Figure Legends

Figure 1. Schematic of Protein Structures, the Retinal Determination Network and a Drosophila Species Tree.

(A) RD proteins. NT: N terminal, CT: C terminal, B: Central linker, PD: Paired domain, HD: Homeodomain, SD: SIX domain, Zn(1-4): Zn finger (1-4), PSQ: Pipsqueak domain, P/S/T: Pst domain, Eya1 and Eya2: Eya 1 and Eya2 Domain; DD1 and DD2; Dac1 and Dac2 domains (B) RD network. Purple connectors indicate genetic interactions; Orange arrows indicate confirmed protein-protein interactions; Blue arrows indicate confirmed transcription factor binding; Green arrows show connections to developmental processes. The X downstream of the Notch pathway shows an as yet unidentified molecule. (C) Phylogenetic tree of the species used in this study.

Figure 2. Divergence of Retinal Determination Genes within Drosophila.

Branch lengths were calculated using distance-based trees generated from *Drosophilid* nucleotide sequences using *Tribolium castaneum* as most recent common ancestor (MRCA). While all the genes have different divergence rates, the genes are not evolving rapidly in any particular lineage.

Figure 3. Phylogenetic Analysis of Paralog Pairs within the Retinal Determination Network.

(A) Ey/Toy. (B) Eyg/Toe. (C) Tsh/Tio. (D) Dan/Danr. (E) So/Optix/DSix4. All duplicate genes are evolving at significantly different rates compared to their sister gene, except for Eyg/Toe and Dan/Danr.

Figure 4. Selective Pressures on Functional Domains and Non-Conserved Regions.

The highly structured regions of the genes in RD cascade (Paired, Homeo, Paired, Zinc1-4, Pipsqueak) are under significantly higher purifying selection than the non-conserved regions (N-terminal, B linker and C- terminal). The error bars indicate standard errors. This is likely to allow the genes to remain connected in a highly regulated network through the DNA-binding and protein-interaction domains, while the N-terminal, B linker and C- terminal segments accumulate mutations and gain new functions.

Figure 4. Selection Signatures of Functional Domains and Non-Conserved Regions within the Retinal Determination Network.

(A) Full length genes. (B,C) Ey/Toy. (D,E) Eyg/Toe. (F,G) Tsh/Tio. (H-J) So/Optix/DSix4 (K,L) Dan/Danr. All genes are under varying degrees of purifying selection. The duplicate genes have different patterns of selection across their coding regions, with the highly structured regions being more constrained than the non-structured regions.

Table 1. d_n/d_s Raw Values (full-length, functional domains, non-conserved segments)

The raw d_n/d_s values obtained for the full-length genes within the melanogaster group as well as those obtained for individual functional domains and non-conserved regions.

Table 2. Number of clade-specific and group-specific changes in residues.

The paralog pairs show some number of residue changes even within the highly structured domains. This is likely to account for the changes in gene regulation and protein binding that we see between the paralog pairs.

Supplementary Figure 1. Higher resolution analysis of Pax6 and Pax6(5a) family C terminals.

Sliding window measurements of d_N/d_S in non-overlapping 100bp intervals across the C terminal domains of ey/toy and eyg/toe. The horizontal dotted line demarcates the average d_N/d_S value across the segment in question.

Supplementary Figure 2. Amino acid alignments of conserved motifs between paralog pairs.

Note that the residues marked in red are clade specific, the residues marked in blue are both clade specific and group specific, and the residues in green are group specific. Note that for So, Optix and

Dsix4, residues in purple indicate that two of the three SIX family genes contain this residue. (a) Ey and Toy homeodomains. (b) Ey and Toy paired domains. (c) Eyg and Toe homeodomains. (d) Eyg and Toe paired domains. (e) So, Optix and DSix4 homeodomains. (f) So, Optix and DSix4 SIX domains. (g) Tsh and Tio Zinc finger domains. (h) Dan and Danr pipsqueak domains.