THE GENOME AND THE ORIGIN OF MAN

Ph.D. Geoff Barnard
Weizman Institute of Science, Israel
University of Cambridge
Wydział Nauk Weterynaryjnych, Wlk. Brytania
gjrb2@aol.com

Abstract

Until recently, disputes over Darwinism had largely focused on various historical evidences, the so-called „icons of evolution". It is now argued that new genomic evidence has settled the case for common ancestry once and for all. In particular, several books have been written from a Christian perspective seeking to adopt all that neo-Darwinism has to offer. In terms of genomic evidence, the authors present very similar arguments. These include the presence of pseudogenes, mobile genomic elements and endogenous retroviruses which are considered clear evidence for common ancestry. The authors also suggest the high probability that human chromosome 2 is a fusion product of two smaller chromosomes possessed by an ancestral hominid.

In the lecture and in subsequent discussions, we will consider much of this information, present counter-arguments and try to see what can be learned from a design perspective. It is necessary, however, to apologies in advance for the technical detail that will be presented in both sound and vision. Nevertheless, in order to do justice to the arguments, dealing with the technicality is pre-requisite. Consequently, it has been necessary to assume a basic level of biological understanding and this may be an unwarranted assumption by the presenter.

Until recently, disputes over Darwinism have focused on various historical evidences which include molecular and structural homologies, missing links in the fossil record, the Cambrian explosion, embryology, peppered moths and Darwin's finches. While all of these icons are still worthy of discussion, the major debate over common descent has moved into the genomic era.

For many years, biologists believed that approximately 98% of the genome was non-functional as it was not transcribed and translated into protein. The term "junk DNA" was in common use. However, times have changed. In 2007, the first

details of the ENCODE Project were published.[1]   It was an unexpected finding that the vast majority of the genome was transcribed into non-protein coding RNA (ncRNA). Since then, this discovery has been confirmed time and time again. For example, Marcel Dinger and his colleagues at the Institute for Molecular Bioscience, University of Queensland, Brisbane published a paper in 2009 in which they state:

> *Genome-wide analyses of the eukaryotic transcriptome have revealed that the majority of the genome is transcribed, producing large numbers of non-protein-coding RNAs (ncRNAs). This surprising observation challenges many assumptions about the genetic programming of higher organisms and how information is stored and organized within the genome.[2]*

This discovery was unexpected but has been confirmed repeatedly. The word "transcriptome" has effectively replaced the obsolete term "junk DNA". This discovery has opened a window onto the unimaginable complexity of the genome.

We now know that protein-coding genes only produce the "nuts and bolts" of the protein machinery of life and much (but not all) protein machinery is common between species. The ncRNAs, however, are involved in the complex regulation and time-management of gene expression and this is very different between species and humans and the great apes are no exception. The non-protein coding genome comprises many distinct elements including introns, pseudo-genes and mobile genomic elements. Specific examples of these, particularly distinguishing the human species from the great apes, will be given in the lecture.

Introns

The initial mRNA transcript (a complementary copy of the DNA) comprises both exons and introns. However, before the mRNA is translated into protein, the introns are cut out (spliced out) of the final mRNA by specific enzymes. This is illustrated in Fig. 1.

Whereas protein-coding RNA (i.e. without introns) comprises 2% of the human genome, intronic sequences make up approximately 25%. The functions of introns are largely unknown but scientists now recognise that introns possess another layer of biological information that has been called the "Splicing Code".

One of the great surprises of the human genome project was the initial finding that there were only approximately 22,000 protein-coding genes. The expectation was that there would actually be hundreds of thousands. However, with the discovery of the splicing code, it is now thought that many subtly different proteins can be produced from the one RNA transcript. All of this is under highly coordinated complex control, which is different in different cells and tissues, with integrated changes throughout the lifetime of a single organism.

---

[1] Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. (2007) Nature 447:799-816.
[2] Barash Y *et al*. (2010) Deciphering the splicing code Nature 465: 53-9.
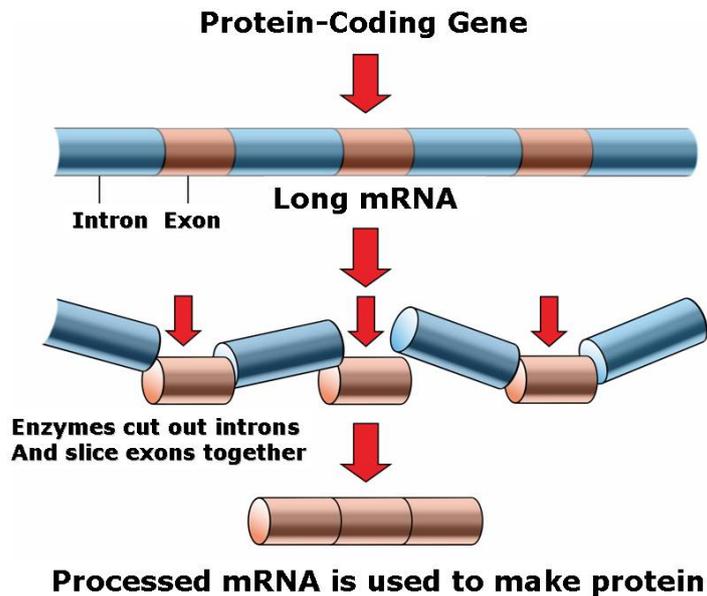
Protein-Coding Gene

Intron Exon   **Long mRNA**

**Enzymes cut out introns
And slice exons together**

Processed mRNA is used to make protein

Fig. 1. Removal of Introns.

Pseudogenes

Pseudogenes are DNA sequences that resemble protein-coding genes but are not transcribed to messenger RNA (mRNA) in a way that could then be translated into some functional protein. Many have suggested that pseudogenes are simply molecular fossils that illustrate and provide evidence for evolutionary history. Implicit in this argument is that pseudogenes are genetic relics that have lost their original protein-coding function, which had been possessed by some ancestral creature. In support of this, evolutionary scientists point to the fact that pseudogenes are scattered throughout the genomes of all higher species (animals and plants) and, in particular, many similar pseudogenes are found in all primates. Biologists have identified two distinct types of pseudogene, often termed "processed" and "unprocessed". These are illustrated in Fig. 2.

As a general rule, processed pseudogenes are usually located on different chromosomes from the protein-coding genes that they resemble. Most biologists believe that they were created by the retro-transposition of the mRNA transcripts from the parent gene. This is because this type of pseudogene lacks introns. Processed pseudogenes also lack the regulatory sequences which are usually found "upstream" of protein-coding genes (before the start sequence), and they have poly-adenine (poly-A) tails which are characteristic of the terminal end of an mRNA. In addition, the pseudogenes are usually flanked by repeat sequences of DNA, which is characteristic of mobile genomic elements (discussed below).
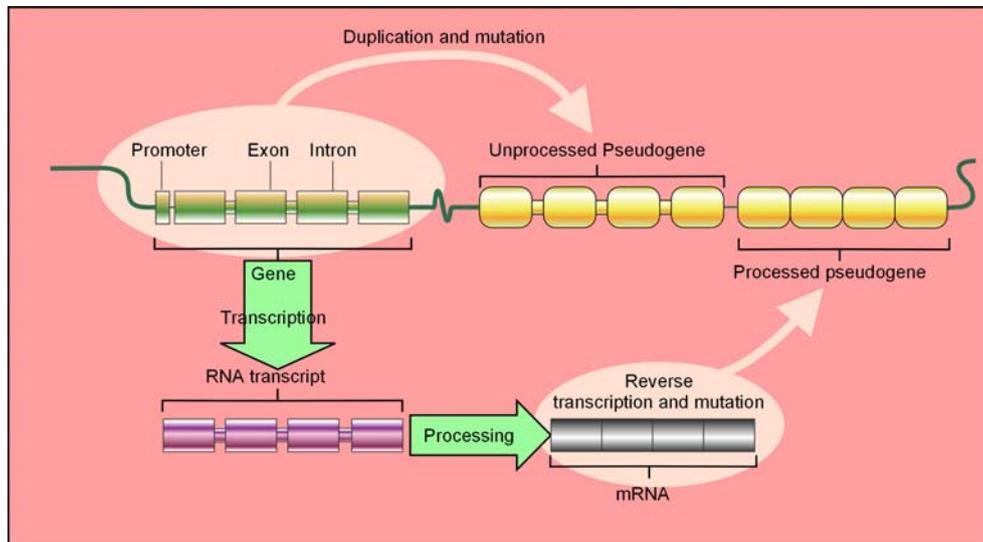
126

Fig. 2: Formation of processed and unprocessed pseudogenes.

Unprocessed pseudogenes, by contrast, are usually found in close proximity to their corresponding protein-coding gene, often on the same chromosome. As a general rule, and unlike processed pseudogenes, they do possess introns and upstream regulatory sequences. Nevertheless, it is believed that the expression of these "genes" is prevented by mutations, deletions and/or insertions of "incorrect" nucleotides. These genetic changes may lead to premature termination or may introduce "frameshifts" that render the message meaningless.

Conservation (i.e. sharing) of similar genetic sequences between species is evidence that indicates that pseudogenes (or any other non-protein coding sequence) possess important biological functions. Such sequences are said to be under purifying (or stabilising) selection, which means that deleterious mutations are removed from the gene pool and genetic diversity is restricted. This is probably the most common role of natural selection, maintaining genetic integrity (and certainly not driving evolutionary change). According to a recent review by Sasidharan and Gerstein:

> *Although pseudogenes have generally been considered as evolutionary 'dead-ends', a large proportion of these sequences seem to be under some form of purifying selection - whereby natural selection eliminates deleterious mutations from the population - and genetic elements under selection have some use.[3]*

Pseudogenes are involved in gene conversion or recombination with functional genes. Pseudogenes exhibit evolutionary conservation of gene sequence,

---

[3] Sasidharan R and Gerstein M (2008) Genomics: protein fossils live on as RNA Nature 453: 729-31.

127

reduced nucleotide variability, excess synonymous over non-synonymous nucleotide polymorphism, and other features that are expected in genes or DNA sequences that have functional roles.

It has been very premature to suggest that pseudogenes are simply genetic fossils. This is not to say that there will never be an example of a pseudogene that is a defunct copy of a protein-coding gene which has lost its activity due to random mutational damage. But it may eventually be necessary to redefine the term "pseudogene" to distinguish between genes that are broken and those genomic elements that possess important roles in gene regulation.

Mobile genetic elements

The genome also contains transposable elements, or transposons. These are sequences of DNA that can move from one position in the genome to another. There are several types of transposon and they are classified according to their mechanism of transposition.

Most retro-transposed genomic elements are DNA sequences known as (1) short interspersed repeated sequences (SINEs), or (2) long interspersed repeated sequences (LINEs). Both types are replicated via RNA intermediates. The majority of the SINEs are the so-called "Alu sequences", which are about 300 base-pairs long, and there are over one million of these in the human genome. They are so named because they can be precisely cut out of the DNA by a specific enzyme (Alu endonuclease) which was isolated from a bacterium (**A**rthrobacter **lu**teus).

It is also very premature to conclude that Alu sequences are just "genetic fossils". Not surprisingly, there have been several recent publications that indicate that Alu sequences may have very important genomic roles. As a general rule, clues to the various roles for Alu sequences are being discovered by the identification of what goes wrong when there is a mutation or inappropriate duplication or deletion. All of this information is circumstantial evidence that normal Alu sequences have important roles.

Endogenous Retroviruses

Retroviruses are viruses that carry their genetic material as RNA rather than DNA. They possess a relatively small number of genes and, like all viruses, cannot replicate without "hijacking" the genetic machinery of the host cell of a higher organism. Retroviruses exploit the enzyme reverse transcriptase to copy their RNA genome into DNA, which is then integrated into the host's DNA genome. From that moment on, the virus replicates as part of the host cell cycle and reproduces by transcription and translation using the cell's own machinery.

It is generally assumed that retroviral genetic insertions have entered the human genome over time and have been passed on from one generation to another. In the case of the human genome, these insertions are known as human endogenous retroviruses (HERVs) and it is thought that they make up between 5% and 8% of the total genome. Most insertions have no known function but it is now understood that at least some HERVs play essential important roles in host biology such as the

control of gene expression, reproduction (e.g. placental function and spermato-genesis) and, indeed, enhancing resistance to infection by pathogenic retroviruses. In addition, we now know that many thousands of "retroviral" promoters are transcribed and initiate transcription throughout the human genome. In a landmark paper entitled *Retroviral promoters in the human genome*, Andrew Conley and co-workers at the Georgia Institute of Technology in the USA reported the existence of 51,197 HERV-derived promoter sequences that initiate transcription within the human genome. These included 1,743 cases where transcription is initiated from HERV sequences that are located in gene promoter regions. In their own words:

> *These data illustrate the potential of retroviral sequences to regulate human transcription on a large scale consistent with a substantial effect of ERVs on the function and evolution of the human genome.[4]*

Although this statement is couched in evolutionary language, these findings also raise the intriguing possibility that the model of an infectious viral origin of HERVs is only partly true. Is it not possible that many HERVs are actually integral functional genetic components which have, as yet, unknown function? The objection to this argument, of course, would be the similarity of the protein coding regions of the HERVs to exogenous retroviruses. However, an alternative hypothesis is that retroviruses might actually have originated as conventional genomic components that "escaped". Only time will tell if there is any substance to this tentative suggestion.

Chromosomal Fusion

Some have suggested that there is incontrovertible genetic evidence that humans and the great apes have descended from the same common hominid ancestor. This argument is supported by referring to chromosomal fusion. The basic facts are these: humans have forty-six chromosomes, while the great apes have forty-eight. The human chromosomes comprise 22 pairs of autosomes and one pair of sex chromosomes. In the female, the sex chromosomes are truly a pair (XX) whereas in the male, there is one X and one Y chromosome. In the case of the great apes (chimpanzees, gorillas and orang-utans) there are 23 pairs of autosomes and one pair of sex chromosomes (again XX female and XY male).

Apart from the structural variation between many of the human and ape chromosomes, the most significant difference is chromosome 2. In the human, there is a single chromosome but in the apes, there are two. Because the evolutionary scenario is now accepted as fact, the ape chromosomes have been re-designated as 2p and 2q rather than their original numbering of 12 and 13.

What evidence is there that a fusion of chromosomes has indeed taken place? Remarkably, there does appear to be some although this is currently under review. Chromosomes only appear as discrete structures in the nucleus when the cell is about to divide by a process known as mitosis. The ends of each

---

[4] Conley AB, Piriyapongsa J, Jordan IK (2008) Retroviral promoters in the human genome. Bioinformatics 24:1563-7.

chromosome are termed telomeres which are repetitive stretches of DNA. The central region of each chromosome called the centromere.

It is clear that the central genomic region in human chromosome 2 appears to contain repeating telomeric sequences. Furthermore, the sequences flanking these telomeric repeats within the centralised region of human chromosome 2 are characteristic of present-day human pre-telomeres which flank the telomeres at the ends of chromosomes. This is shown in Fig. 3.

Fig. 3. Detailed sequences obtained from Human Chromosome 2.

As well as telomeres, centromeres also have a characteristic DNA which has been termed alphoid sequences. If a fusion event has taken place between two ancestral chromosomes, it would be expected to find the evidence of more than one centromere in the fused product.

Accordingly, secondary alphoid DNA in any given chromosome might be seen as a genetic residue left over from a previously functioning centromere on a separate chromosome. Although alphoid DNA is present in human chromosome 2, the situation has become much more difficult to interpret as alphoid regions as well as centralised telomeres have been located in many chromosomes where fusion cannot be the explanation.

130

Meiosis and the Maintenance of Genetic Integrity

Meiosis is a specialised type of cell division that only occurs during the formation of sperm and egg. It is similar to normal cell division (mitosis) except that it also involves a reduction division of chromosomes which results in each gamete (sperm or egg) possessing half the number (haploid number), namely, 23 individual and unpaired chromosomes.

During the first stage of meiosis, homologous chromosomes (e.g. the pairs of chromosome 1, chromosome 2 etc.) accurately line up and become "zipped" together, in a process known as synapsis. Subsequently, chromosomal crossing over or recombination takes place during which individual chromatids of each homologous chromosome exchange segments of genetic information. This is an incredibly precise and accurate process and demands the maintenance of complementarity between the chromatids which are exchanging their DNA sequences. This is illustrated in Fig. 4.
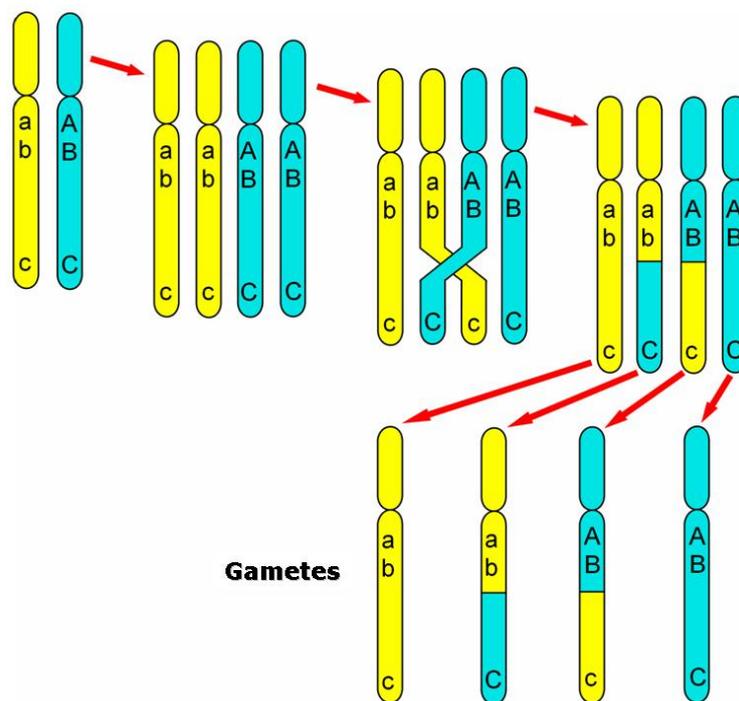

Fig. 4: Synapsis and Recombination in Meiosis.

It is here, more than anywhere else that we come face to face with the Darwinian paradox. By definition, evolution across species must involve gross chromosomal structural changes. The reduction in numbers of chromosomes is

131

actually the least of our problems when we come to consider the supposed common ancestry of humans and chimpanzees. What cannot be tolerated are the wholesale inversions, duplications, deletions, not to mention the insertion of novel lineage-specific genes. Furthermore the position of the centromeres is vitally important in the pairing and processing of homologous chromosomes. Just a cursory glance at the comparison between the human and chimpanzee chromosomes illustrates the problem. Meiosis just cannot happen unless there is synapsis. Variation just cannot happen unless there is recombination.

As a general rule, major chromosomal changes are inevitably damaging and will be eliminated by natural selection. However, let us assume, however, that it might be possible for the supposed genomic alteration either to have no deleterious effect or maybe to be positively beneficial to the individual. Even in this case, for any change to become fixed in the population, a sexual partner will have to be found (presumably by chance) who possesses exactly the same genomic change. Furthermore, the population size must be very small, for if it is not, whatever benefit the change may bring to the individual; it will be "diluted out" within the gene pool and eventually lost. Resolving this issue is not a trivial matter.

Conclusion

To suggest that there is overwhelming evidence in support of the claim that humans and the great apes have common ancestry is, at best, an overstatement. We have dealt with the issue of pseudogenes and mobile genomic elements have shown that such arguments are born out of ignorance. In the second part of the lecture, we considered the argument of chromosomal fusion. It is possible that a chromosomal fusion event has taken place in human history and within the human lineage although we have no need to be dogmatic about this. If it has occurred, it must have taken place when the human population was very small. The implications of this statement are quite profound.

What is certain, however, is that major chromosomal variations that clearly exist between the human and the great apes, such as those that have been well documented in the academic literature and in this lecture, actually dictate against the thesis that these species have common ancestry. In particular, the process of meiosis requires the maintenance of chromosomal structural integrity in order that homologous chromosomes can be perfectly aligned as a pre-requisite for re-combination. Without synapsis, there can be no meiosis. Without meiosis that can be no formation of viable gametes (sperm and egg). Without viable gametes there can be no sexual reproduction and without sexual reproduction there can be no life. It is not the survival of the fittest but the survival of the fertile.