

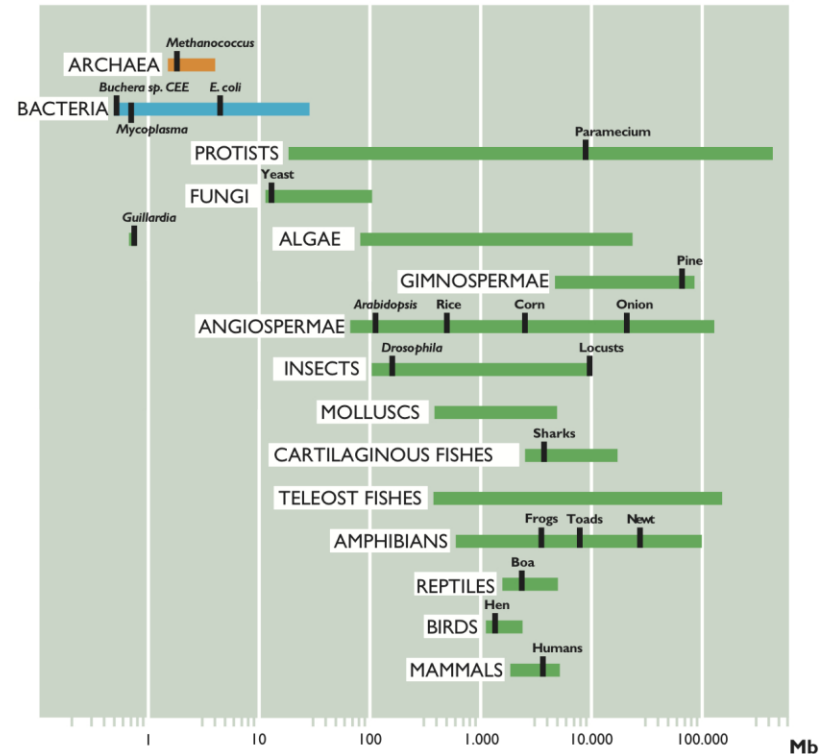
Predicción computacional de genes

La complejidad de los genomas

¿Los organismos más complejos presentan genomas más grandes y contienen más genes?

Organism	Common name or class	Scientific name	Genome size (Mb)	Number of genes	Gene density (genes/Mb)
Eukaryotes					
Baker's yeast		<i>Saccharomyces cerevisiae</i>	12	6,241	480
Nematode		<i>Caenorhabditis elegans</i>	97	18,424	190
Cruciferous		<i>Arabidopsis thaliana</i>	125	25,498	204
Fruit fly		<i>Drosophila melanogaster</i>	180	13,601	75
Pufferfish		<i>Fugu rubripes</i>	400	35,000	100
Rice		<i>Oryza sativa</i>	450		
Sea urchin		<i>Strongylocentrotus purpuratus</i>	900	27,350	30
Maize		<i>Zea mays</i>	2,400		
Human		<i>Homo sapiens</i>	3,400	35,000	10
Onion		<i>Allium cep</i>	18,000		
Amoeba		<i>Amoeba dubia</i>	686,000		
Archaea					
Crenarchaeota		<i>Aeropyrum pernix</i>	1.55	1,522	981
Euryarchaeota		<i>Methanococcus jannaschii</i>	1.66	1,715	1033
Euryarchaeota		<i>Archaeoglobus</i>	2.18	2,420	1110
Bacteria					
Proteobacteria		<i>Buchnera sp. CCE</i>	0.45		
Gram positive		<i>Mycoplama genitalium</i>	0.58	479	831
Proteobacteria		<i>Buchnera sp. APS</i>	0.64	564	881
Gram negative		<i>Haemophilus influenzae</i>	1.8	1,727	959
Cyanobacteria		<i>Synechocystis sp.</i>	3.6	3,168	880
Gram positive		<i>Bacillus subtilis</i>	4.2	4,100	976
Proteobacteria		<i>Escherichia coli</i>	4.6	4,288	932

Table 1. Genome size, gene number and gene density.



Procariotas

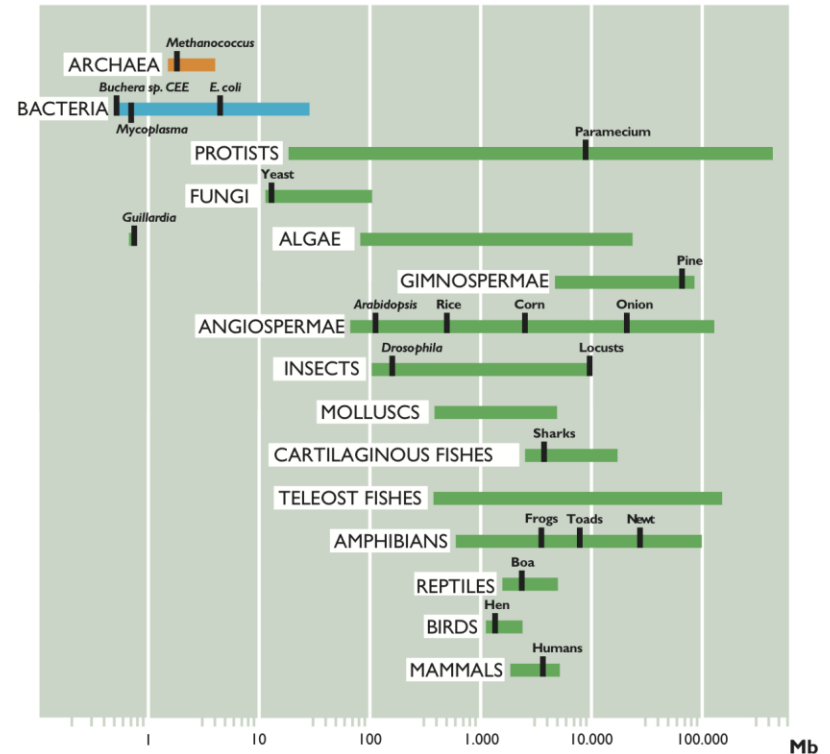
- ~ 900-1000 pbs por gen
- Genomas más grandes, mayor número de genes
- Especialistas más pequeños (parásitos o endosimbiontes)
- Generalistas más grandes (cierto grado de desarrollo)

La complejidad de los genomas

¿Los organismos más complejos presentan genomas más grandes y contienen más genes?

Organism	Common name or class	Scientific name	Genome size (Mb)	Number of genes	Gene density (genes/Mb)
Eukaryotes		<i>Saccharomyces cerevisiae</i>	12	6,241	480
Baker's yeast					
Nematode		<i>Caenorhabditis elegans</i>	97	18,424	190
Cruciferous		<i>Arabidopsis thaliana</i>	125	25,498	204
Fruit fly		<i>Drosophila melanogaster</i>	180	13,601	75
Pufferfish		<i>Fugu rubripes</i>	400	35,000	100
Rice		<i>Oryza sativa</i>	450		
Sea urchin		<i>Strongylocentrotus purpuratus</i>	900	27,350	30
Maize		<i>Zea mays</i>	2,400		
Human		<i>Homo sapiens</i>	3,400	35,000	10
Onion		<i>Allium cep</i>	18,000		
Amoeba		<i>Amoeba dubia</i>	686,000		
Archaea		<i>Aeropyrum pernix</i>	1.55	1,522	981
Crenarchaeota					
Euryarchaeota		<i>Methanococcus jannaschii</i>	1.66	1,715	1033
Euryarchaeota		<i>Archaeoglobus</i>	2.18	2,420	1110
Bacteria		<i>Buchnera sp. CCE</i>	0.45		
Proteobacteria					
Gram positive		<i>Mycoplama genitalium</i>	0.58	479	831
Proteobacteria		<i>Buchnera sp. APS</i>	0.64	564	881
Gram negative		<i>Haemophilus influenzae</i>	1.8	1,727	959
Cyanobacteria		<i>Synechocystis sp.</i>	3.6	3,168	880
Gram positive		<i>Bacillus subtilis</i>	4.2	4,100	976
Proteobacteria		<i>Escherichia coli</i>	4.6	4,288	932

Table 1. Genome size, gene number and gene density.

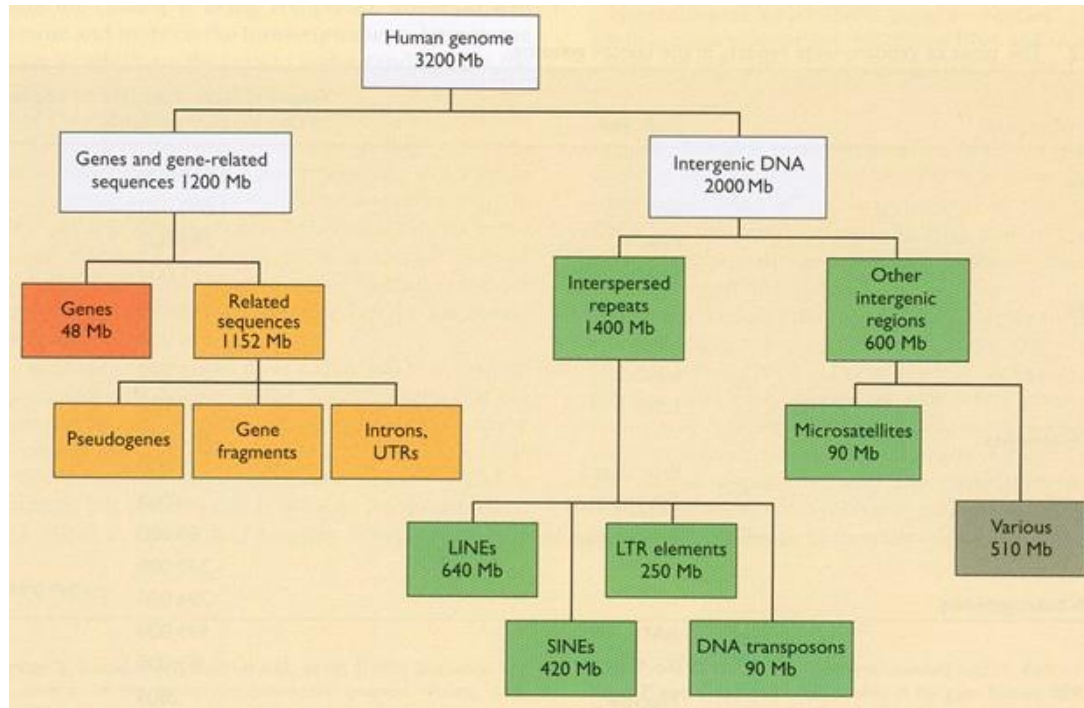


Eucariotas

- Valor-C eucariotas > procariotas (Guillardia thera, alga roja simbiote)
- Variación en valor-C no se corresponde con la complejidad del organismo ni con el número de genes (Paradoja del valor-C)

La complejidad de los genomas

¿Qué explica la paradoja del valor-C?



Complejidad en la regulación y evolución de los genomas

Elementos repetidos

Elementos reguladores

Pseudogenes

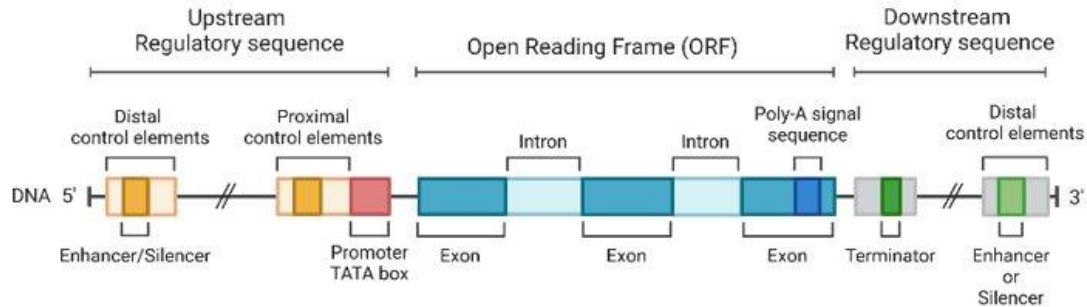
Familias génicas

Estructura génica (Intrones)

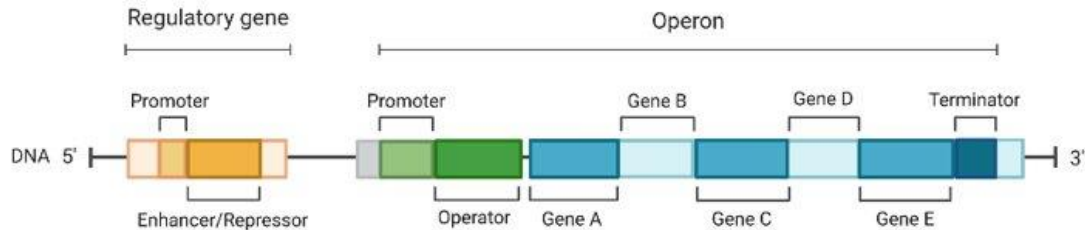
Estructura de genes eucariotas

¿Cuáles son los elementos que componen los genes eucariotas?
¿Qué diferencias observamos con los procariotas?

Eukaryotic Gene Structure



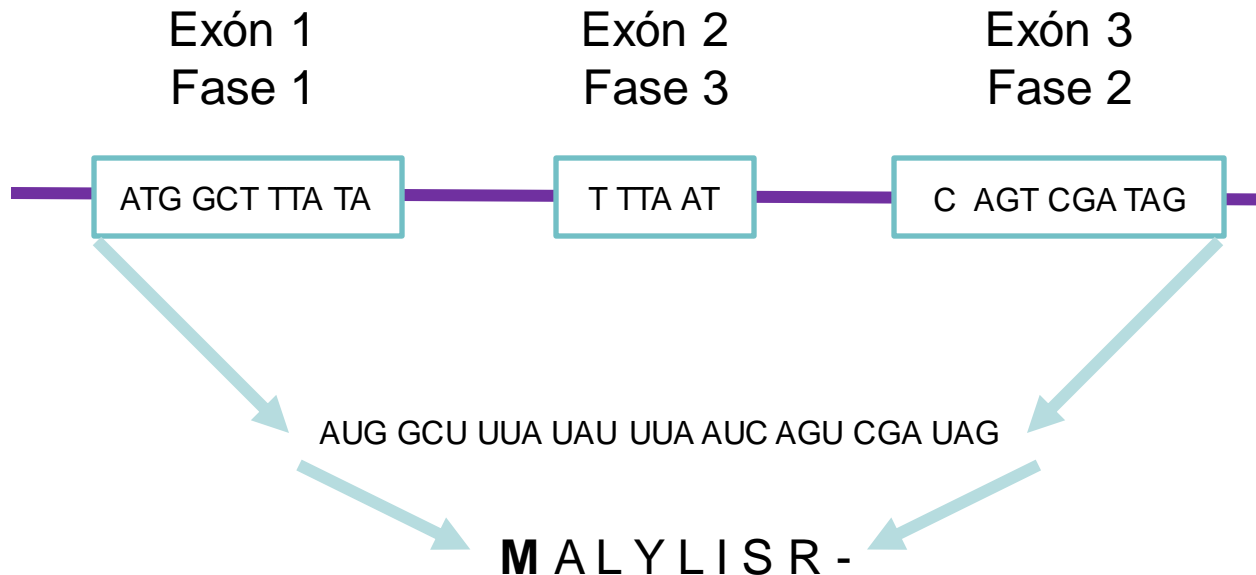
Prokaryotic Gene Structure



Estructura de genes eucariotas

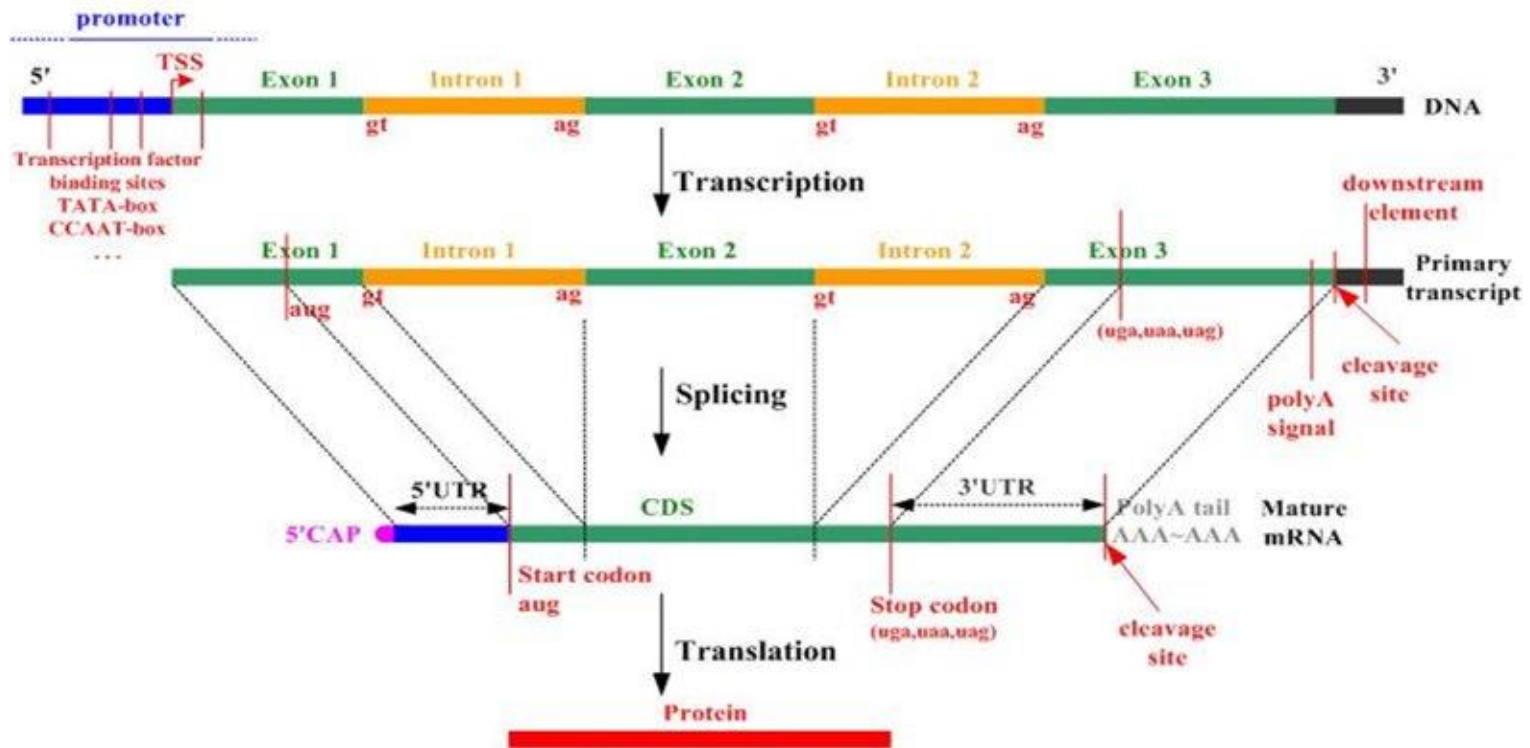
Marco de lectura y exones en genes eucariotas

- Todos los exones de un gen están sobre la misma hebra, pero pueden estar en fases diferentes.
- Los exones de un gen tienen que tener una pauta de lectura consistente.



Predicción de genes eucariotas

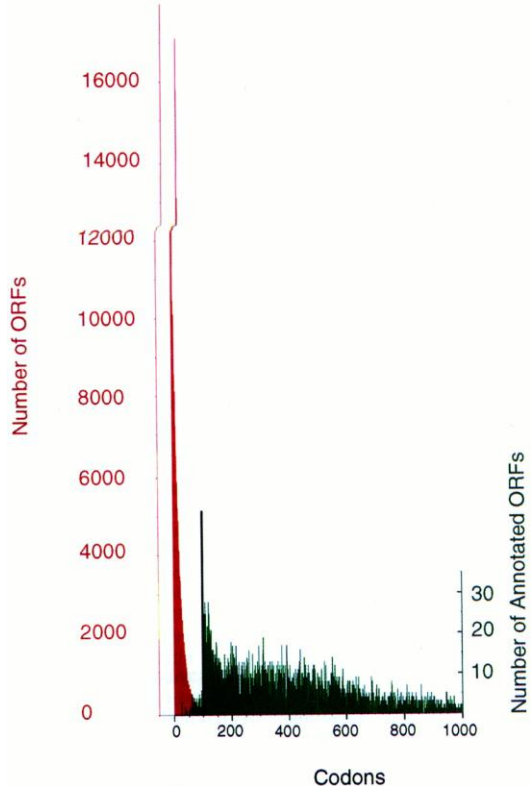
¿Cómo afecta esta estructura a su predicción?



El ORF (Open Reading Frame) no se encuentra en la secuencia de ADN

Predicción de genes eucariotas

¿Cómo afecta esta estructura a su predicción?



ORFs predichos en *S. cerevisiae*

- ~ 260.000 entre 1-99 codones
- ~ 6.000 entre 100-1000 codones

ORFs cortos gran probabilidad de no tener significado biológico



Small Open Reading Frames: Beautiful Needles in the Haystack

Munira A. Basrai, Philip Hieter, and Jef D. Boeke

Predicción de genes eucariotas

¿Cómo podemos buscar/predecir genes que contiene exones?

- Búsqueda a partir de modelos teóricos (*ab initio*)
- Homología y/o emparejamientos de ESTs (Expressed Sequence Tags), cDNA, o genes/proteínas del mismo organismo o de otros relacionados.
- Híbrido, combinando homología y búsqueda *ab initio* (a ser posible incluyendo evidencia experimental)

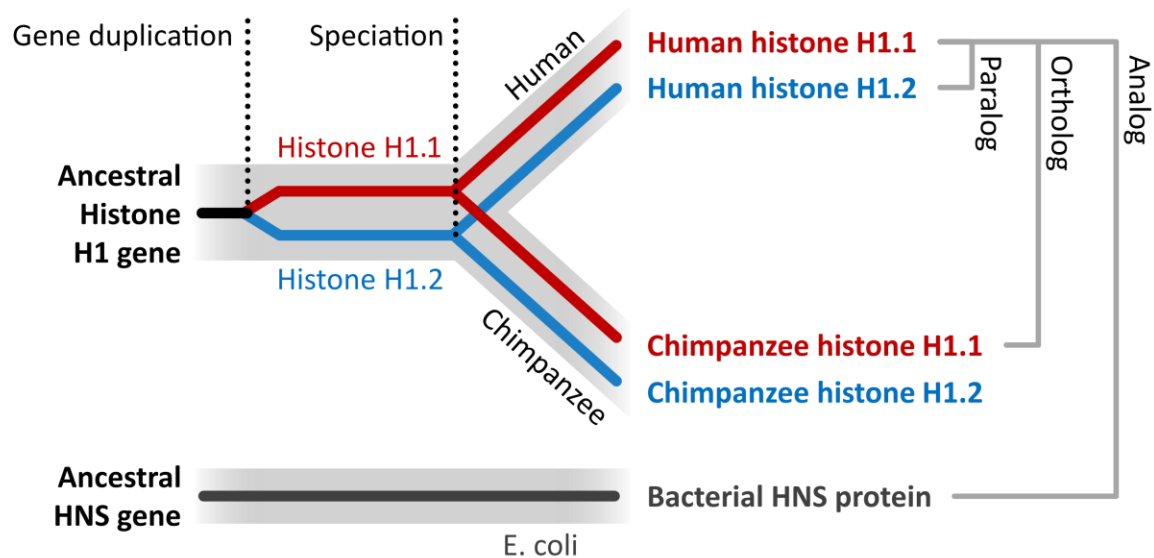
Predicción de genes eucariotas

Modelos *ab-initio*

- Potencial codificador de los exones CDS
 - Los exones tienen una composición de secuencia diferente de las regiones no-exónicas (frecuencia de hexámeros)
 - Modelos de aprendizaje automatizado
- Señales con donadores y aceptores
 - Los 2 primeros nucleótidos de un intrón son en la inmensa mayoría de los casos 'GT' (Donador)
 - Los 2 últimos nucleótidos de un intrón son en la inmensa mayoría de los casos 'AG' (Aceptor)
- Promotores (Islas CpG, composición de secuencia)
 - La búsqueda de genes se puede refinar mediante la detección de regiones o motifs asociados al promotor (TFBS, islas CpG)

Predicción de genes eucariotas

Búsqueda por homología entre especies próximas



Predicción de genes eucariotas

Software predicción *ab-initio*

- Genscan (ab initio)
<http://hollywood.mit.edu/GENSCAN.html>
- GenomeScan (hybrid)
<http://hollywood.mit.edu/genomescan.html>
- Twinscan (hybrid)
<https://mblab.wustl.edu/software.html>
- FGENESH (ab initio)
<http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=gfind>
- GeneMark.hmm (ab initio)
<http://opal.biology.gatech.edu/GeneMark/>
- MZEF (ab initio)
<http://rulai.cshl.org/tools/genefinder/>
- GrailEXP (hybrid)
<http://pbil.univ-lyon1.fr/members/duret/cours/insa2004/exercise4/pgrail.html>
- GeneID (hybrid)
<https://genome.crg.es/geneid.html>