Population Genetics on Gene Regulation and Gene Duplication
Gene expression in balanced state
Gene duplication: perturb the balance
Fitness effect:
  Difference between Regulatory proteins and Structural proteins

Examples: Transcription factors vs olfactory receptors or immunoglobulins
Cis-regulatory elements: Turnover
even in constant expression
within allowed latitude

Example  Ludwig et al. 2000
Stripe2 elements of even skipped
\textit{D. melanogaster vs D.pseudoobscura}
Genomics on expanding gene families
Population genetics: Not simple processes of mutant substitutions

Gene duplication (segmental or transpositional) and gene conversion continuously occur
Examples of gene conversion:
mostly structural proteins

MHC, immunoglobulins,
proteases and their inhibitors

Evolution is characterized by interaction of selection
and gene conversion.
Positive selection on reactive center
Promoter regions:
Selection against spurious binding sites
( Hahn et al. 2003)

Binding sites: underrepresented in genomes of Bacteria and Archaea
Very weakly selected against
Adaptive evolution of transcription factor binding site

Fitness: Function of binding interaction between transcription factors and DNA

Mutation-selection-drift equilibrium
CRP site sequences in E. Coli
mean Ns < 1 0

Mustonen- Laessig 2005
Coevolution between binding sites and transcription factors

Selective constraints at various levels
Effect on the rate of gene duplication in evolution
Drift and selection

New duplications: mostly deleterious

Rare beneficial duplications, and considerable fraction of neutral and nearly neutral duplications
Protein evolution

dN/dS

Human > pig > rodent
    Jorgensen et al. 2005

Human > dog > mouse
    Lindblat-Toh et al. 2005

Hominid > rodent
    The Chimpanzee Sequencing Consortium 2005

Yeast lab. Strains > yeast wild strains
    Gu et al. 2005
Polymorphism

Human SNPs: Abundant rare alleles

Bacteria 84 species  Tajima’s D

Dnon 68% negative
Dsyn 70% positive

Hughes 2005
McDonald-Kreitman test
Sawyer-Hartl test

Human and chimpanzee amino acid change
Positive selection 9.0%
Weak negative selection 13.5%
Range of 2Ns: -10 ~ +10 (peak at 0)

Bustamante et al. 2005
Ratio of gene expression divergence between species to gene expression diversity within species human-chimpanzee

About equal in the brain, heart, kidney, liver but threefold higher in testes
Positive selection in testes, mainly on X chromosome

Brain: Ratio of the change of the human lineage to that of the chimpanzee lineage is larger than the same ratio in the liver or heart
Indication of positive selection in the human lineage

Khaitovich et al. 2006
Lynch-Force 2000
Duplication-Degeneration-Complementation
DDC model: Drift

Our interest
Subfunctionalization: Mainly regulatory proteins
Robustness of networks

Neofunctionalization: Mainly structural proteins
Phenotype Complexity

Cooption-Duplication-Divergence
of mainly regulatory protein genes

Cooption with and without gene duplication

Restriction on no. of binding sites of transcription factors in a regulatory region of one gene?

Later divergence of duplicated genes results in more complexity
Transposable elements: 45% of human genome

Majority: selfish
Minority: gene regulation and other functions
Kidwell-Lisch 2001

Life cycle (DNA mediated type)
  invasive phase
  stage of maturity (stable)
  stage of senescence

Broad spectrum on fitness
  Host recruitments: Regulatory functions
    Enzyme functions
  Restructuring of genome: recombination among elements
Selfish DNA

\[ s < \frac{1}{1+\log_2(1-p)} \]

Outbreeding is necessary.

D. A. Hickey 1982
Regulatory sequences from transposable elements
Jordan et al. 2003

24% of promoter regions of human genome contain TE derived sequences
8% of total nucleotides in promoter regions are LINES and SINES

They correspond to 0.2% of the total genome and are much less than 45%
Micro RNAs derived from genomic repeats
Smalheiser et al. 2005

11 out of 220 miRNA precursors:
   LINES, SINEs, tRNAs and mariner reps.

Driving force to create new miRNAs in mammals
MicroRNA and developmental canalization
Hornstein and Shomron 2006

Buffer activity of genes: stability of phenotype and accumulation of cryptic variation

Conserved miRNA-binding sequences in more than 1/3 of genes

Network motifs like FFL: canalization and robustness
### Genome complexity and Drift

Lynch and Cornery 2003

<table>
<thead>
<tr>
<th></th>
<th>Pop. Size</th>
<th>Genome size</th>
<th>Genome complexity</th>
<th>Mobile elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prokaryotes</td>
<td>large</td>
<td>small</td>
<td>less</td>
<td>less</td>
</tr>
<tr>
<td>Multicellular</td>
<td>small</td>
<td>large</td>
<td>more</td>
<td>more</td>
</tr>
<tr>
<td>eukaryotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Complexity including mobile elements:

Drift is the initial force
Changing fitness landscape: ecology

Genetic backgrounds
Environments
Epigenetics
Joint influence of epistasis and plasticity on mutational effects
Remold-Lenski 2004

Plasticity only 0/18
(environments)

Epistasis only 3/18
(genetic backgrounds)

Plasticity and epistasis 6/18

Neither 9/18

Environments: glucose and maltose
Genetic: glucose-evolved and maltose-evolved
Fitness landscapes have enormous dimensionality

Nearly neutral networks expand through genotype space

Other genotypes are “holes”

Existence of multiple high fitness combinations of genes

Importance of mutational order and random drift (chance and contingency)

Gavrilets 2004
Rapid evolution in small populations

Nature of nearly neutral mutations

Continuity among beneficial, slightly advantageous and slightly deleterious classes