

I GENI DELLE IMMUNOGLOBULINE NEI PRIMATI

IMMUNOGLOBULIN GENES IN PRIMATES

SESSIONE DI LAUREA 2023/2024

CORSO DI LAUREA IN SCIENZE
BIOLOGICHE

LAUREANDO: LORENZO
EVANGELISTI

DOCENTE REFERENTE: PROF.
MARCO BARUCCA



UNIVERSITÀ
POLITECNICA
DELLE MARCHE

Riassunto:

Lo sviluppo di un'applicazione, di un software di apprendimento automatico, basato su reti neurali, ha permesso di scansionare i genomi ed identificare le sequenze di esoni che codificano per i domini CH delle immunoglobuline e da questi geni è possibile risalire alle diverse immunoglobuline che ogni specie possiede. Da questo studio, è stata descritta la distribuzione dei geni delle immunoglobuline tra gli ordini dei primati. Nelle proscimmie, i geni IgD non sono presenti; nei platirrini, un gene è identificato per ciascuna delle classi di immunoglobuline ma il gene IgD non ha l'esone CH2; nella famiglia Cercopithecidae viene rilevato un gene per ogni classe nella famiglia Colobinae, mentre in Cercopithecidae i geni per IgG sono stati duplicati più volte. Negli ominidi, si osserva un maggior numero di duplicazioni che includono i geni che codificano per IgA e IgE. Questi risultati indicano che dalla comparsa dei Cercopithecidae, c'è un'instabilità evolutiva nel locus Ig.

Cosa sono?

Le immunoglobuline, note comunemente con il termine di **anticorpi**, sono proteine globulari coinvolte nella risposta immunitaria, cioè nel legare e neutralizzare le sostanze potenzialmente dannose per l'organismo note con il termine di **antigeni**.

Nei mammiferi esistono cinque classi di immunoglobuline: IgM, IgD, IgE, IgA e IgG.

Negli ultimi anni le sequenze del genoma di molte specie di primati sono diventate disponibili sotto forma di set di dati **WGS** (**Whole Genome Shotgun Sequencing**) assemblati.

Una visione generale del locus IGHC può essere ottenuta in ogni famiglia di primati e sulla base di questo si è sviluppato un software



Chfinder

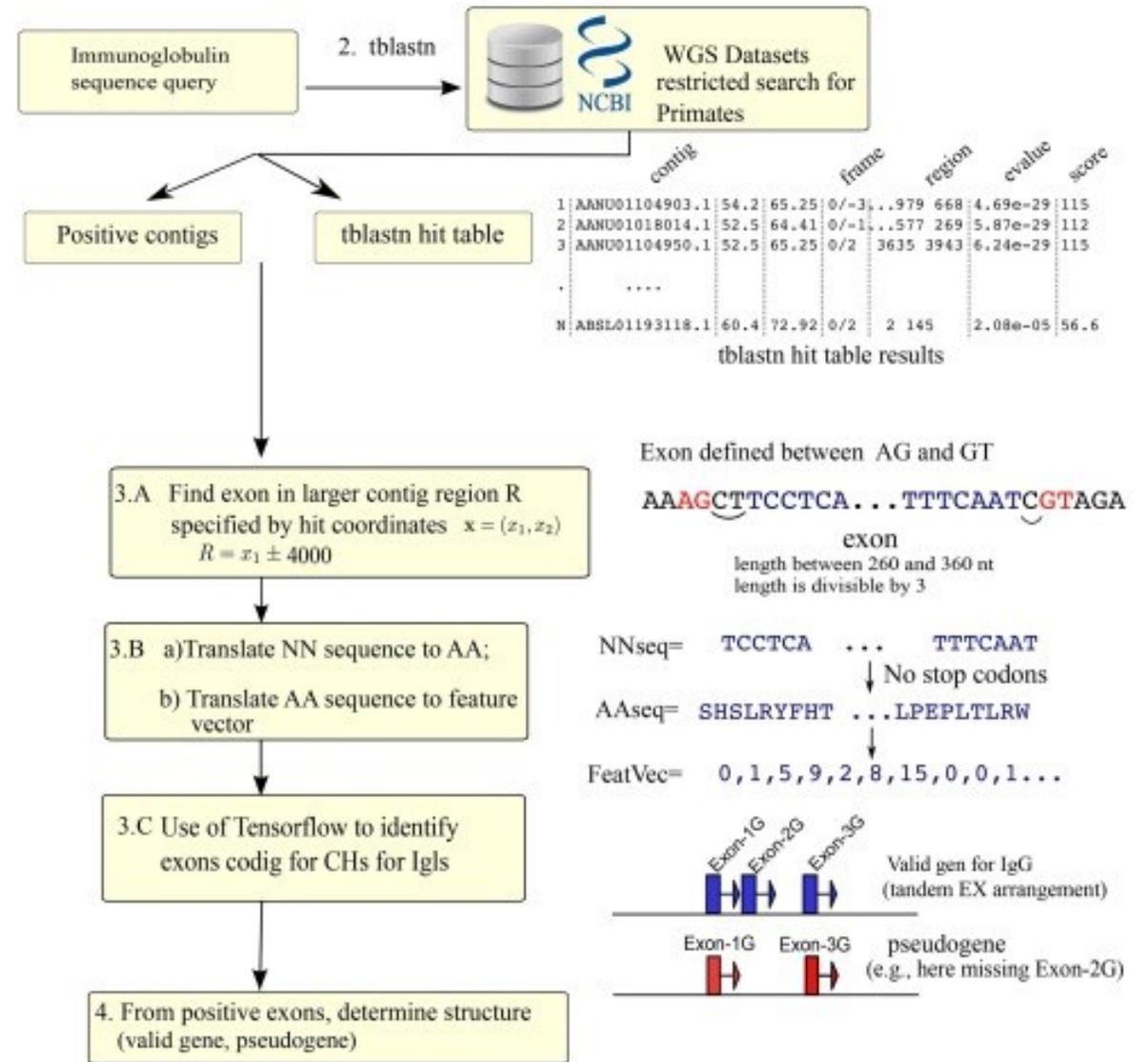


Chfinder è in grado di estrarre le sequenze codificanti (esoni) per i domini costanti delle catene pesanti delle Ig e tali sequenze geniche sono importanti allo scopo di confrontare i geni Ig tra le diverse specie

Un solo requisito deve essere rispettato:



ATTCGATTTCGATTTCGATTTCG
(ripetizione in tandem)



Proscimmie

Nei genomi di sei specie di proscimmie si sono individuati 70 esoni che codificano per i CH di IgM, IgG, IgE e IgA, con l'eccezione di una sola specie...

Table 1
The number of exons that code for immunoglobulin CH domains in Prosimians.

	IgM					IgD				IgG				IgE				IgA				
	CH1	CH2	CH3	CH4	IgM ^a	CH1	CH2	CH3	IgD ^a	CH1	CH2	CH3	IgG ^a	CH1	CH2	CH3	CH4	IgE ^a	CH1	CH2	CH3	IgA ^a
<i>Daubentonia madagascariensis-AGTM01</i>	-	-	1	1	0	-	-	-	0	1	-	1	0	-	-	-	-	0	-	-	1	0
<i>Eulemur flavifrons-LGHW01</i>	1	1	1	1	0	-	-	-	0	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Eulemur macaco-LGHX01</i>	1	1	-	-	0	-	-	-	0	1	1	1	1	1	1	1	1	1	-	1	1	0
<i>Microcebus murinus-ABDC03</i>	1	1	1	1	1	-	-	-	0	1	1	1	1	1	1	1	-	0	1	1	1	1
<i>Otolemur garnettii-AAQR03</i>	1	1	1	1	0	-	-	-	0	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Propithecus coquereli-JZKE01</i>	-	-	-	-	0	-	-	-	0	2	2	2	2	1	1	1	1	1	1	1	1	1
<i>Tarsius syrichta-ABRT02</i>	1	1	1	1	1	-	-	-	0	-	2	2	0	2	1	1	1	1	1	1	1	1

^a Number of molecules deduced by CHfinder when finding the corresponding consecutive exons in the same contig.

File RNAseq utilizzati per determinare la presenza di RNA per le immunoglobuline.

Platirrini

Un isotipo per ogni classe di immunoglobuline da sette file del genoma WGS

Table 3
The number of exons coding for the immunoglobulin CH domains in Platyrrhines.

	IgM				IgD				IgG				IgE				IgA					
	CH1	CH2	CH3	CH4	IgM ^a	CH1	CH2	CH3	IgD ^b	CH1	CH2	CH3	IgG ^b	CH1	CH2	CH3	CH4	IgE ^b	CH1	CH2	CH3	IgA ^b
<i>Aotus nancymaae-JYKP02</i>	1	1	1	1	1	1	-	1	1 ^b	1	1	-	0	1	1	2	1	1	-	1	1	0
<i>Callithrix jacchus-ACFV01</i>	1	1	1	-	0	-	-	1	0	-	1	1	0	-	1	3	1	0	1	1	1	1
<i>Callithrix jacchus-BBXX01</i>	1	1	1	-	0	-	-	1	0	-	1	1	0	1	1	3	1	0	1	1	1	1
<i>Callithrix jacchus-JRUL01</i>	-	-	-	-	0	-	-	-	0	-	-	1	0	-	-	-	-	0	-	-	-	0
<i>Callithrix jacchus-NTCD01</i>	1	1	1	1	1	2	-	-	0	1	1	2	0	2	1	4	1	1	1	1	1	1
<i>Cebus capucinus-L1WQ01</i>	1	1	1	1	1	1	-	1	1 ^b	1	1	1	1	1	2	1	1	1	1	1	1	1
<i>Saimiri boliviensis-AGCE01</i>	1	1	1	1	1	-	-	1	0	-	1	1	0	1	1	1	1	1	1	1	1	1

^a Number of molecules deduced by CHfinder when finding the corresponding consecutive exons in the same contig.

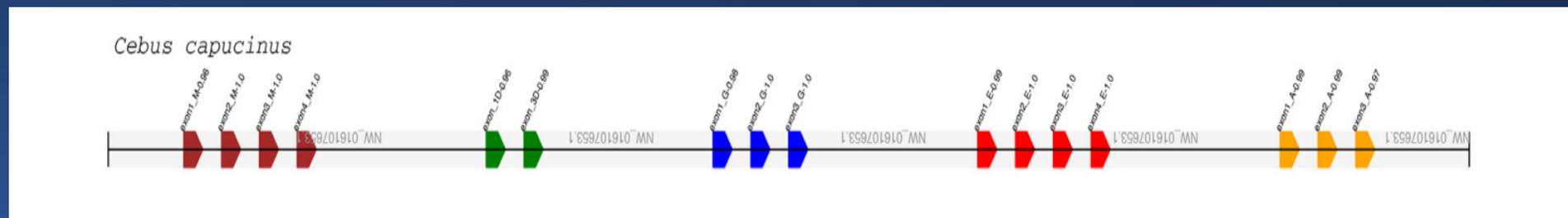
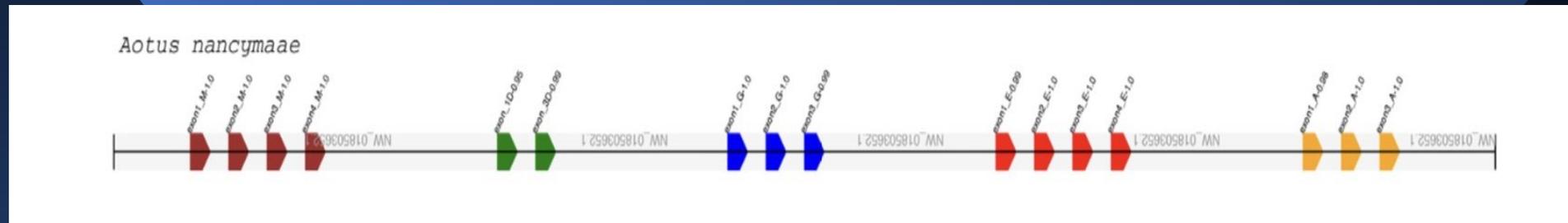
^b Genes apparently viable but only with the domain CH1 and CH3.

In alcuni genomi platirrini è stata individuata la presenza di esoni che si trovano al di fuori del locus

Esoni extra

RETROTRASCRIZIONE

Attraverso il software, l'esone CH2 dei geni IgD non viene rilevato e per vedere se la motivazione a questo fenomeno sia dovuta al fatto che il gene ha solo due esoni:



Cercopithecoidea

Presenza di ulteriori esoni IgE CH2 e duplicazioni di IgG in alcune specie.

Table 4
The number of exons that code for immunoglobulin CH domains in Cercopithecoidea.

	IgM				IgD				IgG				IgE				IgA							
	CH1	CH2	CH3	CH4	IgM ^a	CH1	CH2	CH3	IgD ^b	CH1	CH2	CH3	IgG ^c	CH1	CH2	CH3	CH4	IgE ^d	CH1	CH2	CH3	IgA ^e		
<i>Cercocebus atys</i> -JZLG01	1	1	1	1	1	1	1	1	1	3	3	3	3	1	2	1	1	1	1	1	1	1	1	1
<i>Chlorocebus sabaues</i> -AQIB01	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	2	1	1	1	1	1	1	1
<i>Colobus angolensis</i> -JYKR01	1	1	1	1	1	1	1	1	1	0	2	1	1	1	2	3	1	1	1	1	1	1	1	0
<i>Macaca fascicularis</i> -AEHL01	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
<i>Macaca fascicularis</i> -AQIA01	1	1	1	1	1	1	1	1	1	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1
<i>Macaca fascicularis</i> -CAEC01	-	-	-	1	0	-	1	-	0	2	1	-	0	-	-	-	0	-	0	-	-	-	0	0
<i>Macaca mulatta</i> -AANU01	-	1	1	1	0	-	-	-	0	5	5	5	3	1	2	1	1	1	1	1	1	1	1	1
<i>Macaca mulatta</i> -AEHK01	1	1	1	1	1	1	1	1	1	-	1	2	0	1	2	-	1	1	1	1	1	1	1	1
<i>Macaca mulatta</i> -JSUE03	1	1	1	1	1	1	1	1	1	3	4	4	2	1	2	1	1	1	1	1	1	1	1	1
<i>Macaca nemestrina</i> -JZLF01	1	1	1	1	1	1	1	1	1	0	4	4	3	3	1	2	1	1	1	1	1	1	1	1
<i>Mandrillus leucophaeus</i> -JYKQ01	1	1	1	1	1	1	1	1	0	3	3	3	2	1	1	2	1	1	1	1	1	1	1	1
<i>Nasalis larvatus</i> -JMHX01	-	-	-	1	0	1	-	1	0	1	1	1	1	2	1	-	-	0	1	1	1	1	1	
<i>Papio anubis</i> -AHZZ02	1	1	1	1	1	1	1	1	1	2	2	2	2	1	2	1	1	1	1	-	1	1	0	0
<i>Ptilocolobus tephrosceles</i> -PDMG01	1	1	1	1	1	1	1	1	1	-	1	1	0	2	2	1	1	1	1	1	1	1	1	1
<i>Rhinopithecus bieti</i> -MCGX01	1	1	1	1	1	1	1	1	-	0	1	2	1	1	2	2	1	1	1	1	1	1	1	1
<i>Rhinopithecus roxellana</i> -JABR01	1	1	1	1	1	1	1	1	0	2	1	1	1	2	2	2	1	1	1	1	1	1	1	1

Due copie di IgE di cui una nel locus IGHE e l'altra in contig remoti.

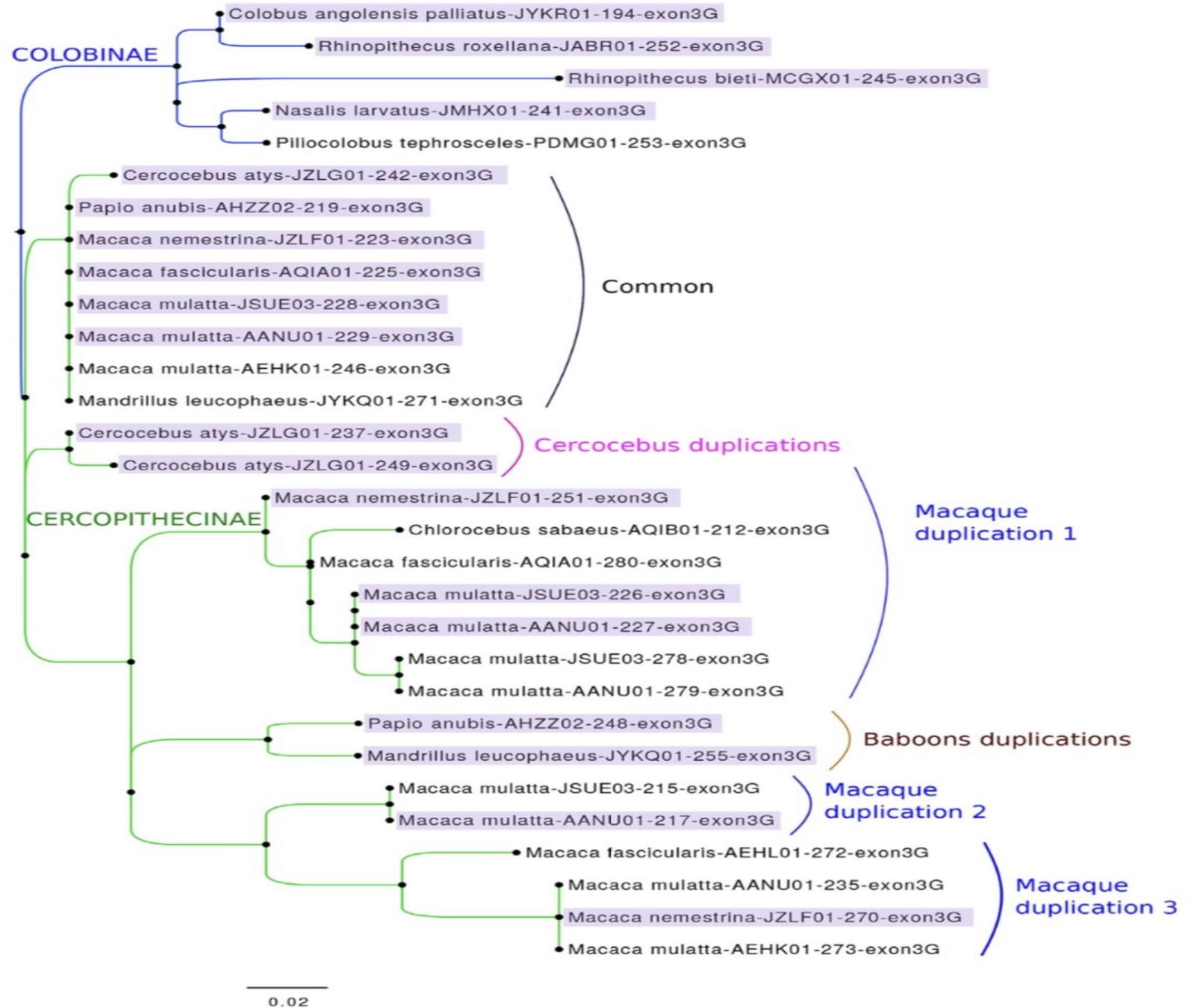
RETROTRASCRIZIONE

una struttura genomica acquisita prima del processo di diversificazione evolutiva.

I Cercopithecoidea si distinguono in due famiglie evolutive, i Cercopithecidae che mostrano duplicazioni del gene IgG e le Colobine che non mostrano quest'alterazione;

differenze nel numero di IgG all'interno della specie *Macaca mulatta*;

presenza di microsatelliti che determina la presenza di contigs non assegnati a nessun cromosoma.



Ominidi

Duplicazioni dei geni IgG, IgA e IgE.

- Nel genoma di *P.troglodytes*, le analisi computazionali due esoni per i domini CH1, CH2 e CH4 dell'IgM;
- sia i gorilla che gli scimpanzè condividono sequenze simili alle quattro IgG degli esseri umani;
- divergenza dell'orango dal resto degli ominidi.

Table 5
The number of exons that code for immunoglobulin CH domains in Hominids.

	IgM				IgD				IgG				IgE				IgA					
	CH1	CH2	CH3	CH4	IgM ^a	CH1	CH2	CH3	IgD ^b	CH1	CH2	CH3	IgG ^c	CH1	CH2	CH3	CH4	IgE ^d	CH1	CH2	CH3	IgA ^e
<i>Gorilla_gorilla-CYU103</i>	1	1	1	1	1	1	1	1	1	5	6	6	2	2	3	2	2	0	1	1	1	1
<i>Gorilla_gorilla-CABD03</i>	1	1	1	1	1	1	1	1	1	4	3	4	2	2	2	2	2	1	1	1	1	1
<i>Nomascus_leucogenys-ADFV01</i>	-	-	1	1	0	1	1	-	0	3	2	2	2	-	1	-	1	0	1	1	2	1
<i>Pan_paniscus-AJFE02</i>	1	1	1	1	1	1	1	1	1	3	6	6	0	1	2	1	2	1	1	1	-	0
<i>Pan_troglodytes-AACZ04</i>	-	-	-	1	0	1	1	1	1	7	5	5	3	1	1	1	2	1	-	1	2	0
<i>Pan_troglodytes-AADA01</i>	1	1	1	1	1	1	1	1	0	3	2	1	1	-	-	-	1	0	-	-	-	0
<i>Pan_troglodytes-NBAG03</i>	2	2	1	2	1	1	1	1	1	7	8	8	7	1	1	1	2	1	2	2	2	2
<i>Pongo_abelii-ABGA01</i>	-	1	1	1	0	2	1	1	0	5	4	4	3	3	3	1	2	1	-	1	1	0
<i>Pongo_abelii-NDHI03</i>	1	1	1	1	1	1	1	1	1	4	4	5	4	4	4	2	3	2	3	3	3	3

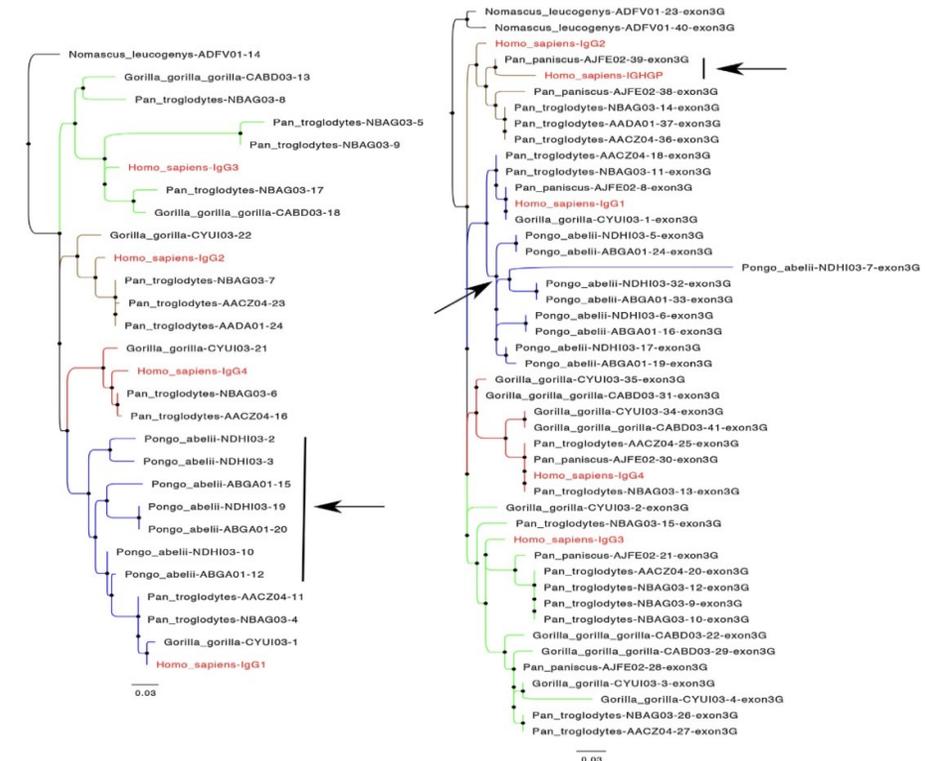
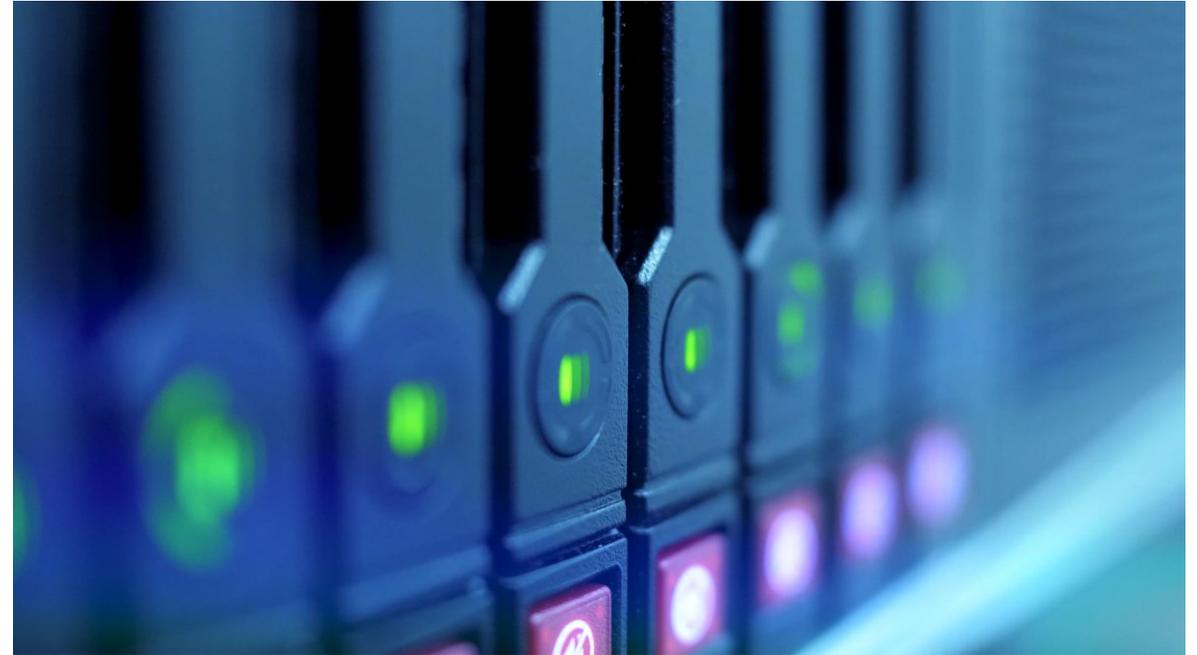


Fig. 7. The tree constructed by aligning the complete IgG sequences (left) and CH3 sequences (right) detected with the *CHfinder* program. To act as markers, the sequences of *H. sapiens* (red) were aligned with the other primate sequences. In each tree, the sequences of *P. abelii* are located in the same clade as the human IgG1 sequences. The human IGHP sequence is also indicated to show its origin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

JGI Genomes OnLine Database (GOLD)

Si tratta di una raccolta aggiornata quotidianamente di progetti genomici accumulati da tutto il mondo.

E' organizzato in quattro livelli (Studi, Organismi, Progetti di Sequenziamento e Progetti di Analisi) e questo database fornisce un'interfaccia web interattiva per la navigazione e la ricerca in un'ampia gamma di progetti e metadati.



Discussione

L'utilizzo di specifici software nell'ambito della biologia computazionale, ha permesso di evidenziare la presenza e struttura dei geni che specificano le varie Ig nei diversi organismi viventi. Considerando l'organizzazione strutturale delle immunoglobuline, il locus delle regioni costanti di queste proteine è cambiato in modo significativo nelle linee evolutive dei vertebrati.

Le regioni costanti degli anticorpi, portano funzioni effettrici che sono diverse in ogni classe di immunoglobuline. Quindi, la conoscenza dettagliata dei cambiamenti evolutivi, cioè generazione o perdita delle classi di Ig e quindi delle diverse sequenze geniche codificanti (esoni), è di grande interesse nell'ambito dell'immunologia comparata.

Riferimenti:

- ❑ Murphy, K., Weaver, C., 2016. Janeway's Immunobiology. Garland Science.
- ❑ Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csàrdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al., 2011. The evolution of gene expression levels in mammalian organs. Nature 478, 343.
- ❑ Flanagan, J., Rabbits, T., 1982. Arrangement of human immunoglobulin heavy chain constant region genes implices evolutionary duplication of a segment containing gamma, epsilon and alfa genes. Nature 300, 709.
- ❑ David N. Olivieri, Francisco Gambòn Deza., 2018. Molecular Immunology 353, 363.
- ❑ Daniel, J., Rigden, X., Fernàndez., 2021. Nucleic Acids Research 723, 733.