This section was removed from the Chalmydomonas genome paper in the final edit to save pages in the supplemental materials section.

S14. Cytochrome P450 genes:

The *CYP* genes represent an interesting family of genes which encode the cytochrome P450s, in Chlamy. There are 246 full-length *CYP* sequences in Arabidopsis, 312 in Populus, 356 predicted in rice and 71 in the moss *Physcomitrella patens*. Plants generally have a large numbers of *CYP* genes, and were anticipated to have more than the algae since they have evolved more complex structures and biochemical pathways. The sequence of the Chlamy genome provides an opportunity to explore the origins of some of the CYP sequences in plants. Chlamy has 39 *CYP* genes and two pseudogenes in 21 different *CYP* families. Some of these genes have not at this point been completely assembled since the encoded proteins are dissimilar to known cytochrome P450 sequences and there are no EST sequences to help with the assembly. The CYP85 clan is critical in land plants because it includes CYPs that are required to make/metabolize brassinosteroids and gibberellins. While Chlamy does not have the CYP85, CYP90 (for brassinsoteroid synthesis), or CYP88 (for kaurenoic acid oxidase synthesis) families, it does have the CYP85 clan sequences CYP737-CYP740, along with CYP739, in a six member cluster on a 40 kb block of scaffold 8. While it is unclear what the substrates are for these CYP85 clan members, sterols remain a strong possibility. Perhaps the most surprising finding in the analysis of the Chlamy CYPs is that only five known CYP families were detected and only four of these are present in other plants. The other 15 CYP families may be unique to Chlamy, although preliminary analyses suggest that close orthologs for most exist in Volvox. Furthermore, there are two groups of Chlamy sequences that are more animal-like than plant-like; they are related to the animal CYP4, CYP5 (thromboxane A2 synthase, only found in vertebrates) and CYP3 clans.

The *CYP* genes are named according to their sequence relatedness. Sequences with greater than 40% amino acid identity are placed in the same family, with exceptions for clear orthologs. The family designation does not identify distant relationships between CYP sequences {Nelson, 1996 #303}. CYP families also cluster together in gene clades. These have been called CYP clans in the P450 nomenclature system. Spermatophytes currently have 63 CYP families in 10 CYP clans. When the 71 moss genome CYPs were identified, there were 16 new families added, and one new clan CYP746 {Nelson, 2006 #265}. Chlamy contains the following seven plant clans: CYP51, CYP72, CYP85, CYP97, CYP710, CYP711, CYP746. It is missing four clans: CYP71, CYP74, CYP86, CYP727. CYP71 is the dominant land plant CYP clan. Even in moss, 41 genes are in the CYP71 clan. This represents a development unique to the land plants. The CYP74 clan includes allene oxide synthase and related sequences such as hydroperoxide lyases and divinyl ether synthases{Howe, 2002 #299}. The absence of these genes in Chlamy suggests that the jasmonate pathway evolved in land plants following the divergence of Chlamy. The CYP86 clan includes some P450s with fatty acid omega hydroxylation activity {Duan, 2005 #298}. The products of these reactions may act as precursors for cutins, the molecule that forms a water barrier on land plants. CYP727 is not found in Arabidopsis, so it is not essential vascular plant survival. CYP751 in moss is in the CYP727 clan, suggesting that this clan is older than the CYP727 family.

CYP51 is a sterol biosynthetic cytochrome P450. It is found in all eukaryotes that do not obtain their sterols strictly from their diet. Some animals and some parasites have lost the *CYP51* gene, but all plants must have this enzyme. The second family shared with plants is CYP97. This family has three subfamilies in all plants and all three subfamiles are encoded on the Chlamy genome. CYP97C1 is an epsilon ring carotenoid hydroxylase that is part of the lutein pathway from alpha carotene in Arabidopsis {Tian, 2004 #304}. CYP97A3 is a beta ring hydroxylase{Kim, 2006 #300}. CYP97B is of unknown function, although it may, also function in carotenoid hydroxylation/oxidation. CYP97 is also found in diatoms, but diatoms contain a plastid originally obtained by secondary endosymbiosis of a red alga, so this endosymbiont may be the source of the CYP97 in diatoms. Each plant species sequenced so far has only one member in each CYP97 subfamily. However, Chlamy has two CYP97A members.

The third family shared with plants is CYP710. CYP710 is related to CYP61 in fungi, a C-22 sterol desaturase that acts after CYP51 in ergosterol biosynthesis. Fungal CYP61 and CYP710B1 do not share any intron boundaries. CYP61 and CYP51 are the only cytochrome P450s found in *Schizosaccharomyces pombe* and they are found in all other fungi indicating a putative mandatory role. There is only one CYP710B1 sequence in Chlamy. Other plants have multiple CYP710s, although Populus has only one, so this CYP may have a fundamental function. CYP710A1 has been identified as a C-22 desaturase for beta-sitosterol in Arabidopsis {Morikawa, 2006 #302}. The red alga *C. merolae* has only five CYP genes and two of them are CYP710 sequences (nearly identical to each other). Red algae also have a CYP51 and two other CYPs that do not match Chlamy CYPs. Red algae lack the *CYP97* genes associated with carotenoid ring hydroxylations. The fourth family, CYP746, is found in moss, but not other plants, and has no known function. The fifth shared family is CYP55, which is of fungal origin. The CYP55 protein is a nitrite/nitrate reductase that is a soluble in fungi (P450nor) {Kizawa, 1991 #301}. All other eukaryotic CYPs are membrane bound, while bacterial CYPs are soluble. The CYP55 family is closely related to the bacterial CYP105 family and almost certainly was acquired by fungi from bacteria. The origin of the Chlamy CYP55 is not clear, however the Chlamy *CYP55B1* gene does share one intron boundary with the *Neurospora crassa CYP55A6* gene, suggesting loss in other lineages.

Moss has 10 of the 11 plant CYP clans. The one missing clan (CYP711) appears to be present in Chlamy, so Physcomitrella probably lost that sequence. The CYP711 clan may be present in other moss species since it was transmitted to angiosperms. CYP711A1 (MAX1) in Arabidopsis has been shown to produce a carotenoid-derived branch-inhibiting hormone. {Booker, 2005 #275}. The CYP97A and CYP97C gene products also use a carotenoid substrate. The two families in the CYP711 clan in Chlamy are CYP743 and CYP744. These two families have 12 members and one pseudogene. They make up one third of all cytochrome P450s in Chlamy. If they metabolize carotenoids, there would be a large number of carotenoid derivatives in Chlamy. There are two groups of Chlamy sequences that are more animal-like than plant-like. The sequences (CYP741,

CYP742, CYP748, CYP767, CYP768, CYP769) are most like CYP4 clan members of animals. When several CYP4 clan members were included in a phylogenetic tree, these Chlamy sequences did not cluster with CYP4 members. Instead, the CYP4 clan sequences clustered with CYP711, CYP743 and CYP744. Blast searches with CYP711A1 show the best hits are all CYP711 members in plants followed by CYP5 sequences in animals. CYP5 is thromboxane A2 synthase, only found in vertebrates. It is a member of the CYP3 clan. CYP744A1 has best matches in the nr section of Genbank to CYP5 followed by CYP3 sequences. The top 100 BLAST hits were all animal sequences. CYP743 had best BLAST hits to CYP3 sequences of animals; once again the top 100 BLAST hits were all animal sequences. CYP743 and CYP744 sequences cluster with CYP711 on trees with plant sequences. They appear to be CYP711 clan members. The CYP711 clan may share a common ancestor with the CYP3/CYP4 clans in animals. This makes the original CYP711 clan member an ancient conserved sequence like CYP51 and CYP61/CYP710.