

Gene Families

1. Gene Duplication
2. Divergence
3. Concerted Evolution
4. Survey of Gene Families

Gene Duplication

A second copy of a gene could arise in several ways:

- Polyploidization—doubling of the whole genome
- Chromosome duplication
- Translocation—gene moves to another location, then the old and new form come together
- Unequal crossing over (recombination)
- Replication slippage
- Insertion of a virus or transposon carrying the gene
- Reverse transcription

Gene Duplication

Once a gene is duplicated, what can happen?

There are now four copies in a diploid, so recessive mutations are thoroughly hidden.

-----A----- -----A-----
-----A----- -----a-----

No matter how sexuality reassorts these, no offspring without a big A are produced. So the recessive is completely safe.

This implies that once there are multiple copies, harmful recessives will tend to creep in, protected by the extra good copies.

Gene Duplication

Consequences of gene duplication:

- One copy may deteriorate and be lost
- One copy may evolve a new function
- Two copies may specialize different parts of old function

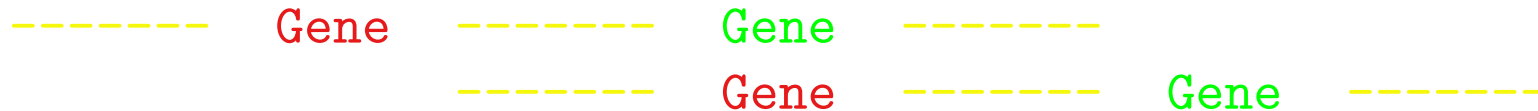
Red-Green Color Vision–Splitting the Old Function

The red-vision and green-vision genes of humans are extremely similar and are believed to be a gene duplication. However, they have specialized on slightly different wavelengths, giving humans more complete color vision than we could have with only one gene.

----- Gene ----- Gene -----

Red-Green Color Vision–Splitting the Old Function

This particular gene family is unstable, because R and G are next to each other. It is easy for a faulty recombination to produce chromosomes with an R/G hybrid gene.



A crossover can produce either of:



Since R and G are on the X chromosome, male humans are haploid for them. This explains the high frequency of R/G colorblindness in men.

HLA examples

HLA region has many gene duplications

- HLA-AR locus—apparent duplication of HLA-A
 - Shows signs of past selection
 - High variability in antigen-binding region
 - Multiple stop codons and other harmful-looking mutations
 - Apparent produces no protein
- HLA-A, -B, and -C
 - Shows signs of past selection
 - High variability in antigen-binding region
 - Clean sequences without stop codons
 - Different expression levels in different tissues
 - Associated with resistance to different diseases
- HLA-C is less strongly expressed and associated with few diseases; it may be on the borderline between useful and redundant

Gene Duplication

What determines whether a duplicated gene will deteriorate or take on a new function?

If the new gene was broken to start with (as reverse transcribed genes often are) its chances are much worse.

Mainly, though, it's luck—does a mutation to a helpful new function occur, and survive drift, before harmful mutations knock one copy out?

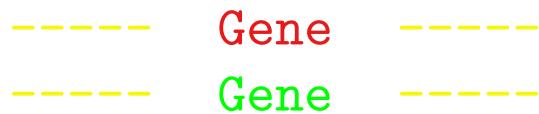
Gene Duplication

The great evolutionary importance of gene duplication is that it allows an organism to evolve a new function without losing the old one.

It also allows a favorable resolution of overdominance. Instead of having two alleles which need to be present together in the heterozygote, you can have two fixed genes, each with a different allele, and everyone can be “heterozygous”. This is the usual explanation for the multiple MHC genes in most mammals.

Gene Duplication

In some primates, there is only one locus for red/green color vision, and red and green are alleles. A female can have human-like color vision:



but many females are colorblind homozygotes, and all males are colorblind.

(Study question: what are the optimal allele frequencies to maximize color-vision females?)

It is easy to see how the human system might be evolutionarily favored, since it produces far fewer colorblind individuals (none, except when illegitimate recombination has messed the genes up).

Concerted Evolution

So far we have considered gene families which evolve away from each other—either one copy dies, or they evolve two separate functions.

However, in some cases unequal crossing over or gene conversion can cause all the members of a family to evolve together. This is called **concerted evolution**.

It is most common when the genes are clustered together on a chromosome, often as a **tandem array** or long series of head-to-tail gene copies.

Concerted Evolution

In a tandem array, unequal crossing over can change the number of copies very easily:

-----gene---gene---gene---gene-----

-----gene---gene---gene---gene-----

leads to:

-----gene---gene---gene---gene---gene-----

-----gene---gene---gene-----

Concerted Evolution

If we show the same picture, but label the genes, we can see that some copies get duplicated and others deleted:

-----gene1---gene2---gene3---gene4-----

-----gene1---gene2---gene3---gene4-----

leads to:

-----gene1---gene2---gene2---gene3---gene4-----

-----gene1---gene3---gene4-----

Concerted Evolution

Repeating this process many times will not only cause the gene family to change size up or down, it will make the copies more similar to each other.

An even stronger force making the copies similar is gene conversion— copying of a small piece from one chromosome to the other.

-----gene1---gene2---gene3---gene4-----

-----gene1---gene2---gene3---gene4-----

leads to:

-----gene1---gene1---gene3---gene4-----

Concerted Evolution

Many of the huge gene families appear to evolve in this way—recombination and conversion make them similar more rapidly than mutation makes them different, so all the copies remain very similar.

This is easier if there is selection for similarity. For example, many organisms have huge numbers of rRNA genes in tandem arrays. At certain stages in the life cycle having many working copies allows very fast synthesis of rRNA, so there is some selection to keep all copies working, and this may help discourage divergent evolution.

Concerted Evolution

The human R/G genes seem to be just at the breakpoint between divergent and concerted evolution.

They are close together and very similar, so the concerted force making them become the same is fairly strong.

But selection probably favors people who are not colorblind, so there is also selection to keep them different.

If they were on separate chromosomes, divergence would probably win—unequal crossing over on different chromosomes is so harmful, such changes would usually not survive.

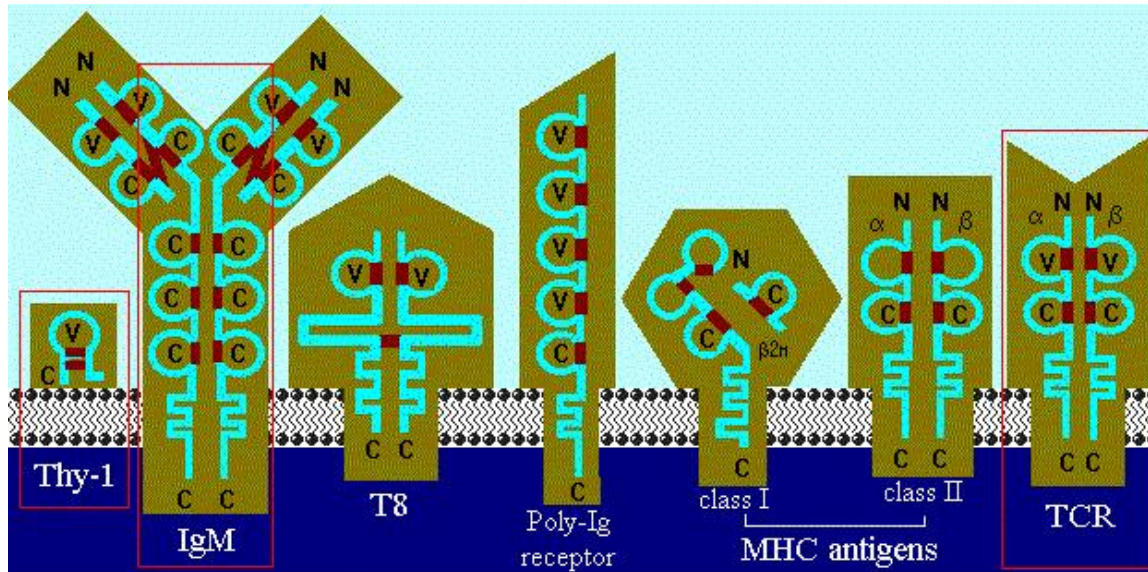
Multi-gene Families

Most genomes are full of multi-gene families of all sizes, from two up to thousands or more.

Some notable human examples:

- Alu transposable elements, tens of thousands of copies.
- HLA loci, around a dozen
- Antibody variable-region loci, dozens to hundreds
- R/G color vision loci, two or three

Multi-gene Families



We can also recognize “superfamilies”, families of families, such as the immunoglobulin genes. They all share part of their structure, but other parts appear unrelated.

Exon shuffling

New functions can arise by combining parts of old genes.

Often this takes the form of attaching a working subunit (domain) from one gene to a working subunit from another.

Introns fairly often fall between domains, rather than in the middle. In other words, exons are often domains.

This could help encourage useful gene-mixing, and some think this is an evolutionary reason for exons. Walter Gilbert has done a lot of work on this.

Exon shuffling

- Pro:
 - Many existing genes seem to be made up of shuffled exons
 - The reading frame of exons is non-random, making it more likely that two random exons will add up to a working protein
 - The very old introns (found throughout a wide range of organisms) are particularly likely to be in reading frame 0.

Exon shuffling

- Con:
 - Organisms with few or no introns seem to evolve successfully
 - The selection suggested seems terribly weak—it's selection against an organism just because it has lost the ability to have a potentially useful mutation in the future. This may not be strong enough to keep the introns.

Introns might be evolutionarily useful by separating exons for reshuffling even if this is not the force that keeps them around.