Evolution of the immune system from model organism to man

Tim Hulsen[1], Wilco W.M. Fleuren[1], Peter M.A. Groenen[2]

[1] CMBI, NCMLS, Radboud University Nijmegen Medical Centre, Nijmegen, NL [2] MDI, NV Organon, Oss, NL

Introduction

The immune system is of major importance since it protects metazoans from infection by pathogenic organisms. Throughout evolution, two major branches have originated: innate and adaptive immunity. The innate immune system exists in a wide range of metazoans, whereas the adaptive immune system is only present in jawed vertebrates. Both the innate and the adaptive immune system are intensively studied by scientists working in the field of drug discovery, since numerous drugs are active in immunologic pathways. However, immunologic drug discovery is difficult since there are sometimes large differences in drug response between model organisms and man. These differences might be explained by studying the evolution of genes involved in the immune system. Here we present an overview of the evolution of the immune system from several model organisms to man, using whole-genome data from a wide range of species.

Figure 1. The Immunophyle web interface.

Methods

We used Ensembl v41 as a starting point for our immunogenomics analysis. This database contains in total 553,721 genes from 26 species: 1 yeast, 6 invertebrate animals, 7 vertebrate non-mammals and 12 mammals, under which numerous species often used as model organisms for man: fruitfly, mouse, rat and macaque. We built phylogenetic lineages, i.e. orthologous groups, using a simple single linkage clustering, in the same way as for the web application PhyloPat [1]. In order to get a immune-specific data set, we gathered all HUGO gene names included in the IRIS database [2]. All phylogenetic lineages connected to one or more of the 1551 immunologic HUGO names were stored in a separate database, named ImmunoPhyle. This database now includes 18,933 genes from the 26 species, including 1,157 genes from H. sapiens. Results are displayed in order from the 'lowest' species S. cerevisiae to the 'highest' species H. sapiens ('low'/'high' corresponding to the longest/shortest evolutionary distance to man). We make use of the classification into 22 categories provided by the IRIS database (table 1). All data is available through the web application Immunophyle (http://www.cmbi.ru.nl/immunophyle, figure 1) [3].

Nr.	Abbrev.	Description	# HUGO IDs	# ImmunoPhyle lineages	# Genes
1	InImm	Innate Immunity	638	272	8640
2	InFlm	Inflammation	314	117	4568
3	Chmtx	Chemotaxis	192	54	2374
4	Phago	Phagocytosis	37	17	890
5	Compl	Complement	62	33	958
6	Cy_Ch	Cytokines and Chemokines	261	109	2947
7	AdImm	Adaptive Immunity	422	140	4983
8	ClRsp	Cellular Response	145	63	2358
9	HmRsp	Humoral Response	98	34	1087
10	BMImm	Barrier and Mucosal Immunity	45	18	713
11	DevIp	Development of Immune System	130	50	2044
12	AgPrc	Antigen Processing	148	31	830
13	PtSig	Immune Pathway or Signalling	470	224	8245
15	Recpt	Receptor	246	118	3506
16	IndIm	Induced by Immunomodulator	200	86	3487
20	ImDef	Involved in Immunodeficiency	71	30	1013
21	AutIm	Involved in Autoimmunity	44	19	530
22	ExplT	Expressed Primarily in Immune Tissues	332	134	3970
23	Other	Other	107	43	1843
25	InKil	Innate NK Killing	82	33	1015
26	RIDis	Related to Disease	172	91	3141
27	Coagl	Coagulation	111	51	2624
0	All	All immunologic lineages	1542	585	18933

Table 1. The IRIS categories linked to the phylogenetic lineages. Green: three largest categories. Red: three smallest categories.

Results

Table 2 shows how many genes are linked to each category, for each of the 26 species in our dataset. From this table, it is obvious that the immune system is largely restricted to vertebrates: *Tetraodon*, the first vertebrate in the list, contains almost four times as many immunorelated genes as *Ciona*, the last non-vertebrate in the list. This can also be concluded from fig. 2, which shows an analysis of the species occurrence in the phylogenetic lineages. The largest differences can be seen in the transition from invertebrates (*C.int*) to vertebrates (*T.nig*) and from non-mammals (*G.gal*) to mammals (*M.dom*), depicted by arrows. Moreover, this figure shows that the opossum, elephant and rabbit have a large number of deletions. This probably points to the lesser quality of the genome assembly rather than to any real evolutionary deletions.

Cat.	Sc	Ce	Ag	Aa	Dm	Cs	Ci	Tn	Tr	Ol	Ga	Dr	Xt	Gg	Md	Dn	Bt	Cf	Et	La	Rn	Mm	Oc	Mm	Pt	Hs	Total
InImm	17	51	77	89	81	81	93	351	355	339	351	420	304	295	466	355	566	435	416	384	517	571	539	384	535	568	8640
InFlm	13	34	45	57	43	53	55	202	200	194	197	237	179	150	227	200	267	221	215	197	263	302	271	194	265	287	4568
Chmtx	4	12	18	24	18	22	28	107	118	112	122	125	96	69	157	90	135	121	112	103	124	147	132	86	141	151	2374
Phago	1	4	9	10	10	8	10	46	43	41	47	50	32	31	42	34	51	45	46	42	49	58	44	33	51	53	890
Compl	0	3	13	7	7	11	19	45	41	43	43	54	37	31	50	36	58	45	48	34	60	62	55	43	54	59	958
Cy_Ch	2	11	14	20	18	18	18	122	119	124	119	148	92	120	144	106	219	173	143	133	175	187	195	143	190	194	2947
AdImm	17	44	37	40	48	59	62	212	207	204	219	253	158	170	246	188	330	260	225	223	276	315	244	324	303	319	4983
ClRsp	6	26	20	23	22	36	41	106	101	102	100	119	78	96	116	93	138	112	105	104	124	148	148	111	137	146	2358
HmRsp	3	9	8	8	9	8	9	48	46	48	45	47	37	40	49	43	60	58	55	50	61	65	76	68	65	72	1087
BMImm	0	1	10	9	15	2	4	20	25	17	27	24	18	11	33	30	68	42	38	32	48	58	47	25	52	57	713
DevIp	5	18	23	25	23	22	29	109	90	89	96	124	64	72	106	74	109	108	103	86	114	122	116	92	108	117	2044
AgPrc	3	8	9	11	11	10	12	34	31	36	38	56	22	25	39	40	49	35	37	36	39	40	63	35	54	57	830
PtSig	13	63	70	87	81	93	102	400	381	382	390	480	301	296	446	302	454	415	371	344	459	508	489	337	480	501	8245
Recpt	2	18	16	20	18	18	24	148	151	150	158	187	125	124	165	141	205	191	170	154	227	240	226	156	231	241	3506
IndIm	7	23	28	25	40	29	32	172	163	159	175	200	129	122	171	130	224	184	154	154	198	218	197	151	193	209	3487
ImDef	4	8	15	12	9	18	28	44	44	41	38	64	35	42	48	34	61	45	45	43	54	56	58	52	56	59	1013
AutIm	0	1	12	6	1	5	3	23	24	26	23	32	18	20	25	19	29	30	28	22	29	30	31	29	31	33	530
ExplT	11	22	36	27	31	32	42	157	160	153	173	186	137	118	249	155	238	201	179	170	257	274	260	158	261	283	3970
Other	9	28	32	31	37	23	25	99	82	80	90	105	86	78	84	78	98	82	75	74	93	92	96	73	94	99	1843
InKil	1	4	9	9	6	6	8	31	37	28	32	35	44	23	38	49	70	49	52	50	74	88	72	38	80	82	1015
RIDis	6	14	32	28	32	25	34	151	143	133	144	176	115	128	159	131	190	159	153	133	170	184	184	145	182	190	3141
Coagl	5	25	36	44	33	31	33	132	123	122	124	154	114	81	124	108	141	123	121	109	145	162	139	106	138	151	2624
All	54	156	192	211	214	219	239	876	830	824	855	1015	686	685	969	740	1121	948	870	802	1070	1163	1131	818	1087	1157	18933

Table 2. Numbers of genes per category and per species. Green: three highest numbers within each category. Red: three lowest numbers within each category.

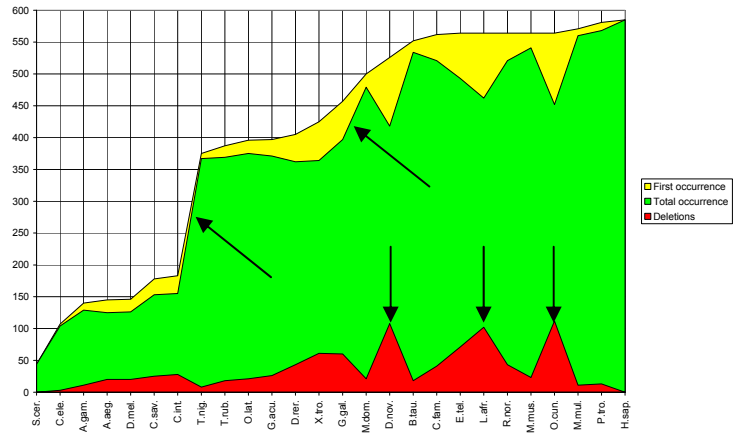

Figure 2. Analysis of the occurrence of the 26 species in the 585 ImmunoPhyle phylogenetic lineages. Yellow: nr of lineages that started in the corresponding species, or earlier. Green: total nr of lineages which contain gene(s) from the corresponding species. Red: nr of deletions in the corresponding species.

Discussion

We give the first real overview of the molecular evolution of the immune system from model organisms to man. Our analysis gives general insights in this evolution and offers a framework for further investigation of interesting observations. General trends, such as the emergence of the adaptive immune system and the decline of the innate immune system, can be observed very easily. As seen in some case studies (data not shown here), this approach can also be used to zoom in on specific gene families or pathways. However, in order to explain differences in drug response between a certain model organism and man, usually more data is needed than just orthology data. A combination of orthology data, expression data, protein interaction data and structural data as used in recent other studies might help solving the problems that are encountered when transferring experimental results from model organism to man.

References

1. Hulsen T, de Vlieg J, Groenen PM: **PhyloPat: phylogenetic pattern analysis of eukaryotic genes.** *BMC Bioinformatics* 2006, 7:398.
2. Kelley J, de Bono B, Trowsdale J: **IRIS: a database surveying known human immune system genes.** *Genomics* 2005, 85(4):503-511.
3. Hulsen T, Fleuren WWM, Groenen PMA: **Evolution of the immune system from model organism to man.** *Manuscript in preparation*.