The difference between "homeobox" & "Hox" genes

Jennifer Harrison Aug 13, 2020

https://geneticjen.medium.com/the-difference-between-homeobox-and-hox-genes-e73d7926eca1

1

Update: Here are some facts to demonstrate the extent of this issue in higher education. Since publishing this post it has become the 2nd most popular result on Google when searching "homeobox vs hox", second only to the Wikipedia entry for homeobox genes. The post is currently recommended to biology students by educational awarding bodies like the Oxford Cambridge and RSA (OCR). Years later, I still get emails and social media messages from grateful students about this piece. They tell me they're learning something here that they're expected to understand but aren't properly taught by their professors or learning materials. I'm glad it's helping people. I've switched blogging websites several times, reposting this piece and requiring links to be updated. It's possible I'll move again. However, if I do move to another platform, I will leave this version here on Medium so people can always link to it. Thanks.

You're here because you want to know what the difference is between homeobox and Hox genes. Let's get straight to business: the terms "homeobox" and "Hox" are *not* interchangeable; they mean different things. It's correct to say that the lancelet Amphioxus (*Branchiostoma lanceolatum*) has 15 Hox genes but it's equally correct to say it has over 130 homeobox genes.

This is a confusing topic for biology students because almost all trustworthy sources in their education will use the terms interchangeably. Professors, science communicators, biology documentaries, professional publications, popular science YouTubers, and bloggers will either switch between the terms or present one as a shortened version of the other.

If you're a student you may have watched science videos by YouTubers like Hank Green and seen the "Homeobox (Hox)" terminology that suggests one term is shorthand for the other. This otherwise decent glossary at Epigenesys manages to dump the terms homeotic, homeobox, and Hox into one single paragraph and glossary entry, which is of little help to a confused student seeking clarity. The top Google result for "homeodomain" (ignoring Wikipedia) is R&D Systems saying, "The DNA sequence that encodes the homeodomain is called the homeobox and homeobox-containing genes are known as hox genes". This is wrong. A homeobox-containing gene is not necessarily a Hox gene. Let's clear this up.

The homeobox: It makes sense to define the terms properly before we discuss where the confusion comes from. In the 1960s, scientists discovered there are some genes that share a highly conserved region of DNA we now call the homeobox. Any genes that contain the homeobox region in their DNA are homeobox genes.

When I say very conserved, I really mean it. *You* have homeobox genes, the *birds* outside do, the *grass* on your lawn does… even *yeast* does. A seagull and a blade of grass are very different organisms yet they have plenty of genes that share the homeobox, which has barely changed over eons. The origin of homeobox genes is truly ancient, pre-dating the origin of animals.

The 180-base-pair homeobox codes for a 60-residue chain known as the homeobox domain (or homeodomain). In plain English: the region of the gene is known as a homeobox and the region of the resulting protein is the homeodomain.

The explanation for *why* it is so conserved across organisms, through hundreds of millions of years of independent evolution, is that its function restricts its evolution. The homeobox domain binds DNA (or RNA), allowing a protein with a homeodomain to act in gene regulation. For example, these proteins can be used to turn genes on and off during development. It's an invention of evolution that has persisted through the origin of the fungi, plants, and us animals, without the homeobox itself changing much at all. If the homeobox region changed too much, the resulting homeodomain of the protein wouldn't be able to bind to DNA and RNA, which would be catastrophic for the organism.

To summarise, a homeobox gene isn't one *specific* gene; it's a huge and ancient group of genes that all contain the homeobox, a region of DNA that codes for a domain which can bind to DNA. It allows genes to result in proteins that can bind to other genes, which is essential in development.

Hox genes: Every Hox gene *is* a homeobox gene, but not every homeobox gene is a Hox gene. The homeobox genes have diversified so much through evolutionary history that there are now distinct classes of them and the Hox genes are the most famous family of homeobox genes. The homeobox itself might be highly conserved, but the rest of DNA in homeobox-containing genes can have more freedom to evolve independently.

In us bilaterian animals, one of the main roles of the Hox genes is to specify anteroposterior identity to your body. It's a complicated system but a simplified explanation would be that the Hox genes play a role in determining which body parts grow and where on the body. By messing with their DNA, you can cause limbs to grow in the wrong places.

However, there are plenty of other non-Hox homeobox genes, including entirely different families with entirely different roles. The Hox genes control the body plan along the anterior to posterior axis in us bilaterian animals, but there's still some uncertainty over their precise role in non-bilaterian animals. The Hox genes appear to be unique to animals so you won't find them in plants or fungi. Of course you still find those organisms have *other* homeobox genes; just not the Hox genes, which appear to have arisen very early in animal evolution (there is evidence that sponges had Hox genes too, but have since lost them).

In summary, Hox genes are a family of homeobox genes that play a role in determining which body parts grow and where on the body they grow from.

Why the confusion? These terms and the confusion surrounding them make sense when you consider their history. When scientists first discovered the homeobox domain, they found it because they happened to be studying animals growing body parts in the wrong place (e.g., legs growing where antenna should be). They described them as "homeotic mutants". This was the first of these related terms to be coined.

When scientists identified the specific genes causing the homeotic mutations, they discovered that they shared a common motif, so they named it the homeobox. This was one of the most incredible discoveries in biology, as they soon realised the homeobox was found in genes from humans, flies, jellyfish, daffodils, yeast, and so on. However, the specific genes they had discovered were a *distinct group* of homeobox genes, which they named the Hox genes. They certainly *are* homeobox genes, and they regulate other genes. Consider again the timing of events:

- Scientists were studying homeotic mutants (body parts in wrong places) and discovered that the genes involved all had a highly conserved region, so it made perfect sense to call it the homeobox.
- Those specific genes involved in homeotic mutants came to be known as Hox genes, which also makes perfect sense. Mutations in the Hox genes cause homeotic mutants. There's a theme here.
- However, scientists didn't immediately understand that the homeobox motif would be found in many genes that *aren't* Hox genes. Many homeobox genes have absolutely nothing to do with body parts growing in the right or wrong places (i.e., nothing to do with homeotic mutants) but when they named the homeobox, they only knew of the Hox genes they were discovering via the homeotic mutants.

This is where almost all the confusion stems from. Despite all homeobox genes having "homeo-" in their name, most don't cause homeotic mutants if modified. The Hox genes, a *specific family* of homeobox genes, are famous examples of genes that *can* cause homeotic mutants and that's all these scientists knew at the time.

There are similar situations in astronomy. Planetary nebulae are huge shells of gas ejected from red giant stars. Why "planetary"? When they were first observed using early telescopes, astronomers thought they looked like planets. Today we know they have nothing to do with planets but the name has already stuck. Similarly, most homeobox-containing genes have nothing to do with homeotic mutants but the name has already stuck. The difference between these two situations is that planetary nebulae *never* have anything to do with planets; while biologists in the 1960s really were looking at genes causing "homeotic" mutations. They weren't wrong; they just didn't have the full picture yet.

Today we know so much about homeobox genes, especially the Hox cluster, that they could be discussed for hours on end. The evolution of the Hox, ParaHox, and NK clusters is fascinating, as are the roles of these gene families in a developing animal. Understanding the interplay between development and evolution can provide unique insight into both, which is incredibly exciting, so please don't be put off by the unfortunate naming conventions and mistakes made by others. Homeobox genes are at the heart of some of the most incredible current research in biology

Hox genes are homeobox genes as they contain the homeobox, but homeobox genes include Hox genes, ParaHox genes, NK genes etc. The terms are not interchangeable. It's such an easy mistake to make that it appears in books, academic websites, and helpful videos on YouTube. If you're studying biology, just keep the differences in mind and focus on what exactly is being discussed in your lectures, exam questions, or scientific papers. It's not necessarily wrong to describe a mobile phone as technology but the terms aren't interchangeable. You can't go around describing all technology as mobile phones. It makes no sense to say, "the electron microscope is a wonderful mobile phone". Homeobox and Hox genes work the same way.