterial hit is to Myxococcus xanthus, a carotenoid producing bacterium at an expect value of e-60 (34% identity). The CYP97 clan is most similar to the CYP72 clan. CYP102A1, the bacterial fatty acid hydroylase clusters with the CYP72 clan on this tree. This reinforces the possible bacterial origin of the CYP97/CYP72 cluster.

This brings us to the last short section of my talk. [slide 20] I show this picture of a man standing on his head in an unlikely position because I want to propose a P450 that is standing on its head. When annotating several fungal genomes this past April, I came across these five CYPs from Aspergillus species. [slide 21] Note that there is no significant heme signature region, even though the EXXR and PERF motifs are present. This was very odd so I looked at the alignment in more detail. This is what I found at the I-helix region. [slide 22] Instead of the typical Thr residue there was a Cys. Considering the usual geometry of the Thr at this site, pointing at the heme iron, I am speculating that the Cys may be liganded to the iron and the P450 may be an inverted structure with the Cys on the other side of the heme. Nature it seems is very adventurous. To ponder that ,

I leave you with this quote from the famous evolutionist and actress Mae West. [slide 23]
“I’ll try anything once, twice if I like it, three times to make sure.”