Evolutionary and polymorphism analyses reveal the central role of BTN3A2 in the concerted evolution of the BTN3 gene family

Hassnae Afrache, Pierre Pontarotti, Laurent Abi-Rached, Daniel Olive

Evolutionary and Polymorphism Analyses Reveal the Central Role of *BTN3A2* in the Concerted Evolution of the *BTN3* Gene Family

Hassnae Afrache[1], Pierre Pontarotti, Laurent Abi-Rached[3*], Daniel Olive[1 *]

[1] Centre de Recherche en Cancérologie de Marseille (CRCM), Inserm, U1068, Marseille, F-1300 9, France, CNRS, UMR7258, Marseille, F-13009, Institut Paoli-Calmettes, Marseille, F-13009, F rance, Aix-Marseille University, UM 105, F-13284, Marseille, France

[2] Institut de Mathématiques de Marseille, UMR 7373

[3] ATIP CNRS, Institut de Mathématiques de Marseille, UMR 7373

[*] **lar and do are co-senior authors**

Correspondence should be addressed to Daniel Olive, e-mail daniel.olive@inserm.fr

Conflict-of-interest disclosure: Olive D. is a founder and shareholder of Imcheck Therapeutics (Marseille, France).

Running title: evolution of BTN3A gene family

Abstract (230 words)

The butyrophilin 3 (BTN3) are receptors of the immunoglobulin superfamily implicated in the T lymphocytes regulation and present a wide plasticity in mammals, being absent in rodent but present with three copies in human. In order to understand how these genes have been diversified and what forces guiding this diversification, we studied their evolution and show that the three human *BTN3* genes are the result of two successive duplications in Primates and that the three

genes are present in Hominoids and the Old World Monkey groups. A thorough phylogenetic analysis reveals a concerted evolution of *BTN3* characterized by a strong and recurrent homogenization of the region encoding the signal peptide and the IgV domain in Hominoids. During these homogenizations the sequences of *BTN3A1* or *BTN3A3* are replaced by *BTN3A2* sequence. In human, the analysis of the diversity of the three genes in 1683 individuals representing 26 worldwide populations shows that the three genes are polymorphic, with more than 46 alleles for each gene, but they are also marked by extreme homogenization of the IgV sequences. The same analysis performed for the *BTN2* genes that are also diversified in Primates, shows also a concerted evolution, however it is not as strong and recurrent as for *BTN3*. This study shows that BTN3 receptors are marked by extreme concerted evolution at the IgV domain and that *BTN3A2* plays a central role in this evolution.

Introduction

The butyrophilins 3 are receptors of the Immunoglobulins superfamily ubiquitously expressed on T and B lymphocytes, monocytes, NK cells and dendritic cells(1-3). They form a group of three proteins BTN3A1, 3A2 and 3A3 with an extracellular region composed of two domains IgV and IgC extremely conserved between the three proteins(4) but with a significant difference at the cytoplasmic region where BTN3A1 and 3A3 contain a B30.2 type domain which is absent in BTN3A2. BTN3 are also butyrophilin family members, a multigene family which is structurally homolog at the extracellular level to the co-signaling B7 family(5) and with several members implicated in immune response regulation. It is the case of Btnl2 regulating the murine T lymphocytes activation and that seems to be a regulator of the intestinal mucosal immunity(6, 7).

The BTN3 are also important regulators of the immune response. Their function, initially suggested by specific stimulations that could inhibit or activate T lymphocytes in function of the antibodies used(2, 3, 8), is currently clearly described for a specific cytotoxic T lymphocytes population, the Vg9Vd2 T cells(9). The activation of Vg9Vd2 T cells requires the presence of phosphoantigens (pAg), phosphorylated metabolites accumulated in infected and tumor cells(10). The role of each BTN3 has been investigated through shRNA invalidation and the use of B30.2 mutants and BTN3A1 has been identified as a key player in this activation which is B30.2 dependent(9). However, the function of the others BTN3 remains elusive. A recent structural and functional study has shown that BTN3A1 B30.2 domain binds directly the pAg through a positively charged surface pocket that contains a Histidine residue critical for the binding and which is absent in BTN3A3. When introduced into the B30.2 domain of BTN3A3, the Histidine confers pAgs binding ability and Vg9Vd2 T lymphocytes activation(11). Another mechanism of activation has been described: a direct binding of the pAg by the IgV domain of BTN3A1 describing a stimulation model analog to the classical antigen presentation by CMH molecules (12).

At the genomic level, the *BTN3* are located at the extended HLA class I region on the chromosome 6 within a gene cluster composed of four additional *BTN* genes: *BTN1* and three *BTN2 (2A1-2A3)*. Interestingly, this cluster is plastic in mammals and *BTN3* are present in one copy in horses while absent in rodents(13). Additionally, association studies in human show that genes like *BTN3A2* are polymorphic with some SNP that are related to susceptibility to diseases like type 1 diabetes(14). Similarly, the polymorphism of other *BTN* such as *BTN2A1* is related to susceptibility to type 2 diabetes (15), chronic kidney disease (16, 17), metabolic syndrome (18), dyslipidemia (19) or myocardial infarction Moreover (20, 21). A recent study has described the

association of the polymorphism of the *3A3-2A1* region with rubella virus-specific cellular immunity following vaccination (22). Given the importance of *BTN3* for the immune response and their apparent plasticity in mammals and in human populations, we conducted a detailed study of the evolution of these genes in primates and their polymorphism in human populations.

Material and methods

Data for the *BTN3* evolution analysis

Sequences used for the analysis were obtained by BLAST search of 'nr', 'est', 'htgs' and 'wgs' databases of the National Center for Biotechnology Information (NCBI).

For the analysis of the *BTN3* transcripts, the *Gorilla gorilla BTN3A1* sequence was incomplete and was re-predicted with the program FGENESH^{+} (23) using the human protein sequence as reference. Similarly, *Colobus guereza BTN3* sequences were predicted using the BAC (Bacterial Artificial Chromosome) genomic sequences CH272-166E5 (Genbank accession number AC210406).

BTN3 genomic sequences were obtained from the NCBI "Gene" database. The *BTN3A1* sequence of *Gorilla gorilla* and *Saimiri boliviensis* and the *BTN3A3* sequence of *Macaca fascicularis* were of low quality (large non-sequenced regions) and were excluded. As the New World Monkey group contained only one genomic sequence, the *BTN3* evolution could not be studied in this taxonomic group.

Data for the *BTN2* evolution analysis

Two of the three *BTN2* genes were analyzed (*BTN2A3* was excluded as it is a pseudogene) and the sequences were obtained as it was described for *BTN3* sequences.

The *BTN2A1* transcript sequence of *Saimiri boliviensis* was incomplete and was re-predicted using genomic sequences similarly to the prediction of the *Gorilla gorilla BTN3A1* sequence.

The quality of the two BTN2 genomic sequences of *Callithrix jacchus* was verified using the transcribed sequences available in the NCBI 'est' database and reassembling the genomic sequences generated within the genome sequencing project of *C.jacchus* by the 'Washington University Genome Sequencing Center' and the 'Baylor College of Medicine'. For the latter analysis, BLAST searches were made of the NCBI 'Trace Archive' database using both *BTN2* genomic sequences of *C.jacchus*: over 170 sequences have been obtained and reassembled with the 'STADEN package'(24).

Phylogenetic analysis

Nucleotide sequences were aligned using MAFFT (25) and QuickProbs (26) and corrected manually. Phylogenetic analyses were conducted using three methods: Neighbor Joining (NJ), Parsimony and Maximum likelihood (ML). The NJ phylogenies were performed with MEGA6 (27) using the Tamura-Nei method with 500 replicates. For parsimony analyses, PAUP*4.0b10

(28) and the tree bisection-reconnection branch swapping algorithm were used with 500 replicates and a heuristic search. RAxML8 (29) was used to perform ML phylogenies with the GTR+G model with 500 replicates (rapid bootstrapping).

***BTN2* and *BTN3* genotyping data in human populations**

Exome sequencing data generated by the 1000 Genomes project (30) were analyzed for 1,683 individuals representing 26 populations. RNA sequencing data generated by the European medical sequencing consortium GEUVADIS (31) were downloaded from the website http://www.geuvadis.org/web/geuvadis/RNAseq-project. They correspond to mRNA sequencing data of 295 among the 1683 individuals analyzed for exome data.

The 26 populations were divided into five groups: African origin populations, European origin populations, South Asian origin populations, East Asian origin populations and populations of the Americas (Table SI).

In silico* genotyping of *BTN2* and *BTN3

The genotyping of the five BTN genes analyzed (*BTN3A1-3A3*, *BTN2A1-2A2*) was performed as previously described (*BTNL2 article*). Compared to this approach, an additional analysis was developed for *BTN3* genotyping as the first analyzes showed the presence of a highly conserved region at the exon encoding the IgV domain. The presence of such a region leads to loss of coverage because the sequences are eliminated to the specificity step since they are not specific to a single gene. In order to verify for each individual, if the low covered or non-covered region contains SNPs or not, a "combined" analysis was performed: it consists in using the sequences *BTN3A1, 3A2* and *3A3* for the conserved region to identify all the variations. The results of the combined analysis were compared to the result of the individual analysis for each gene and the

cases where a variation is present in the combined analysis but absent in the individual analysis were isolated. They correspond to variations missed in the individual analysis. 172 problematic cases were detected: for 35 cases, RNA sequencing data of the same individuals were useful to resolve the genotype. For the remaining 141 genotypes (including 28 genotypes partially resolved: one allele identified) they were resolved statistically (see *BTN3* genotypes phasing).

***BTN2* and *BTN3* genotypes phasing**

Genotypes phasing was performed using PHASE 2.1 (32, 33). The input files were prepared using SNPtools (34). Information for associated SNPs was specified using the option –k. Unresolved genotypes in the *BTN3* conserved region were determined statistically using "?" in place of the missing position. The reconstruction of allele sequences was done using reference coding sequence of each gene: ENST00000312541 (*BTN2A1*), ENST00000244513 (*BTN2A2*), ENST00000289361 (*BTN3A1*), ENST00000356386 (*BTN3A2*) and ENST00000244519 (*BTN3A3*).

Results:

Primates have variable number of *BTN3* genes

In order to identify the relation between Primates *BTN3* sequences, a phylogenetic analysis of coding sequences has been conducted. The three *BTN3* genes are present in the Catarrhini group: Hominoids and Old World monkeys (**Figure 1A**) whereas the Prosimians group represented by *Otolemur garnettii* has only one ortholog and the New World monkeys (NWM) group has two *BTN3* genes: one gene ortholog to *BTN3A1* and one co-ortholog to *3A2* and *3A3*. Thus, the three human *BTN3* genes are the result of two successive duplications: a first duplication at the basis of

153 the Simiiformes group gave rise to *BTN3A1* and an ancestral sequence of *3A2* and *3A3* and a
154 second duplication at the basis of the Catarrhini group gave rise to *3A2* and *3A3*.

155 The domain architecture of the three BTN3 is conserved in the Catarrhini group with two
156 extracellular domains IgV and IgC, a transmembrane domain TM and a B30.2 intracellular
157 domain which is absent in 3A2 sequences (**Figure 1E**). The presence of the B30.2 domain in the
158 *O.garnettii* and the NWM sequences shows that the ancestral structure of the three BTN3 is IgV-
159 IgC-TM-B30.2 and that the B30.2 domain has been lost specifically in BTN3A2 of the Catarrhini
160 group.

161 **The BTN3 IgV domain in Hominoids has undergone several homogenizations**

162 Interestingly, a domain by domain phylogenetic analysis has shown that the BTN3 domains
163 evolved differently (**Figure 1A-D**). While the phylogenies of the IgC, the TM and the B30.2
164 domains were compatible with the full sequence phylogenies, the phylogeny of the IgV domain
165 shows the presence of paralog groups in human, orangutan and gibbon (**Figure 1B**), which mean
166 that at the IgV domain level, the paralog sequences *BTN3A1, 3A2* and *3A3* are more closely
167 related to each other than they are to their orthologs counterparts in a closely related species.
168 Such phylogenetic profile suggests homogenizations of the IgV sequences have occurred in each
169 of these species. The absence of this phylogenetic profile in the OWM group suggests this
170 homogenization is specific to Hominoids. As the NWM group contains few sequences, the
171 interpretation was difficult even if the presence of a group with *BTN3A1* and *3A2/3A3* sequences
172 is compatible with an ancient homogenization (**Figure 1B**).

173 The analysis of the groups present in the IgV phylogeny shows that all Hominoids *BTN3A1*
174 sequences are grouped in the *BTN3A2/3A3* group (**Figure 1B**), suggesting the homogenizations
175 replaced the *BTN3A1* IgV sequences with the IgV sequences of *3A2* or *3A3*. In order to identify

more precisely the relations between the sequences and the regions homogenized, we extended the analysis to the genomic sequences.

The homogenization targets different regions in the three *BTN3* genes

The phylogenetic analysis of the genomic regions (introns/exons) was performed using variable size windows and reveals different relations between the *BTN3* groups depending on the genomic region considered (**Figure S1**). Particularly, this analysis shows that *BTN3A3* is a chimeric gene resulting from a recombination between the 5' region of a *BTN3A1-like* gene and the 3' region of a *BTN3A2-like* gene, with a recombination zone located at the UTR regions between the positions 4,000 and 4,025 (**Figures 2A, S1B-C**).

The homogenization detected at the IgV level with the coding sequences analysis was also confirmed. However, it is not limited to the IgV exon but covers a large region. In fact, we detected a specific homogenization between *BTN3A2* and *BTN3A3* that starts at the 5' regulatory regions (the second UTR exon) and extends at the 3' region up to about 100 bp after the IgV exon (**Figures 2B, S1D-G**). A second homogenization region has been detected between *BTN3A2* and *BTN3A1* that starts at the 5' side of the IgV exon and continues up to about 200 bp after this exon (**Figure S1H-I**). In both cases the topologies of the phylogenetic trees allow the identification of the homogenization direction and show that *BTN3A2* is the donor gene.

The homogenization between *BTN3A2* and *3A3* in the Hominoids targets a part of the 5'UTR and the signal peptide (**Figure 2B**) leading to highly similar protein sequences at the peptide signal level but divergent from those of *BTN3A1* (**Figure 2C**). However, the *BTN3A2* and *3A3* sequences of the OWM group, which are not homogenized, are different. Similarly, the homogenized IgV domain between the three *BTN3* is conserved in the Hominoids group while the sequences of the OWM group, which are not homogenized, are different (**Figure 2C**).

Although the Hominoids and the OWM groups contain the three *BTN3* genes, this analysis shows the presence of a homogenization phenomenon targeting specifically the IgV domain in the first but not the second group (**Figure 3**). We were interested to verify whether such homogenization exists in human populations.

The human BTN3 are polymorphic but with homogenized IgV sequences

To investigate the diversity of human *BTN3* we re-analyzed the exomes sequencing data generated by the 1000 Genomes consortium for >1,680 individuals representing 26 populations. Careful reconstruction of allelic sequences of the three *BTN3* genes reveals that *BTN3* are polymorphic with, 57, 47 and 61 alleles encoding 33, 23 and 34 allotypes for BTN3A1, 3A2 and 3A3 respectively (**Figure 4A**). However, the majority of allotypes are uncommon with only 8, 5 and 2 allotypes having a frequency > 1% in a region of the world for BTN3A1, 3A2 and 3A3 respectively (**Figure 4A, D, F and H**).

The comparison of the *BTN3* allelic sequences shows that the non-synonymous variations are more concentrated on the IgC domain, in particular for *BTN3A2*, and on the cytoplasmic region (**Figures 4C, E, G**). On the contrary, the IgV domain is highly conserved: for example there is no common allotypes (frequency ≥ 1% in a region of the world) for BTN3A2 and 3A3 with a variation on the IgV domain (**Figure 4C, E and F**). Confirming this conservation are the results of the non-synonymous nucleotide diversity analysis for the IgV, IgC and B30.2 exons between alleles of each gene. We observed a loss of diversity for the IgV exon which is extremely striking between *BTN3A1* and *3A2* (**Figure 4B**). Indeed, while *BTN3A1* and *3A2* have duplicated tens of millions of years ago, the IgV diversity for *3A1-3A2* pair is of the same order of magnitude as the allelic diversity of each gene and about three orders of magnitude lower than the *3A1-3A2* diversity for IgC. Similarly, the other pairs (*3A1-3A3* and *3A2-3A3*) and *3A1-3A2-3A3* trio have

diversity for the IgV lower by about one order of magnitude compared to the diversity for the IgC.

Thus, these data show the variation between *BTN3A1* and *3A2* at the IgV level is of the same order of magnitude as the allelic variation of each gene for this domain. This result is consistent with an active homogenization in humans of the IgV between these two *BTN3*. This model is also supported by the existence of several variable positions that are shared between *BTN3A1* and *3A2* or *3A3* or represent the reference position in *3A2* and / or *3A3* (**Table I**). This is the case of the position -14 in the signal peptide and 37 and 39 in the IgV which are polymorphic in *BTN3A1* but conserved in *3A2* and / or *3A3*. The positions -14 and 39 are very common and are carried by alleles in different populations while position 37 is more specific to populations of African origin. Taken together this strongly suggests that there is always an exchange between genes in this region and that the historical homogenization detected in hominoids still exists in human populations.

The homogenization characterizes also the Primate *BTN2* genes evolution

In order to verify if the homogenization observed for *BTN3* genes was unique to this group, we preformed the same analyses for *BTN2* genes, another phylogenetic group with more than one gene. We first analyzed the relation between the two human *BTN2* genes (BTN2A1 and 2A2) and their counterparts in primates and rodent: the two genes were present in the different primate groups while rodents have only one copy of the gene (**Figure 5A** and **S2**). The two *BTN2* encode for protein with two extracellular domains IgV and IgC separated from a B30.2 cytoplasmic domain with a transmembrane region that differs between the two genes with *BTN2A2* having two TM domains in the Simiiformes group.

The phylogenetic analysis of coding sequences shows the presence of a homogenization between the two genes in the Prosimians and the NWM groups (**Figure S2**). This result was confirmed by the genomic sequences analysis which illustrates three homogenizations: one at the basis of the Catarrhini group, one in the Prosimians group and one in the ancestor of the NWM group (**Figure 5B and S3**). These homogenizations have specifically target the exons encoding the signal peptide, the IgV domain and the B30.2 domain (three homogenizations) but not the exon encoding the TM and a part of the cytoplasmic region (**Figure S3**). Interestingly, sequence alignment of the TM region reveals the presence of an important divergence between the two BTN2 at the beginning and at the end of the TM exon. In fact, *BTN2A2* contains an insertion of 39 bp at the 5' side of the exon and a deletion of 54 bp at the 3' side compared to *BTN2A1*. This difference leads to a large transmembrane region predicted to contain two TM and a shorter cytoplasmic region upstream B30.2 (**Figure 5C**). Only the divergence at the 3' side of the TM exon has been conserved between Prosimians *BTN2* sequences.

Similarly to *BTN3*, the IgV domain has been targeted by several homogenizations in Primates. in order to test if, like for human *BTN3*, the IgV domain is conserved between the two *BTN2*, the allelic diversity has been characterized in the same populations as for *BTN3* and nucleotide diversity for the IgV, IgC and B30.2 domains has been estimated for each *BTN2* gene as well as for the combination *BTN2A1-2A2* (**Figure 5D**). We found that the IgV domain presents the lowest nucleotide diversity among the three domains for the analysis of individual genes. However, the diversity for the *2A1-2A2* analysis for the IgV domain (**Figure 5D**) is much higher (between one and three orders of magnitude) than that observed for the combined analysis between the three *BTN3* (**Figure 4B**), despite the fact that the same analysis to the IgC domain produces comparable results for *BTN2* and *BTN3*.

Taken together, these data demonstrate that the homogenization is not limited to *BTN3* group but seems to be characteristic of duplicated *BTN* groups. However, the homogenization is much less pronounced for *BTN2* and concerns the extracellular and intracellular region, while for *BTN3* it is recurrent in hominoids with extreme homogenization at the IgV domain level especially for human *BTN3A1* and *3A2*.

Discussion

Our study of the emergence and diversification of *BTN3* in Primates shows that these genes are characterized by recurrent homogenization between paralog genes. This type of evolution called concerted evolution is concentrated on a large genomic region from the 5' UTR to the end of the IgV exon. Although the three *BTN3* are present in both OWM and Hominoids, the homogenization is observed only in hominoids suggesting the presence of specific evolution constraints on *BTN3* in this taxonomic group. Moreover, these constraints seem to be different between the three genes as the homogenization targets distinct regions. In fact, for *BTN3A2* and

3A3 the homogenization is concentrated on the 5'UTR region and the exon encoding the signal peptide. The recurrence of this phenomenon in several species (human, orangutan and gibbon) suggests that strong selection pressure maintains the homogenization of the regulatory elements of the expression these two genes and cellular distribution of the proteins. However, for the IgV domain the homogenization concerns the three *BTN3* and it is particularly extreme for *BTN3A1* and *3A2* in human which explains the apparent conservation of the IgV of the human BTN3. Given that the IgV domain is implicated in the interaction with ligand at the cell surface suggests the three proteins interact with the same ligand.

Interestingly, our data show that *BTN3A2* is a donor gene in the historic homogenizations in Hominoids but also in human populations with some variations that are polymorphic in *BTN3A1* but conserved in *BTN3A2*. These data are consistent with a model where *BTN3A2* evolves with a partner that varies and then *BTN3A2* adaptations arising spread to the other two genes through the homogenization. The major difference between BTN3A2 and the other BTN3 is the absence of the cytoplasmic domain B30.2 indispensable for the function of the other BTN3. In fact, the well-known function of the BTN3 is the stimulation of the TVg9Vd2 cytotoxic lymphocytes in presence of phosphoantigens (pAgs), phosphorylated metabolite accumulated in infected and stressed cells. This stimulation is dependent of BTN3A1 and its B30.2 domain (9) that can bind directly the pAg(35). Previous study has shown that the stimulation of BTN3 with monoclonal antibodies specific of the extracellular region of the three proteins increased the cytokine production by $TCD4^+$ ant $CD8^+$ lymphocytes but not by NK cells. This has been explained by the differential expression level of the transcripts of the three genes: while *BTN3A1* was majorly expressed in different T cells *BTN3A2* was most expressed in NK cells. Furthermore, stimulation assays using cell lines expressing NKp-30 receptor, implicated in interferon gamma (IFNg)

311 production, and transfected with *BTN3A1* or *BTN3A2* have shown differential regulation of
312 cytokine production depending on the BTN3 expressed. The co-stimulation with anti-NKp-30
313 and anti-BTN3A2 antibodies decreases the IFNg production while the co-stimulation with an
314 anti-BTN3A1 antibody increases the IFNg production(2).

315 Thus, *BTN3A1* and *3A2* seem to have distinct structures and functions but their evolution is
316 extremely correlated at the IgV domain level, typical characteristics of paired receptors. Paired
317 receptors are closely related receptors including inhibitors and activators members and which are
318 encoded by genes organized into clusters(36). They are characterized by very similar
319 extracellular regions with distinct cytoplasmic regions leading to opposite intracellular
320 signals(37). Concerted evolution is responsible for the evolution of certain paired receptors
321 interacting with ligands that are rapidly evolving like the family of KLRC (killer cell lectin-like
322 receptors subfamily C) receptors in Primates implicated in the regulation of the activation and the
323 cytotoxicity of NK cells(38). Although the BTN3 ligand is not yet identified, these data suggest
324 that depending on the cell types BTN3A2 would be a receptor with opposite cellular effects of
325 those of 3A1 or a decoy receptor interacting with the ligand but with no intracellular signaling.
326 Thus, BTN3A1 and 3A2 would be of potential receptors evolving in pairs in a concerted manner
327 at the level of their IgV domain.

328 The presence of homogenization between *BTN2*, which is less recurrent than that observed for
329 *BTN3*, shows that concerted evolution is not limited to *BTN3* but is rather a characteristic of the
330 evolution of *BTN* groups that duplicated. The major difference between *BTN2* at the TM exon
331 which is accompanied by a lack of homogenization in all species requires further investigation to
332 verify the impact of the presence of such a broad membrane region for BTN2A2 and which is
333 associated with a shorter cytoplasmic region upstream of the B30.2 domain, on its expression,

structure and function. MOG is a member of the BTN family that has a second TM domain: this domain does not cross through the membrane but is semi-integrated into the lipid bilayer, and lying partially buried in the membrane, exposing the C-terminal part of the protein lacking the B30.2 domain(39). Thus, the second TM of BTN2A2 could adopt the same conformation or entirely cross the membrane to expose the B30.2 domain at the cell surface. The absence in the galago (*O.garnettii*) of the difference simple versus double TM, conserved in other primates, suggests that it is not as critical as the difference in the short cytoplasmic region upstream of the B30.2 domain which is conserved in all species studied. This region upstream of the B30.2 domain is present in all *BTN* having a B30.2 exon (*BTN3A2* does not have a B30.2 domain due to an insertion of an Alu repeat sequence interrupting the reading frame(40)) with a length varying between groups. However, no function has been assigned to this region and the functional data described so far involving the cytoplasmic region are all related to the B30.2 domain(9, 35).

In conclusion, we have shown that concerted evolution by homogenization marks the *BTN3* history. These data pave new ways for studying the regulation of the BTN3A2 and 3A3 expression or the potential function of BTN3A2 and 3A1 as paired receptors interacting with the same ligand via the IgV domain but with different cellular effects across the cytoplasmic region.

Aknowledgements : DO laboratory is supported by Fondation pour la Recherche Médicale (Equipe FRM DEQ20140329534). DO is a senior scholar of the Institut Universitaire de France.

Figures legends

Figure 1. The *BTN3* repertoire is different between Primates with different evolution profiles between domains. **A)** Neighbor joining (NJ) phylogeny of complete coding sequences. **B)** NJ phylogeny of the IgV coding sequences. **C)** NJ phylogeny of the IgC coding sequences. D) NJ phylogeny of the B30.2 domain coding sequences of 3A1 and 3A3. E) Presentation of the *BTN3* genes in the principal Primates groups and the protein domain architecture. Abbreviations: Hsa : *Homo sapiens* ; Ptr : *Pan troglogdytes* ; Ppa : *Pan paniscus* ; Pab : *Pongo abelii* ; Nle :

Nomascus leucogenys ; Pan : *Papio anubis* ; Mmu : *Macaca mulatta* ; Mfa : *Macaca fascicularis* ; Cgu : *Colobus guereza* ; Oga *: Otolemur garnettii*.

Figure 2. The BTN3 5' region is marked with a recombination and several homogenizations. A) Illustration of the generation of *BTN3A3* through a recombination between a *BTN3A1-like* gene and a *3A2-like gene*. **B)** Illustration of the homogenization between the *BTN3*. **C)** Alignment of Primates BTN3 protein sequences of the signal peptide, the IgV domain and the beginning of the IgC domain. Hominoid sequences are marked in turquoise and the Old World Monkeys sequences in gray. Abbreviations: U: UTR; L: leader peptide; IgV: IgV domain; IgC: IgC domain; TM: transmembrane domain; B30.2: B30.2 domain.

Figure 3. Loss of the diversity at the exon encoding the IgV domain of *BTN3*. The nucleotide diversity normalized by base pair (bp) was calculated by a window of 100bp with a step of 25 bp using the program DnaSP version 5.10.1 (41). Abbreviations: LP: leader peptide; IgV: IgV domain; IgC: IgC domain; TM: transmembrane domain; B30.2: B30.2 domain.

Figure 4. The *BTN3* are polymorphic in human populations but with an extremely conserved IgV domain. A) Summary of alleles and allotypes number for each *BTN3*. **B)** Values of nucleotide diversity (π) by exon for the alleles of each *BTN3*. **C), E)** and **G)** Summary of non-synonymous variations for allotypes with a frequency $\geq$ 1% in at least one of the regions studied. Allele nomenclature is based on the HLA convention: the first two digits (*xx) correspond to amino acid differences, subsequent digits correspond to synonymous variations. Q denotes alleles with a "questionable" structure due to a non-sense mutation. In gray are marked the sequences obtained if the exon with the non-sense mutation is spliced. Abbreviations: LP: leader peptide; IgV: IgV domain; IgC: IgC domain; CR: cytoplasmic region between the TM and the B30.2

domain; TM: transmembrane domain; B30.2: B30.2 domain; n: number of times the allotype is observed; ter: termination codon; NS: non-synonymous.

Figure 5. The evolution of *BTN2* in Primates is also marked by homogenizations. A) Presentation of the BTN2 of the principal Primates groups and their domain architecture. **B)** Illustration of the homogenizations for each group and their localizations. **C)** Alignment of BTN2 protein sequences for the TM and the beginning of the cytoplasmic region. **D)** Comparison of the nucleotide diversity (π) values calculated for each exon between the two *BTN2*. Abbreviations: U: UTR; L: leader peptide; IgV: IgV domain; IgC: IgC domain; TM: transmembrane domain and B30.2: B30.2 domain. Abbreviations for the species names are the same as for the Figure 1.

References:

1. Compte, E., P. Pontarotti, Y. Collette, M. Lopez, and D. Olive. 2004. Frontline: Characterization of BT3 molecules belonging to the B7 family expressed on immune cells. *European Journal of Immunology* 34: 2089-2099.
2. Messal, N., E. Mamessier, A. Sylvain, J. Celis-Gutierrez, M. L. Thibult, B. Chetaille, G. Firaguay, S. Pastor, Y. Guillaume, Q. Wang, I. Hirsch, J. A. Nunès, and D. Olive. 2011. Differential role for CD277 as a co-regulator of the immune signal in T and NK cells. *European Journal of Immunology* 41: 3443-3454.
3. Yamashiro, H., S. Yoshizaki, T. Tadaki, K. Egawa, and N. Seo. 2010. Stimulation of human butyrophilin 3 molecules results in negative regulation of cellular immunity. *J. Leukoc. Biol.* 88: 757-767.

4. Abeler-Dörner, L., M. Swamy, G. Williams, A. C. Hayday, and A. Bas. 2012. Butyrophilins: an emerging family of immune regulators. *Trends Immunol.* 33: 34-41.
5. Linsley, P. S., R. Peach, P. Gladstone, and J. Bajorath. 1994. Extending the B7 (CD80) gene family. *Protein Sci.* 3: 1341-1343.
6. Arnett, H. A., S. S. Escobar, E. Gonzalez-Suarez, A. L. Budelsky, L. A. Steffen, N. Boiani, M. Zhang, G. Siu, A. W. Brewer, and J. L. Viney. 2007. BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. *J. Immunol.* 178: 1523-1533.
7. Nguyen, T., X. K. Liu, Y. Zhang, and C. Dong. 2006. BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J. Immunol.* 176: 7354-7360.
8. Simone, R., B. Barbarat, A. Rabellino, G. Icardi, M. Bagnasco, G. Pesce, D. Olive, and D. Saverino. 2010. Ligation of the BT3 molecules, members of the B7 family, enhance the proinflammatory responses of human monocytes and monocyte-derived dendritic cells. *Mol. Immunol.* 48: 109-118.
9. Harly, C., Y. Guillaume, S. Nedellec, C.-M. Peigné, H. Mönkkönen, J. Mönkkönen, J. Li, J. Kuball, E. J. Adams, S. Netzer, J. Déchanet-Merville, A. Léger, T. Herrmann, R. Breathnach, D. Olive, M. Bonneville, and E. Scotet. 2012. Key implication of CD277/butyrophilin-3 (BTN3A) in cellular stress sensing by a major human γδ T-cell subset. *Blood* 120: 2269-2279.
10. Bonneville, M., and J.-J. Fournié. 2005. Sensing cell stress and transformation through Vgamma9Vdelta2 T cell-mediated recognition of the isoprenoid pathway metabolites. *Microbes and infection / Institut Pasteur* 7: 503-509.
11. Sandstrom, A., C.-M. Peigné, A. Léger, J. E. Crooks, F. Konczak, M.-C. Gesnel, R. Breathnach, M. Bonneville, E. Scotet, and E. J. Adams. 2014. The intracellular B30.2 domain of butyrophilin 3A1 binds phosphoantigens to mediate activation of human Vγ9Vδ2 T cells. *Immunity* 40: 490-500.
12. Vavassori, S., A. Kumar, G. S. Wan, G. S. Ramanjaneyulu, M. Cavallari, S. El Daker, T. Beddoe, A. Theodossis, N. K. Williams, E. Gostick, D. A. Price, D. U. Soudamini, K. K. Voon, M. Olivo, J. Rossjohn, L. Mori, and G. De Libero. 2013. Butyrophilin 3A1 binds phosphorylated antigens and stimulates human γδ T cells. *Nature Immunology* 14: 908-916.
13. Afrache, H., P. Gouret, S. Ainouche, P. Pontarotti, and D. Olive. 2012. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics* 64: 781-794.
14. Viken, M. K., A. Blomhoff, M. Olsson, H. E. Akselsen, F. Pociot, J. Nerup, I. Kockum, A. Cambon-Thomsen, E. Thorsby, D. E. Undlien, and B. A. Lie. 2009. Reproducible association with type 1 diabetes in the extended class I region of the major histocompatibility complex. *Genes Immun.* 10: 323-333.
15. Hiramatsu, M., M. Oguri, K. Kato, T. Yoshida, T. Fujimaki, H. Horibe, K. Yokoi, S. Watanabe, K. Satoh, Y. Aoyagi, M. Tanaka, H. Yoshida, S. Shinkai, Y. Nozawa, T. Murohara, and Y. Yamada. 2011. Association of a polymorphism of BTN2A1 with type 2 diabetes mellitus in Japanese individuals. *Diabetic medicine : a journal of the British Diabetic Association* 28: 1381-1387.
16. Yoshida, T., K. Kato, H. Horibe, M. Oguri, M. Fukuda, K. Satoh, Y. Aoyagi, S. Shinkai, Y. Nozawa, and Y. Yamada. 2011. Association of a genetic variant of BTN2A1 with chronic kidney disease in Japanese individuals. *Nephrology* 16: 642-648.
17. Yoshida, T., K. Kato, M. Oguri, H. Horibe, T. Kawamiya, K. Yokoi, T. Fujimaki, S. Watanabe, K. Satoh, Y. Aoyagi, M. Tanaka, H. Yoshida, S. Shinkai, Y. Nozawa, and Y. Yamada. 2011. Association of a polymorphism of BTN2A1 with chronic kidney disease in individuals with or without hypertension or diabetes mellitus. *Experimental and therapeutic medicine* 2: 325-331.
18. Oguri, M., K. Kato, T. Yoshida, T. Fujimaki, H. Horibe, K. Yokoi, S. Watanabe, K. Satoh, Y. Aoyagi, M. Tanaka, H. Yoshida, S. Shinkai, Y. Nozawa, D. J. Shin, J. H. Lee, Y. Jang, and Y. Yamada. 2011.

Association of a genetic variant of BTN2A1 with metabolic syndrome in East Asian populations. *Journal of medical genetics* 48: 787-792.

19. Fujimaki, T., K. Kato, M. Oguri, T. Yohida, H. Horibe, K. Yokoi, S. Watanabe, K. Satoh, Y. Aoyagi, M. Tanaka, H. Yoshida, S. Shinkai, Y. Nozawa, D. J. Shin, J. H. Lee, Y. Jang, and Y. Yamada. 2011. Association of a polymorphism of BTN2A1 with dyslipidemia in East Asian populations. *Experimental and therapeutic medicine* 2: 745-749.
20. Yamada, Y., T. Nishida, S. Ichihara, M. Sawabe, N. Fuku, Y. Nishigaki, Y. Aoyagi, M. Tanaka, Y. Fujiwara, H. Yoshida, S. Shinkai, K. Satoh, K. Kato, T. Fujimaki, K. Yokoi, M. Oguri, T. Yoshida, S. Watanabe, Y. Nozawa, A. Hasegawa, T. Kojima, B. G. Han, Y. Ahn, M. Lee, D. J. Shin, J. H. Lee, and Y. Jang. 2011. Association of a polymorphism of BTN2A1 with myocardial infarction in East Asian populations. *Atherosclerosis* 215: 145-152.
21. Yoshida, T., K. Kato, M. Oguri, H. Horibe, T. Kawamiya, K. Yokoi, T. Fujimaki, S. Watanabe, K. Satoh, Y. Aoyagi, M. Tanaka, H. Yoshida, S. Shinkai, Y. Nozawa, and Y. Yamada. 2011. Association of polymorphisms of BTN2A1 and ILF3 with myocardial infarction in Japanese individuals with different lipid profiles. *Molecular medicine reports* 4: 511-518.
22. Kennedy, R. B., I. G. Ovsyannikova, I. H. Haralambieva, N. D. Lambert, V. S. Pankratz, and G. A. Poland. 2014. Genetic polymorphisms associated with rubella virus-specific cellular immunity following MMR vaccination. *Human genetics*.
23. Solovyev, V. 2007. Handbook of Statistical genetics. 3rd ed. C. C. M. B. D. Balding, ed. Wiley-Interscience. 1616.
24. Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package, 1998. *Methods in molecular biology (Clifton, N.J.)* 132: 115-130.
25. Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30: 3059-3066.
26. Gudys, A., and S. Deorowicz. 2014. QuickProbs--a fast multiple sequence alignment algorithm designed for graphics processors. *PloS one* 9: e88901.
27. Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729.
28. Swofford, D. L. 2001. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer, Sunderland, Massachusetts.
29. Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.
30. Genomes Project, C., G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, and G. A. McVean. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.
31. Lappalainen, T., M. Sammeth, M. R. Friedlander, P. A. t Hoen, J. Monlong, M. A. Rivas, M. Gonzalez-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira, M. Barann, T. Wieland, L. Greger, M. van Iterson, J. Almlof, P. Ribeca, I. Pulyakhina, D. Esser, T. Giger, A. Tikhonov, M. Sultan, G. Bertier, D. G. MacArthur, M. Lek, E. Lizano, H. P. Buermans, I. Padioleau, T. Schwarzmayr, O. Karlberg, H. Ongen, H. Kilpinen, S. Beltran, M. Gut, K. Kahlem, V. Amstislavskiy, O. Stegle, M. Pirinen, S. B. Montgomery, P. Donnelly, M. I. McCarthy, P. Flicek, T. M. Strom, C. Geuvadis, H. Lehrach, S. Schreiber, R. Sudbrak, A. Carracedo, S. E. Antonarakis, R. Hasler, A. C. Syvanen, G. J. van Ommen, A. Brazma, T. Meitinger, P. Rosenstiel, R. Guigo, I. G. Gut, X. Estivill, and E. T. Dermitzakis. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501: 506-511.
32. Stephens, M., and P. Scheet. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American journal of human genetics* 76: 449-462.

33. Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *American journal of human genetics* 68: 978-989.
34. Chen, B., S. Wilkening, M. Drechsel, and K. Hemminki. 2009. SNP_tools: A compact tool package for analysis and conversion of genotype data for MS-Excel. *BMC Res Notes* 2.
35. Sandstrom, A., C. M. Peigne, A. Leger, J. E. Crooks, F. Konczak, M. C. Gesnel, R. Breathnach, M. Bonneville, E. Scotet, and E. J. Adams. 2014. The intracellular B30.2 domain of butyrophilin 3A1 binds phosphoantigens to mediate activation of human Vgamma9Vdelta2 T cells. *Immunity* 40: 490-500.
36. Yamada, E., and D. W. McVicar. 2008. Paired receptor systems of the innate immune system. *Current protocols in immunology / edited by John E. Coligan ... [et al.]* Chapter 1: Appendix 1X.
37. Lanier, L. L. 2001. Face off--the interplay between activating and inhibitory immune receptors. *Current opinion in immunology* 13: 326-331.
38. Fossum, S., P. C. Saether, J. T. Vaage, M. R. Daws, and E. Dissen. 2011. Paired opposing leukocyte receptors recognizing rapidly evolving ligands are subject to homogenization of their ligand binding domains. *Immunogenetics* 63: 809-820.
39. della Gaspera, B., D. Pham-Dinh, G. Roussel, J. L. Nussbaum, and A. Dautigny. 1998. Membrane topology of the myelin/oligodendrocyte glycoprotein. *European journal of biochemistry / FEBS* 258: 478-484.
40. Tazi-Ahnini, R., J. Henry, C. Offer, C. Bouissou-Bouchouata, I. H. Mather, and P. Pontarotti. 1997. Cloning, localization, and structure of new members of the butyrophilin gene family in the juxta-telomeric region of the major histocompatibility complex. *Immunogenetics* 47: 55-63.
41. Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

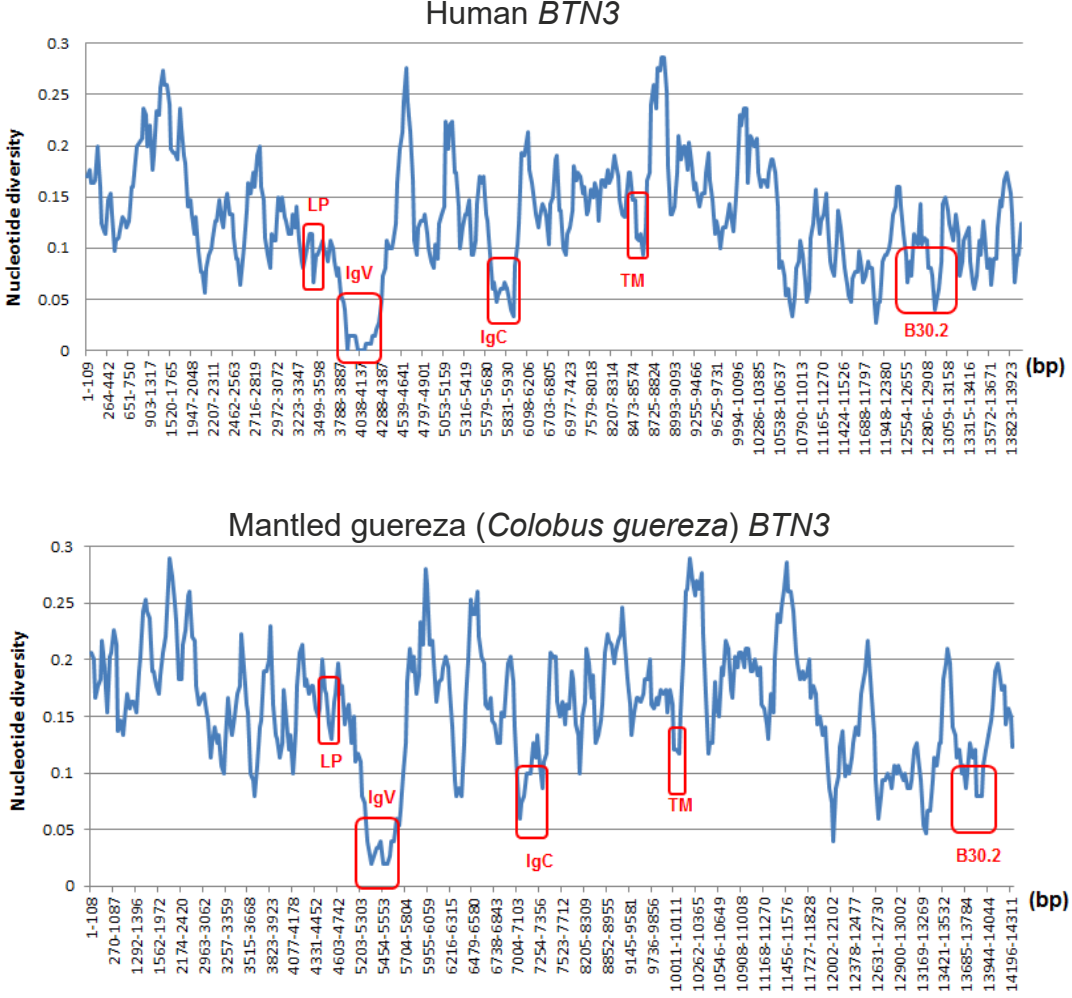

Figure 3. Loss of the diversity at the exon encoding the IgV domain of BTN3.

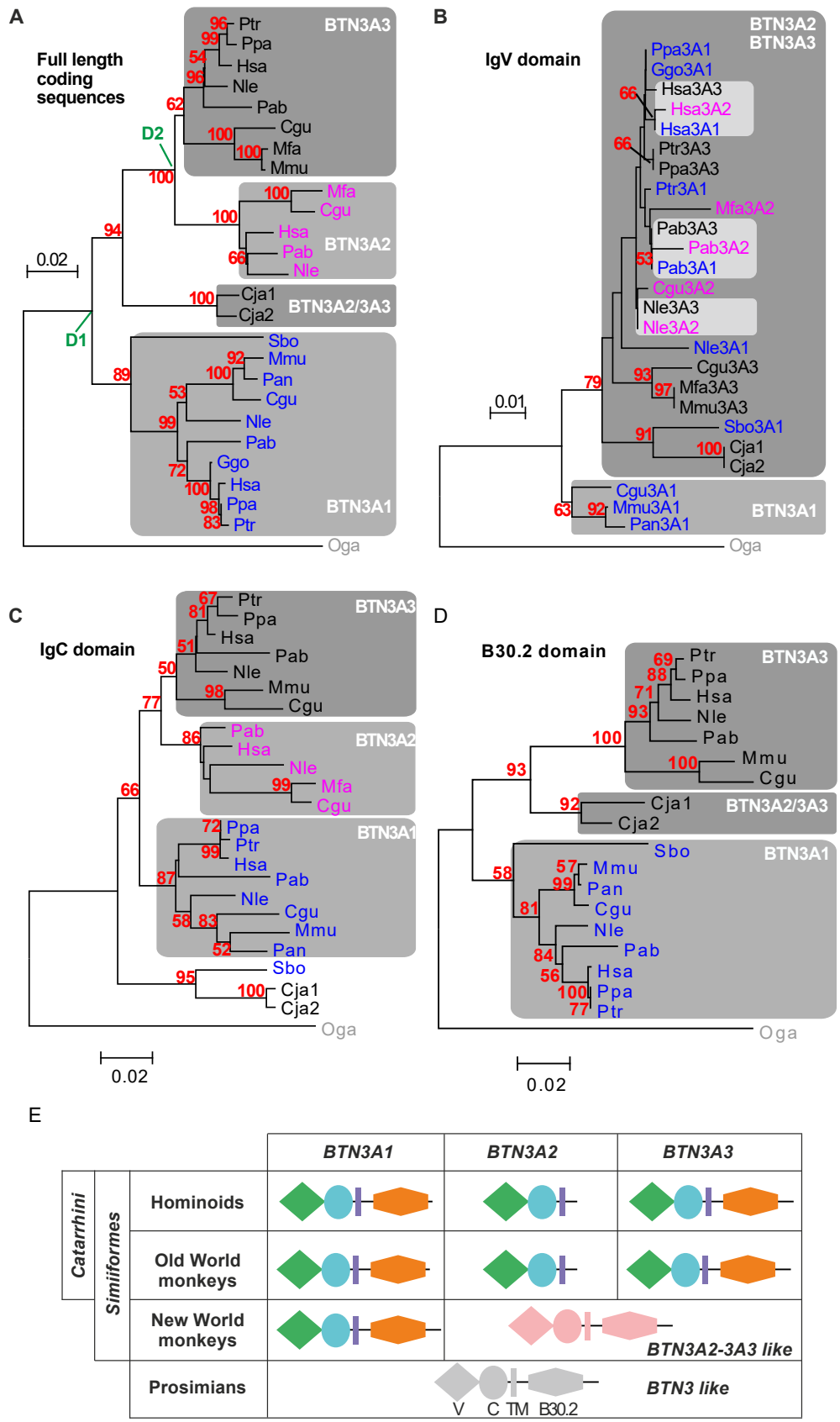

Figure 1:The *BTN3* repertoire is different between Primates with different evolution profiles between domains

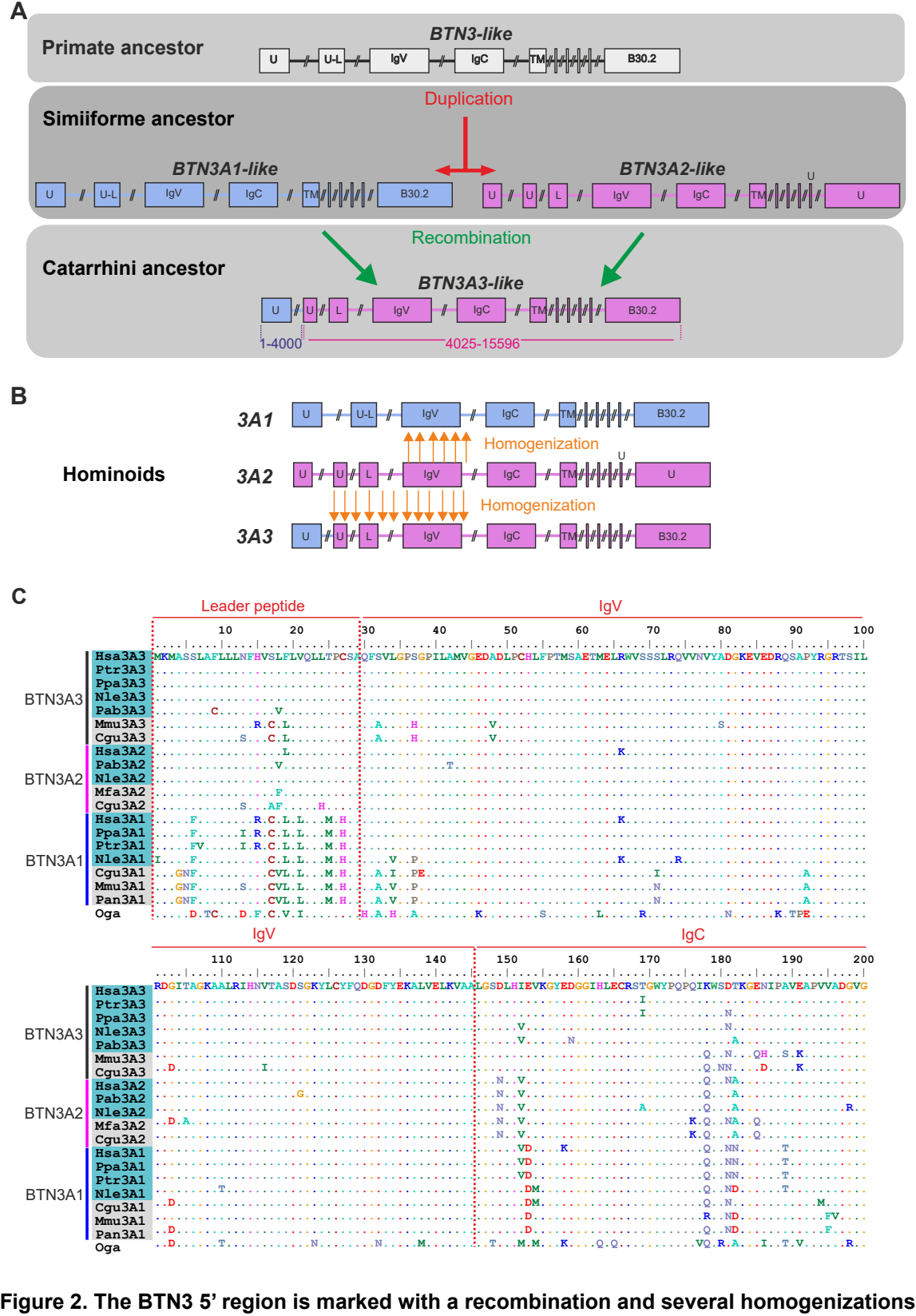

Figure 2. The BTN3 5' region is marked with a recombination and several homogenizations