1st Lecture: What is a gene?

Sonja J. Prohaska

Computational EvoDevo
Institute of Computer Science
University of Leipzig

October 14, 2011

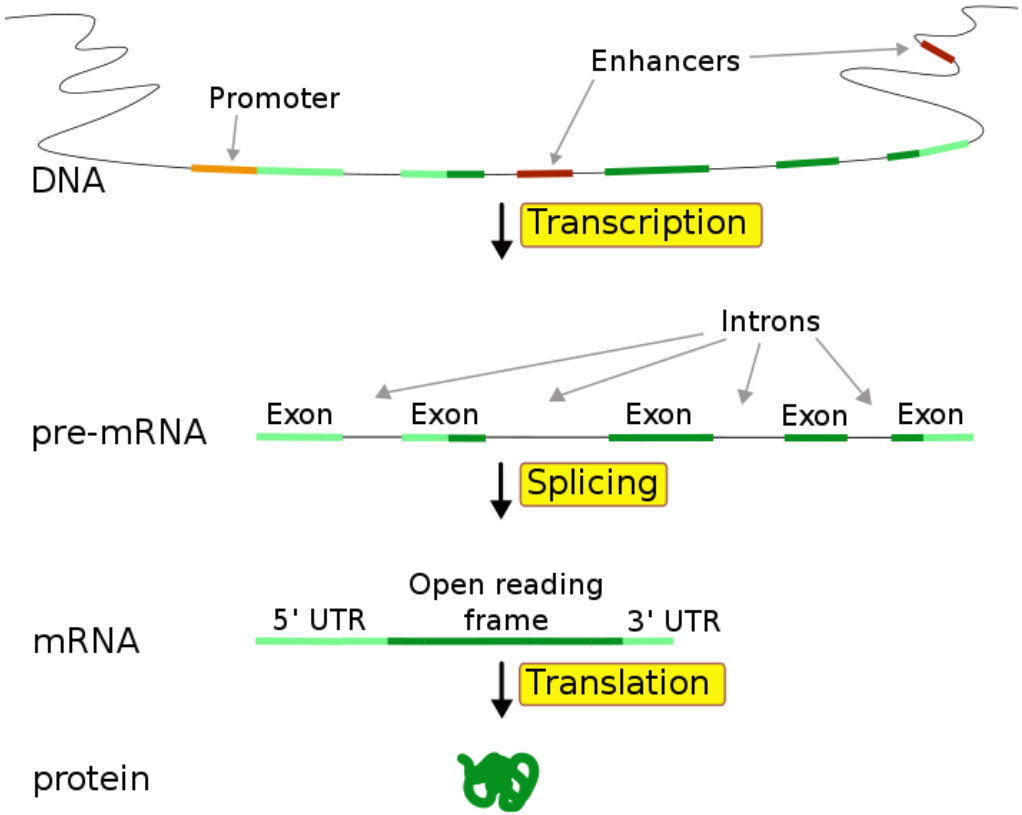
What is a gene?

- a stretch of nucleic acid that codes for a functional (gene) product?
- the necessary information to build and maintain a (several) genetic trait(s)?
- a molecular unit of heredity?

Bottum up: The structural gene definition

A gene is defined as a single **stretch of nucleic acid sequence** or several stretches in a specific order and orientation with a **specific structure**, marked by sequence elements that might differ from species to species.

A eukaryotic gene has the following archetypic structure:



Bottum up: The structural gene definition

The gene sequence can be partitioned into **exonic sequences** and **intronic sequences**. **Exons** are found on the RNA sequence after splicing while **introns** are removed. Exons can be (protein) coding, non-(protein) coding or partially (protein) coding. Non-coding, exonic sequences are found at the 5' or 3'-end of the RNA sequence and are called **untranslated regions**, or short **UTRs**. The sequence between the UTRs is refered to as **coding sequence**. If it forms a contiguous stretch of nucleotides which is a multiple of 3, starts with a start codon (e.g. AUG) and ends with a STOP codon (e.g. UAG) it is called the **open reading frame**. It is translated into a **polypeptide chain** (protein sequence) codon by codon. Promotor and enhancer elements ensure that the gene sequence is transcribed the gene is expressed.

Top down: The functional gene definition

A gene is defined as the **"image" of a functional product** on the genetic material. Knowing a **trait** and the functional proteins involved in its establishment as well as the expression cascade, the coding sequence of each protein is inferred. Similarly, the mRNA and the **genomic locus/loci** serving as a **template** for protein production can be derived.

Biological traits (Merkmale):

- visible traits, e.g. eye color, hair color
- experimentally accessable traits, e.g. blood type, lactose tolerance
- complex traits, e.g. suseptibility for a certain disease, intelligence

gene defintion used by population geneticists

A gene is an abstract construct that might (or might not) be linked to a phenotypic trait. It takes the form of one of multiple variants. These variants are called **allels**. In diploid organisms to alles of each gene are present.

gene	eye color
variant 1	blue
variant 2	green
variant 3	brown

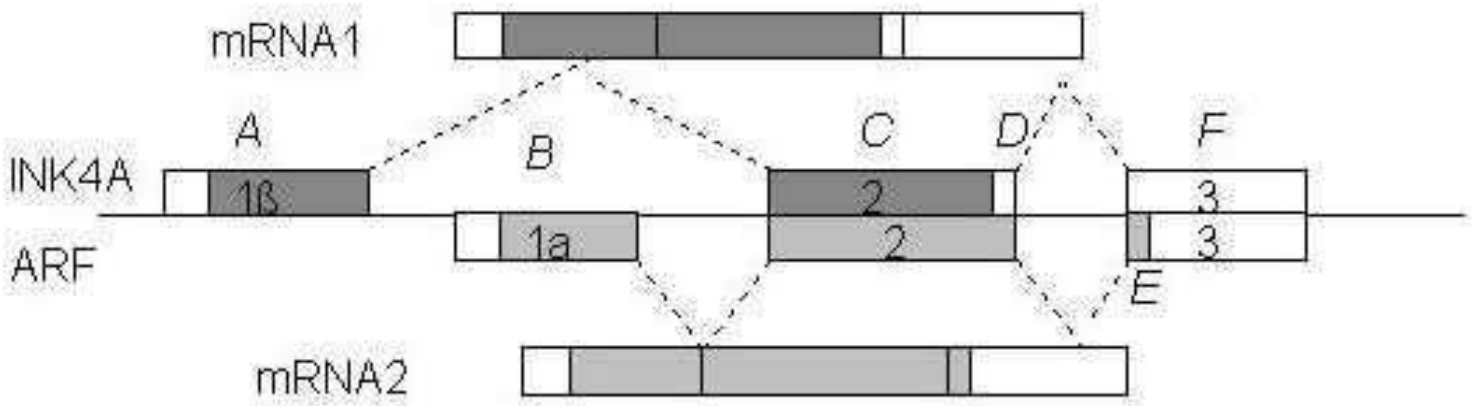
Uniting the structural and functional gene defintion

"one gene → one enzyme → one function"

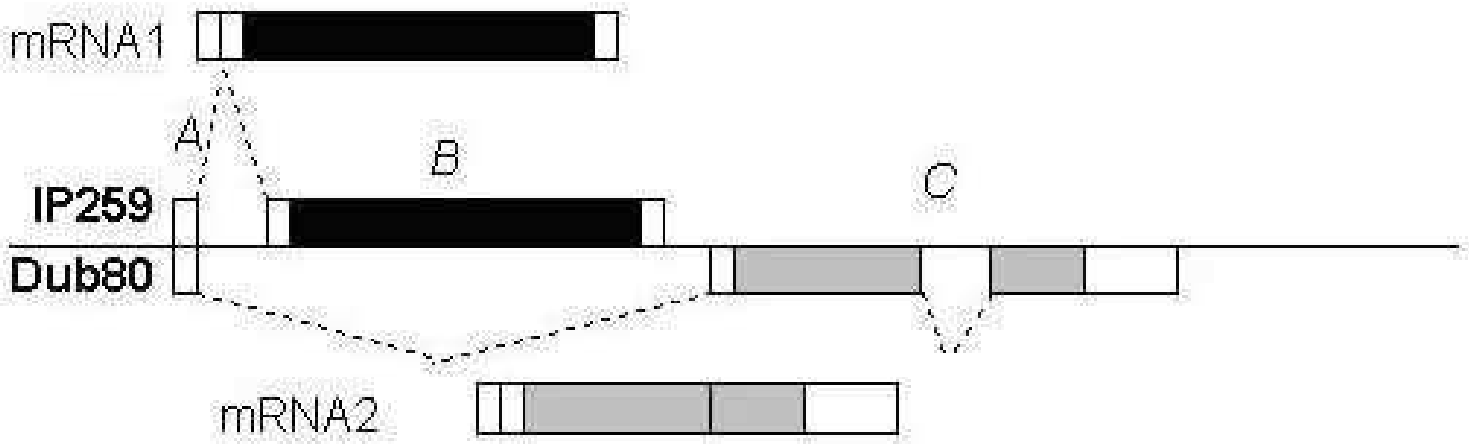
Complications occure when

- a structurally defined gene has no function (e.g. pseudo-genes)
- a species has a very derived and unknown gene structure (e.g. ciliates)
- a functionally defined gene has not the expected structure (e.g. trans-spliced genes)
- a gene locus is alternativly processed into several functional products
- counting genes

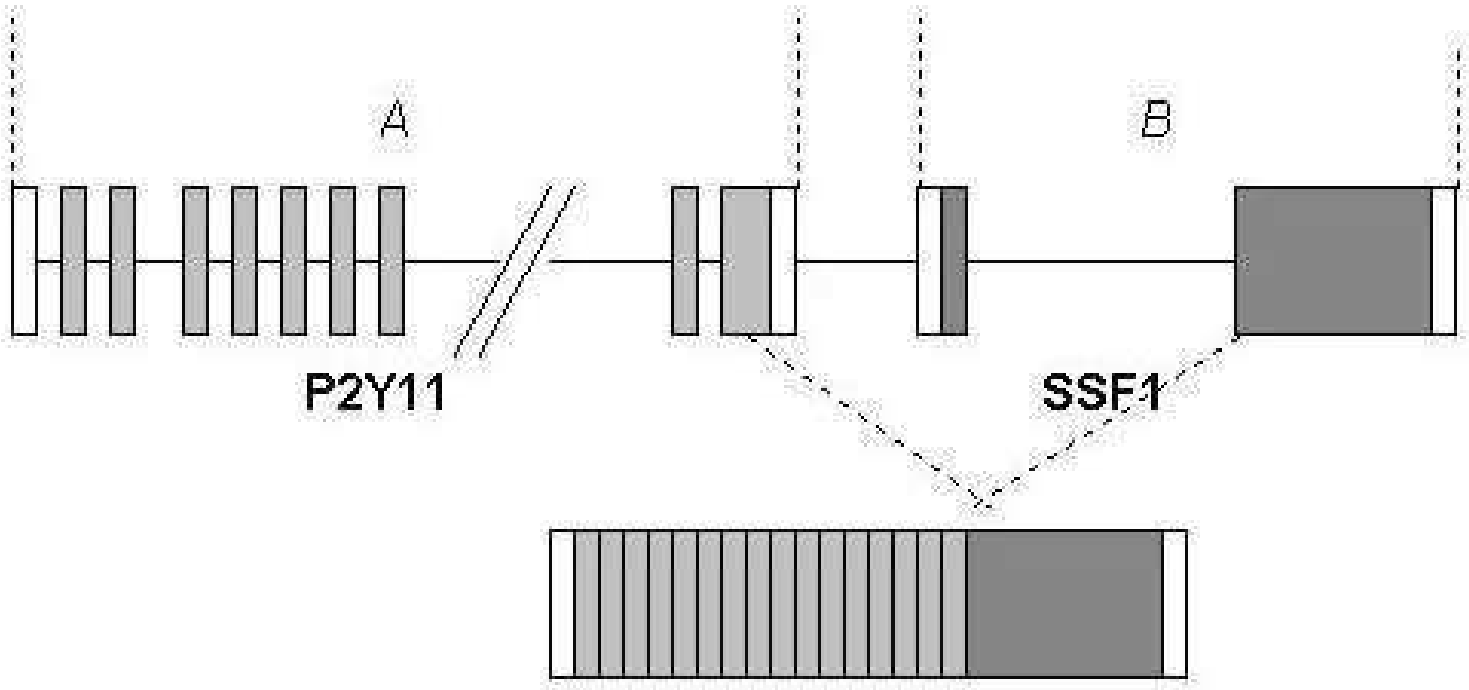
One or Two Genes? Case 1



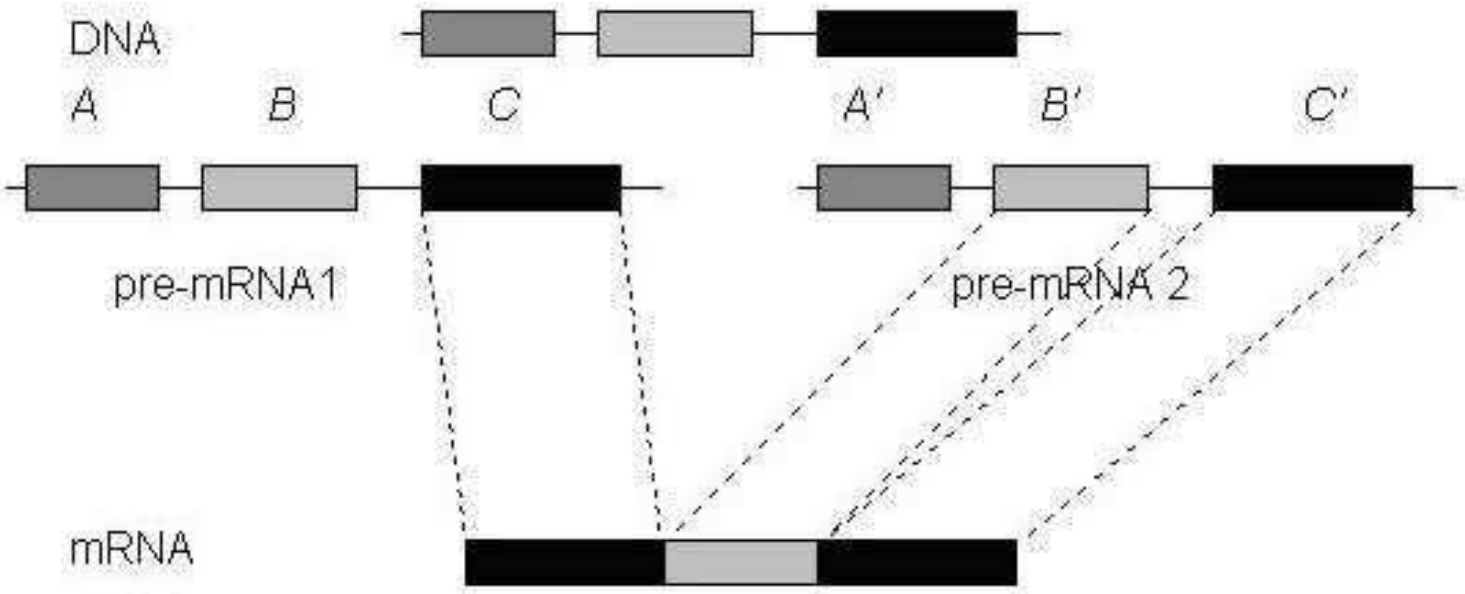
One or Two Genes? Case 2



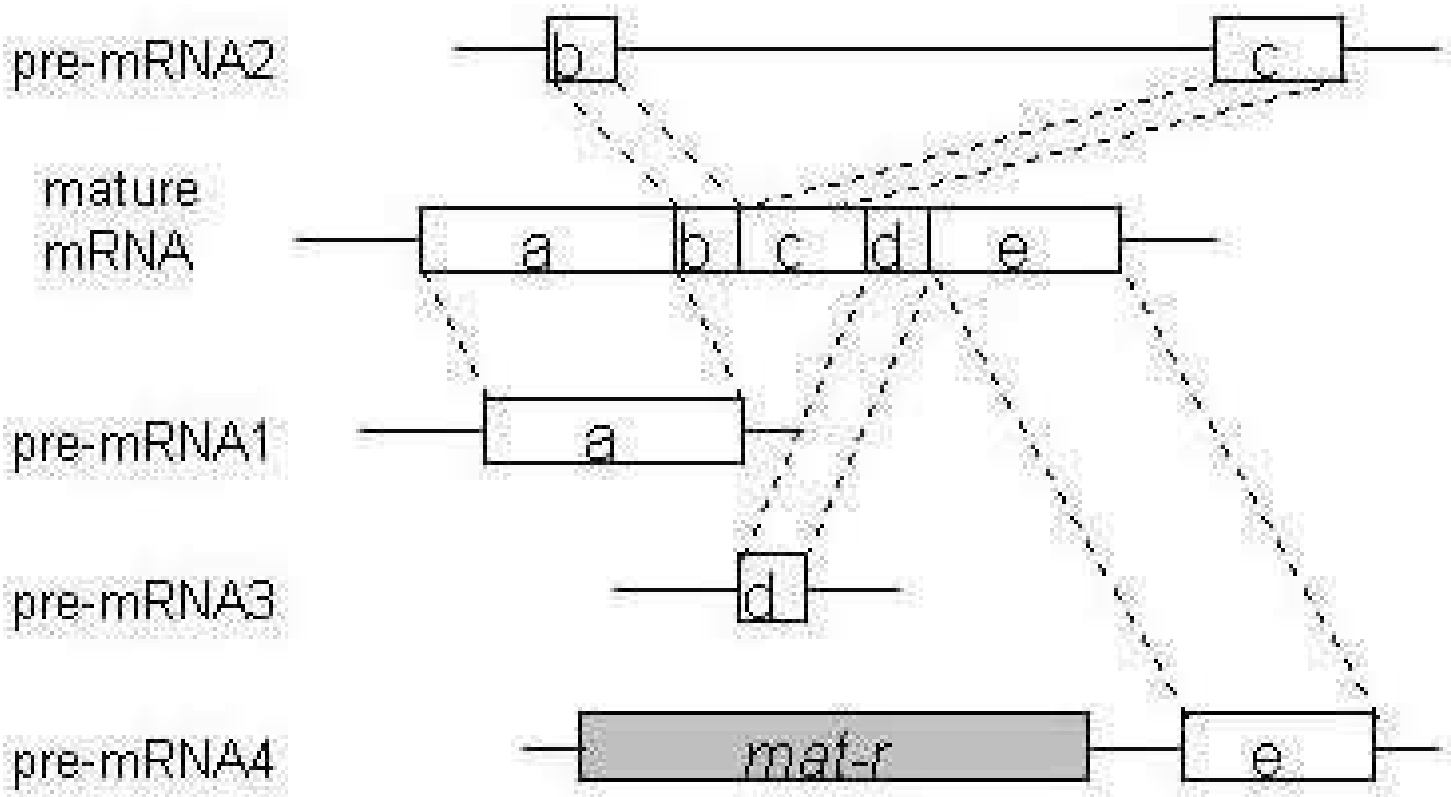
One or Two Genes? Case 3



One or Two Genes? Case 4



One or Two Genes? Case 5



[Griffiths, 2006] Griffiths P. E. and Stotz K. *Genes in the postgenomic era*. Theor Med Bioeth. 2006;27(6):499-521

Currently, a gene is...

- molecular biology
  anything that looks like a traditional gene and is an image of the gene product in the DNA
- bioinformatics
  anything an automated gene discovery tool returns as gene with good reliability
- evolutionary biology and population genetics
  anything which causes a phenotypic difference and behaves like a Mendelian allele

The current usage of the term "gene" differs between disciplins, is vague, and causes frequent misunderstanding. Substitute the word "gene" by a well defined term whenever possible!

[Griffiths, 2006] Griffiths P. E. and Stotz K. *Genes in the postgenomic era*. Theor Med Bioeth. 2006;27(6):499-521

The Gene - The Early Concept and Terminology

Gregor Mendel

1866, St. Thomas in Brno

working with pea plants, Mendel discovered that traits are determined by discrete units that are passed on from generation to generation. Independent sorting of these units is possible.

Hugo de Vries

Wilhelm Johannsen

1889, Hugo de Vries coined the term "pangen" for "the smallest particle [representing] one hereditary characteristic". About a decade later, Wilhelm Johannsen abbreviated this term to "gene".

What is the cellular basis of heredity?

The Gene - a Bead on a String

Thomas A. Morgan

1910, Columbia University

working on Drosophila, Morgan was able to demonstrate that genes are carried on chromosomes.

How are genes organized on a chromosome?

Genes on different chromosomes follow the rules of Mendelian inheritance. Genes on the same Chromosome are linked, however they sort independently with a certain rate. This recombination rate, measured in "Morgan", was then used to calculate distances between genes and to draw genetic maps of joined genes.

The molecular basis of heredity was still unclear.

The Gene - Finding Missing Links

Frederick Griffith

1928, England

working on bacteria, Griffith showed that the material of heredity can be transfered.

Alfred Hershey and Martha Chase

1952, Cold Spring Harbor Laboratory

working on Phages, they showed that DNA (rather than protein) is the molecular basis of heredity.

The Gene - Discoveries Change the Picture

about 1960, transcription of genes into mRNA was discovered.

Phillip A. Sharp

Sir Richard J. Roberts

1977, Cold Spring Harbor Laboratory and Massachusetts Institute of Technology

working with viruses infecting the cells of higher organisms Sharp and Roberts independently discovered that *"genes in eukaryotes are not contiguous strings but contain introns, and that the splicing of messenger RNA to delete those introns can occur in different ways, yielding different proteins from the same DNA sequence."*

The Gene - Changes to the Gene Concept

1978, Harvard University

Walter Gilbert

used the findings about splicing to redefine the gene.

"The notion of the cistron [...] must be replaced by that of a transcription unit containing regions which will be lost from the mature messenger - which I suggest we call introns (for intragenic regions) - alternating with regions which will be expressed - exons."

Cistron: unit of genetic function. (Benzer)

Property: Encoding a Function

A gene encodes a gene product that has a function.

- Many elements (e.g. “genomic locus”, “transcribed region”, “coding sequence”) are already defined that can substitute the term “gene” in a structural context.
- Definition of arbitrary structural elements yields biologically irrelevant annotation.
- The term “gene” should be reserved to express linkage of genomic sequences to a phenotypic traits.

Property: Unit of Function

A gene is a unit of function.

- The gene product should show stability in isolation. This renders it accessebly to experimental analysis. (E.g. It can be cloned, expressed and purified.)
- The gene is a unit selection can act on. (Selection acts on traits, traits are passed on to the next generation as DNA sequence. Only changes to the DNA that translate into a change of a trait can be selected.)
- As a functional unit, the gene/ gene product is a building block for more complex cellular functions.

It's the gene product that is a unit, the structual representation might be dispersed on the genome!

Property: Heritability

A gene is heritable.

- A gene's image on the genetic material can be viewed as the transferable instruction to implement a function.
- It follows the rules of inheritance. (E.g. Genes must be replicated and segregated as part of the process of inheritance.)

Taken together

A gene is a **heritable unit of function**.

- It combines the **functional unit** with the **image on the genetic material**.
- A gene links the genotype with the phenotype.