Evolutionary and expression analyses reveal a pattern of ancient duplications and functional specializations in the diversification of the Downstream of Kinase (DOK) genes

Geoffrey Guittard, Pierre Pontarotti, Samuel Granjeaud, Magda Rodrigues, Laurent Abi-Rached, Jacques A. Nunès

Evolutionary and expression analyses reveal a pattern of ancient duplications and functional specializations in the diversification of the Downstream of Kinase (DOK) genes

Geoffrey Guittard [a], Pierre Pontarotti [b], Samuel Granjeaud [c], Magda Rodrigues [a] , Laurent Abi-Rached [d,*] and Jacques A Nunès [a,*]

[a] Centre de Recherche en Cancérologie de Marseille, Immunity and Cancer team ; Institut Paoli-Calmettes ; Inserm, U1068 ; CNRS, UMR7258 ; Aix-Marseille Université UM 105, Marseille, France.

[b] Aix Marseille Université, CNRS, Centrale Marseille, I2M UMR 7373, équipe évolution biologique modélisation, Marseille, France.

[c] Centre de Recherche en Cancérologie de Marseille, CiBi Platform ; Institut Paoli-Calmettes ; Inserm, U1068 ; CNRS, UMR7258 ; Aix-Marseille Université UM 105, Marseille, France.

[d] Equipe ATIP, Unité de Recherche sur les Maladies Infectieuses Tropicales Emergentes (URMITE); CNRS, UMR 7278 ; IRD, 198 ; Inserm, U1095 ; IHU Méditerranée Infection ; Aix-Marseille Université UM63, Marseille, France.

Running title: Evolutionary and expression analyses of DOK genes

[*] To whom the correspondence should be addressed, Laurent Abi-Rached, Unité de Recherche sur les Maladies Infectieuses Tropicales Emergentes, IHU Méditerranée Infection, Aix-Marseille Université, 19-21 boulevard Jean Moulin, 13385 Marseille Cedex 5, France. Tel: +33 (0)6 52 49 74 39, E-mail address: Laurent.Abi-Rached@univ-amu.fr and Jacques Nunès, Centre de Recherche en Cancérologie de Marseille, CS 30059, 27 bd Leï Roure, 13273 Marseille Cedex 09, France. Tel: +33 (0)486 977 274, Fax: +33 (0)486 977 499, E-mail address: jacques.nunes@inserm.fr

Abstract

Downstream of Kinase (DOK) proteins represent a multigenic family of adaptors that includes negative regulators of immune cell signaling. Using phylogenetics and intron/exon structure data, we show here that the seven human *DOK* genes (*DOK1* to *DOK7*) form three highly divergent groups that emerged before the protostome-deuterostome split: *DOK1/2/3*, *DOK4/5/6*, and *DOK7*. For two of these three groups (*DOK1/2/3* and *DOK4/5/6*), further gene duplications occurred in vertebrates and so while chordates only have three *DOK* genes, vertebrates have seven *DOK* genes over the three groups. From our expression analysis in humans, we show that each group of *DOK* genes has a distinct pattern of expression. The *DOK1/2/3* group is immune specific, yet each of the three genes in the group has a distinct pattern of expression in immune cells. This immune specificity could thus be ancestral, with the *DOK1/2/3* gene also being immune-related in protostomes. The *DOK4/5/6* and *DOK7* groups represent genes that are much less expressed in immune system than the *DOK1/2/3* group. Interestingly, we identify a novel tyrosine based motif that is specific to the vertebrate *DOK4*/5/6 sequences. The evolution of the *DOK* genes is thus marked by a pattern of ancient duplications and functional specializations.

Keywords : *DOK* genes; immune cell signaling; adapter molecules.

1. Introduction

Downstream of Kinase (DOK) proteins represent a multigenic family of adaptors that includes regulators of immune cell signaling. To date, there are seven identified members: DOK1 to DOK7. The DOK proteins play a central role in the assembly of the binding partners in different cell types especially upon receptor tyrosine kinase and immunoreceptor triggering (Jordan et al., 2003; Mashima et al., 2009). Some of these adaptor proteins are preferentially expressed in hematopoietic cells. All of them share structural similarities characterized by an amino-terminal part with both a pleckstrin-homology (PH) and a phosphotyrosine binding (PTB) domain that is essential for their plasma membrane recruitment and a carboxy-terminal part containing tyrosine residues that can be phosphorylated, allowing the recruitment of Src-homology-2 (SH2) domains bearing proteins. DOK1/2/3 are phosphorylated upon triggering of immunoreceptors, such as T-cell receptor (TCR) (Dong S et al., 2006), B-cell receptor (BCR) (Lemay S et al., 2000) or some Fc receptors (FcRs) (Ott VL et al., 2002). In addition, mouse models show the important role of these DOK proteins in immune responses (Celis-Gutierrez et al., 2014; Ng et al., 2007; Yasuda et al., 2007). These adapter proteins are key players of the negative regulation of signaling pathways in immune cells. On the top of these three DOK family members, four other DOK proteins were identified, initially outside of the immune system, and were named DOK4-7 (Crowder et al., 2004; Grimm et al., 2001; Okada et al., 2006). Among them, we reported that *DOK4* and *DOK5* genes are expressed in human T cells (Favre et al., 2003) and subsequently demonstrated that DOK4 is a negative regulator of T cell activation (Gerard et al., 2009). To further investigate the putative role of all DOK family members in immune cells, we performed an investigation of their evolution and expression patterns.

2. Materials and methods

2.1 Genomic structure

Exon/Intron structures of *DOK* genes were obtained using the Ensembl website (*www.ensembl.org/*).

2.2 *Dok* gene dataset

To assemble the *Dok* dataset used in the phylogenetic analysis, BLAST (Altschul et al., 1990) searches were performed to screen NCBI's non-redundant (NR), expressed sequence tags (EST), and whole genome shotgun (WGS) databases. *Branchiostoma belcheri* sequences were also obtained from the Chinese Amphioxus Genome Project database (http://mosas.sysu.edu.cn/genome/). For the WGS database, gene structures were reconstructed using FGENESH (Salamov and Solovyev, 2000). Because *DOK* genes are present in most bilaterian lineages, a set of sequences representing the main taxonomic groups was selected to avoid overrepresentation of certain groups (such as mammals). In doing so, and wherever possible, we used the same reference species for a given taxonomic group. Yet, in some cases, it was not possible to find a complete sequence for a given species (for example *Danio rerio* DOK5 and DOK7 or *Xenopus tropicalis* DOK3). In such cases however we cannot conclude that the gene was lost in those species as it could just be a case of partially-sequenced genomes.

2.3 Sequence alignment and phylogenetic analysis

Initial sequence alignment for the region covering the DOK proteins was generated with MUSCLE (Edgar, 2004) and improved using SATé-II (Liu et al., 2012) and manual correction. Phylogenetic analyses were conducted using maximum-likelihood (ML), Neighbor-Joining (NJ), and parsimony methods. NJ analysis was performed with MEGA6 (Tamura et al., 2013) using a Poisson correction distance with pairwise deletion and 500

bootstrap replicates. PAUP*4.0b10 (Swofford et al., 2001) and the tree bisection-reconnection branch swapping algorithm were used for parsimony analyses with 500 bootstrap replicates and a heuristic search. ML analysis was performed with RAXML8 (Stamatakis, 2014) under the WAG+G+F model with 500 bootstrap replicates (rapid bootstrapping) (Stamatakis and Ott, 2008).

2.4 Accession codes

RNAseq profiles used to assess *DOK* gene profiles in human immune cells were collected from GEO (http://www.ncbi.nlm.nih.gov/geo/), accession GSE60424. The number of reads shown in a Supplemental Table represents the average of counts in four healthy donors and twelve to sixteen patients. Standard deviations (sd) as well as coefficients of variation (CV) are also provided. RNAseq profiles from mouse immune cells were collected from the Immunological Genome project (ImmGen, http://www.immgen.org) using the RNA-seq Skyline tool.

2.5 Analysis procedures

Human expression levels are the counts of reads aligned the genome and expressed as count per million. Values were extracted from supplementary data of the GSE60424. The providers analyzed the sequencing libraries on an Illumina HiScan, with a target read depth of ~20M reads. Then, they aligned the reads to Hg19 via Omicsoft sequence aligner (OSA) version 2.0.1. Gene counts were generated by HTSeq version 0.5.4p3 and normalized using the Trimmed Mean of M-values procedure (EdgeR package, Bioconductor), which considers that most genes are invariant among different experiments. We re-assembled clinical data to expression levels and annotated the ensembl identifiers by use of R and Bioconductor packages (GEOquery and org.Hs.eg.db). Mouse expression levels result from a similar

processing of reads involving the comparable DESeq2 package. The reads were generated using ImmGen's "Ultra Low Input" RNAseq pipeline, starting from 1,000 sorted cells.

3. Results

3.1 Three ancient groups of DOK sequences

To further classify the DOK family, we performed phylogenetic analyses for all DOK family members. This analysis shows that bilaterians possess three distinct groups of *DOK* sequences that emerged before the protostome-deuterostome split: a *DOK1/2/3* group (first group), a *DOK4/5/6* group (second group), and a *DOK7* group (Figure 1A). The phylogenetic tree shows that the *DOK7*, *DOK1/2/3* and *DOK4/5/6* genes were present in a metazoan ancestor. The *DOK7* gene remained single copy throughout metazoan evolution while *DOK1/2/3* and *DOK4/5/6* duplicated in vertebrates to give rise to *DOK1-3* and *DOK4-6*, respectively. Interestingly these gene duplications all occurred before the mammal-bony fish split, so that while chordates only possess three *DOK* genes, vertebrates possess seven *DOK* genes.

3.2 Each of the three DOK groups has a unique genomic structure

To understand the evolution of eukaryotic genes, it is important to investigate exon/intron structures. A gene prediction program was used in combination with *DOK* coding sequences to analyze the seven human *DOK* genes and deduce exon/intron structures. This analysis shows that the three groups identified in the phylogenetic analysis all display a unique exon/intron structure (Figure 1B). Yet, for the two groups with three *DOK* genes (*DOK1/2/3* and *DOK4/5/6*) the structures are conserved within each group. We already showed that *DOK* genes from the first group are composed of four or five putative exons and two *DOK* genes from the second group, *DOK4* and *DOK5* are composed of eight putative exons (Favre et al., 2003). Interestingly *DOK6* showed the same exon/intron structure compared to other members of this group, *DOK4* and *DOK5*. The *DOK*7 exon/intron structure showed a dramatic difference compared to *DOK* genes from the first or second group. *DOK*7 is encoded

by 8 exons. Analysis of phylogenetic trees and exon/intron structure of *DOK* family members are thus congruent, since the same groups (*DOK1/2/3*, *DOK4/5/6* and *DOK*7) can be defined in both experimental approaches.

3.3 *DOK1*, *DOK2* and *DOK3* are differentially expressed in immune cells

In order to study *DOK1* to *DOK7* expression pattern, we analysed data processed into RNA sequencing (RNAseq) (GEO, Accession GSE60424) (Linsley et al., 2014). PBMCs samples from twenty donors were analysed (suppl. Table). In bone marrow, the hematopoietic stem and progenitor cells (HSPCs) give rise to the mature blood cells. Gene expression for *DOK1* to *DOK7* has been reported in HSPCs (Coppin et al., 2016; Dutta et al., 2017; Itoh-Nakadai et al., 2017; Izadpanah et al., 2008; Kunimoto et al., 2017; Will et al., 2012). Here, we studied expression level in different mature hematopoietic cells (neutrophils, monocytes, B cells, CD4 and CD8 T cells, natural killer (NK) cells). Expression pattern of human *DOK* genes across different immune cell types shows that the *DOK* transcripts are heterogenously expressed in the immune system (Table S1, Figure S1). It is interesting to note that *DOK1/2/3* genes are immune specific at least in human and mouse. This lead us to predict that this represents an ancestral pattern for all vertebrate *DOK1/2/3* sequences and that the *DOK1/2/3* genes in protostomes are also involved in immune functions. Immune specific functions of the ancestral *DOK1/2/3* gene have not been currently assessed (i.e. in protostomes). In fruit flies as *Drosophila melanogaster*, the expression pattern of this ancestral *DOK1/2/3* gene, *Ddok* is quite similar to the pattern of a gene encoding for a Syk family kinase, *Shark* (Biswas et al., 2006). Ddok interacts physically with Shark (Biswas et al., 2006) and both fly Src family kinases (SFK), Src42A and Src64B (Giot et al., 2003). In *Drosophila*, a receptor containing an immunoreceptor tyrosine-based activation motif (ITAM), Draper promotes phagocytosis in hemocytes/macrophages through an ITAM-domain-SFK-Syk-mediated signaling cascade

(Ziegenfuss et al., 2008). This signaling pathway is homologous to the well-established SFK-ITAM-Syk-signaling pathway used in vertebrate adaptive/innate immune responses. Taken together, the role of Ddok in this signaling axis should be addressed to potentially define DOK immune functions in *Drosophila*.

Interestingly *DOK2* is the most expressed gene in all hematopoietic cells, except for neutrophils and B cells. Inversely, *DOK3* gene is highly expressed in B cells and neutrophils. This observation shows that although closely related, genes from group A (*DOK1/2/3*) are expressed in different cell types. It would indicate that the *DOK2* and *DOK3* genes are expressed in different lymphocyte subsets. Concerning *DOK1*, this gene is broadly expressed among hematopoietic cells but always at a lower level compared to the *DOK2* or *DOK3* expression patterns. Thus, *DOK1* may have a redundant role in cells when co-expressed with *DOK2* or *DOK3*. In accordance, the DOK1 and DOK2 proteins have a redundant role in T cells (Yasuda et al., 2007) or myeloid lineage (Yasuda et al., 2004). Likewise, both DOK1 and DOK3 proteins could reach similar signaling endpoints in B cells, but it is less obvious (Ng et al., 2007; Yamanashi et al., 2000a). Although these studies show differential responses, it might be of interest to study mice lacking both *Dok1* and *Dok3* to investigate the potential redundancy of these proteins in B cells or in neutrophils.

Concerning *DOK4*, *DOK5* and *DOK6* genes (group B), only *DOK4* and *DOK6* are expressed in human immune cells (Table S1). Both human and mouse *DOK4/Dok4* gene are expressed in a majority of tested immune cell subsets (Figure S1). Consistent to that, previous studies showed that DOK4 could be important in T cells (Gerard et al., 2009), and human *DOK5* gene expression is induced upon T cell activation (Favre et al., 2003). Upon viral infection in mice, activated CD8+ T cells habor a slight induction of *Dok5* and *Dok6* similar to *Il2ra* encoding for the IL-2 receptor α chain (Figure S2).

Finally, *DOK7* (group C) gene expression is detected in human B and at lower level, in human NK cells, suggesting a potential role of DOK7 at least in these human cell types. This point should be moderated by the fact that *Dok7* gene expression was not detected in mouse B and NK cells.

3.4 Identification of a newly conserved region in the group B *DOK* gene

From the sequence alignment of all DOK proteins, we could identify a conserved motif sequence within group B members (Figure 2). This conserved sequence is localized in the carboxy-terminal part of all vertebrate DOK4/5/6 proteins, and such conservation is consistent with an interaction with partners that are also present and conserved in all vertebrates. Interestingly, a point mutation of the tyrosine residue located in this motif of the DOK4 protein dramatically decreased neurite outgrowth upon GNDF stimulation compared to wild-type DOK4 (Uchida et al., 2006). This observation reinforces the possibility that this conserved motif is important for DOK4 function and probably DOK5 and DOK6 proteins as well. Further studies are necessary to identify potent interactor(s) of this highly conserved motif in DOK proteins across vertebrates. Since this region is not conserved in the first DOK group (group A: *DOK1/2/3*), we suggest that this motif could be a subject of difference between the group A and group B in immune functions. Yet, studies showed that DOK4 could have a role in immune cells and, intriguingly, that DOK4 and DOK5 can be expressed upon stimulation in immune cells (Favre et al., 2003; Gerard et al., 2009). Thus we cannot exclude a role for this second DOK group (group B: *DOK4/5/6*) and this conserved region in immune cells upon stimulation.

4. Conclusion

Using phylogenetic analyses of *DOK* gene family we showed that the seven human DOK genes (*DOK1-7*) form three highly divergent groups that emerged before the protostome-deuterostome split: *DOK1/2/3*, *DOK4/5/6*, and *DOK7*. The *DOK7* gene remained single copy throughout metazoan evolution while *DOK1/2/3* and *DOK4/5/6* duplicated in vertebrates to give rise to *DOK1-3* and *DOK4-6*, respectively. Consistent with this, the three phylogenetic groups all display a distinct exon/intron structure.

We confirmed using RNA sequencing data that the first DOK group (group A: *DOK1-3*) is more expressed in immune cells in vertebrates (at least in human and mouse). Our observation that *DOK3* is highly expressed in both human and mouse neutrophils, allow us to predict a role for DOK3 in this cell type. The collected RNA sequencing data (Figure S1 & S2) are showing that some gene expression for group B and C can be detected in immune cells. When these *DOK* genes are detected across the different immune cell subsets both in human and mouse, the human and the orthologous mouse gene are co-clustering, that is the case for the DOK group A and the first member of the group B, *DOK4* (Supplementary Figure). *DOK4-6* genes (second group, group B) seem to get a limited in immune cells. However *DOK4* and *DOK5* are expressed upon T-cell receptor (TCR) stimulation in T cells and DOK4 protein has a role in immune cells (Favre et al., 2003; Gerard et al., 2009). Thus, although this second group gene expression is low in immune cells at steady state, it will be interesting to test *DOK4-6* gene expression in activated immune cells. Unexpectedly, we also show that *DOK7* was expressed in human B cells at a level comparable to that of *DOK1*. An inhibitory role for Dok1 in B cell signaling was demonstrated so it could be of interest to similarly investigate DOK7 function in B cells (Yamanashi et al., 2000b). Interestingly, DOK3 influences B cell signaling pathways in a non-redundant way (Ng et al., 2007).

Finally, the identification of a conserved motif in the carboxy-terminal part of group B DOK (DOK4-6) in vertebrates suggests the binding of a vertebrate protein that is important for their functions. Interestingly, a previous study showed that this motif was important in DOK4 function (Uchida et al., 2006). Future studies are required to know whether this motif could be important in DOK5 and DOK6 functions. Since DOK4 and DOK5 may be expressed upon immune cell stimulations, this may be of interest to study the role of this motif in this context.

Our study thus reinforces the importance of combining evolutionary and expression studies to understand gene family organization. This approach would predict that the protein encoded by a particular DOK gene would deliver similar functions across the immune system of the different species. However, the endpoint will be to carry functional experiments in key animal models to decipher the role of DOK proteins during the evolution of the immune system.

Funding: This work was supported by institutional grants from the Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique and Aix-Marseille Université to CRCM, by the fondation ARC Recherche sur le Cancer (# PJA20161204835). G. Guittard is supported by a fellowship from the fondation ARC Recherche sur le Cancer.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Biswas, R., Stein, D., Stanley, E.R., 2006. Drosophila Dok is required for embryonic dorsal closure. Development 133, 217-227.

Celis-Gutierrez, J., Boyron, M., Walzer, T., Pandolfi, P.P., Jonjic, S., Olive, D., Dalod, M., Vivier, E., Nunes, J.A., 2014. Dok1 and Dok2 proteins regulate natural killer cell development and function. The EMBO journal 33, 1928-1940.

Coppin, E., De Grandis, M., Pandolfi, P.P., Arcangeli, M.L., Aurrand-Lions, M., Nunes, J.A., 2016. Dok1 and Dok2 Proteins Regulate Cell Cycle in Hematopoietic Stem and Progenitor Cells. J Immunol 196, 4110-4121.

Crowder, R.J., Enomoto, H., Yang, M., Johnson, E.M., Jr., Milbrandt, J., 2004. Dok-6, a novel p62 Dok family member, promotes Ret-mediated neurite outgrowth. J Biol Chem.

Dutta, A., Hutchison, R.E., Mohi, G., 2017. Hmga2 promotes the development of myelofibrosis in Jak2(V617F) knockin mice by enhancing TGF-beta1 and Cxcl12 pathways. Blood 130, 920-932.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Favre, C., Gérard, A., Clauzier, E., Pontarotti, P., Olive, D., Nunès, J.A., 2003. DOK4 and DOK5: new Dok-related genes expressed in human T cells. Genes Immun 4, 40-45.

Gerard, A., Ghiotto, M., Fos, C., Guittard, G., Compagno, D., Galy, A., Lemay, S., Olive, D., Nunes, J.A., 2009. Dok-4 is a novel negative regulator of T cell activation. J Immunol 182, 7681-7689.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C.A., Finley, R.L., Jr., White, K.P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R.A., McKenna, M.P., Chant, J., Rothberg, J.M., 2003. A protein interaction map of Drosophila melanogaster. Science 302, 1727-1736.

Grimm, J., Sachs, M., Britsch, S., Di Cesare, S., Schwarz-Romond, T., Alitalo, K., Birchmeier, W., 2001. Novel p62dok family members, dok-4 and dok-5, are substrates of the c-Ret receptor tyrosine kinase and mediate neuronal differentiation. J.Cell Biol. 154, 345-354.

Itoh-Nakadai, A., Matsumoto, M., Kato, H., Sasaki, J., Uehara, Y., Sato, Y., Ebina-Shibuya, R., Morooka, M., Funayama, R., Nakayama, K., Ochiai, K., Muto, A., Igarashi, K., 2017. A Bach2-Cebp Gene Regulatory Network for the Commitment of Multipotent Hematopoietic Progenitors. Cell Rep 18, 2401-2414.

Izadpanah, R., Kaushal, D., Kriedt, C., Tsien, F., Patel, B., Dufour, J., Bunnell, B.A., 2008. Long-term in vitro expansion alters the biology of adult mesenchymal stem cells. Cancer Res 68, 4229-4238.

Jordan, M.S., Singer, A.L., Koretzky, G.A., 2003. Adaptors as central mediators of signal transduction in immune cells. Nat Immunol 4, 110-116.

Kunimoto, H., McKenney, A.S., Meydan, C., Shank, K., Nazir, A., Rapaport, F., Durham, B., Garrett-Bakelman, F.E., Pronier, E., Shih, A.H., Melnick, A., Chaudhuri, J., Levine, R.L., 2017. Aid is a key regulator of myeloid/erythroid differentiation and DNA methylation in hematopoietic stem/progenitor cells. Blood 129, 1779-1790.

Linsley, P.S., Speake, C., Whalen, E., Chaussabel, D., 2014. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. PLoS One 9, e109760.
Liu, K., Warnow, T.J., Holder, M.T., Nelesen, S.M., Yu, J., Stamatakis, A.P., Linder, C.R., 2012. SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. Syst Biol 61, 90-106.
Mashima, R., Hishida, Y., Tezuka, T., Yamanashi, Y., 2009. The roles of Dok family adapters in immunoreceptor signaling. Immunological reviews 232, 273-285.
Ng, C.H., Xu, S., Lam, K.P., 2007. Dok-3 plays a nonredundant role in negative regulation of B-cell activation. Blood 110, 259-266.
Okada, K., Inoue, A., Okada, M., Murata, Y., Kakuta, S., Jigami, T., Kubo, S., Shiraishi, H., Eguchi, K., Motomura, M., Akiyama, T., Iwakura, Y., Higuchi, O., Yamanashi, Y., 2006. The muscle protein Dok-7 is essential for neuromuscular synaptogenesis. Science 312, 1802-1805.
Salamov, A.A., Solovyev, V.V., 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res 10, 516-522.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
Stamatakis, A., Ott, M., 2008. Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. Philos Trans R Soc Lond B Biol Sci 363, 3977-3984.
Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst Biol 50, 525-539.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-2729.
Uchida, M., Enomoto, A., Fukuda, T., Kurokawa, K., Maeda, K., Kodama, Y., Asai, N., Hasegawa, T., Shimono, Y., Jijiwa, M., Ichihara, M., Murakumo, Y., Takahashi, M., 2006. Dok-4 regulates GDNF-dependent neurite outgrowth through downstream activation of Rap1 and mitogen-activated protein kinase. J Cell Sci 119, 3067-3077.
Will, B., Zhou, L., Vogler, T.O., Ben-Neriah, S., Schinke, C., Tamari, R., Yu, Y., Bhagat, T.D., Bhattacharyya, S., Barreyro, L., Heuck, C., Mo, Y., Parekh, S., McMahon, C., Pellagatti, A., Boultwood, J., Montagna, C., Silverman, L., Maciejewski, J., Greally, J.M., Ye, B.H., List, A.F., Steidl, C., Steidl, U., Verma, A., 2012. Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. Blood 120, 2076-2086.
Yamanashi, Y., Tamura, T., Kanamori, T., Yamane, H., Nariuchi, H., Yamamoto, T., Baltimore, D., 2000a. Role of the rasGAP-associated docking protein p62(dok) in negative regulation of B cell receptor-mediated signaling. Genes Dev 14, 11-16.
Yamanashi, Y., Tamura, T., Kanamori, T., Yamane, H., Nariuchi, H., Yamamoto, T., Baltimore, D., 2000b. Role of the rasGAP-associated docking protein p62(dok) in negative regulation of B cell receptor-mediated signaling. Genes Dev. 14, 11-16.
Yasuda, T., Bundo, K., Hino, A., Honda, K., Inoue, A., Shirakata, M., Osawa, M., Tamura, T., Nariuchi, H., Oda, H., Yamamoto, T., Yamanashi, Y., 2007. Dok-1 and Dok-2 are negative regulators of T cell receptor signaling. Int Immunol 19, 487-495.
Yasuda, T., Shirakata, M., Iwama, A., Ishii, A., Ebihara, Y., Osawa, M., Honda, K., Shinohara, H., Sudo, K., Tsuji, K., Nakauchi, H., Iwakura, Y., Hirai, H., Oda, H., Yamamoto, T., Yamanashi, Y., 2004. Role of Dok-1 and Dok-2 in myeloid homeostasis and suppression of leukemia. J Exp Med 200, 1681-1687.

Ziegenfuss, J.S., Biswas, R., Avery, M.A., Hong, K., Sheehan, A.E., Yeung, Y.G., Stanley, E.R., Freeman, M.R., 2008. Draper-dependent glial phagocytic activity is mediated by Src and Syk family kinase signalling. Nature 453, 935-939.

Figure legends

Figure 1. Phylogenic and exon/intron analyses identify three ancient groups of *DOK* sequences **A**. Phylogenic analysis of *DOK* family members. The analysis was performed on the complete coding sequences of *DOK* genes. Bootstrap support at nodes is given as a percentage. **B**. Schematic diagram showing the comparison of the genomic structure of the human *DOK* genes. Exons are shown by open boxes, and introns by the connecting lines. Numbers inside boxes indicate exon lengths in base pairs. For the smallest exon lenghts, the number of base pairs is indicated using an asterisk (*,**) in the right side of the structures. The 5' UTR and 3' UTR extremities are not represented and the introns are not drawn to scale. The intron phases are given by the numbers 0, 1, 2 inside the shematic exon/intron structures.

Figure 2. Identification of a motif in the second group of DOK proteins (DOK4/5/6). Upper panel, shematic representation of human DOK proteins from the phylogenic group B, DOK4, DOK5 and DOK6. The number of amino acids (aa) contained in each DOK sequence is indicated in the right part of the panel. These proteins are containing a tandem of a pleckstrin homology (PH) and a phosphotyrosine binding (PTB) domain in the amino-terminal region. The carboxy-terminal region harbors several several tyrosine (Y) residues. Y in red (upper panel) denotes the central tyrosine residue in the conserved peptidic motif (lower panel). This tyrosine residue is corresponding to the position 270 (Y270) in human DOK4 structure and the position 268 (Y268) in human DOK5 or DOK6 structure. This motif is a conserved sequence of eleven amino acids across vetebrates species. Around the central tyrosine residue, the most frequent upstream sequence is leucine – proline – arginine – serine – alanine and the downstream sequence is tryptophan – (histidine or glutamine) – histidine – isoleucine – threonine, giving the following signature: LPRSAYW(H/Q)HIT.