Evolutionary history of Ku proteins: evidence of horizontal gene transfer from archaea to eukarya

Ashmita Mainali
Kathmandu University

Sadikshya Rijal
Kathmandu University

Hitesh Kumar Bhattarai (hitesh321@gmail.com)
Kathmandu University  https://orcid.org/0000-0002-7147-1411

Research article

Keywords: Double Stranded Breaks, Non-homologous End Joining, Maximum Likelihood Phylogenetic tree, domains, Ku protein, origin, evolution

DOI: https://doi.org/10.21203/rs.3.rs-58075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

The DNA end joining protein, Ku, is essential in Non-Homologous End Joining in prokaryotes and eukaryotes. It was first discovered in eukaryotes and later by PSI blast, was discovered in prokaryotes. While Ku in eukaryotes is often a multi domain protein functioning in DNA repair of physiological and pathological DNA double stranded breaks, Ku in prokaryotes is a single domain protein functioning in pathological DNA repair in spores or late stationary phase. In this paper we have attempted to systematically search for Ku protein in different phyla of bacteria and archaea as well as in different kingdoms of eukarya.

Result

From our search of 116 sequenced bacterial genomes, only 25 genomes yielded at least one Ku sequence. From a comprehensive search of all NCBI archaeal genomes, we received a positive hit in 7 specific archaea that possessed Ku. In eukarya, we found Ku protein in 27 out of 59 species. Since the entire genome of all eukaryotic species is not fully sequenced this number could go up. We then drew a phylogenetic maximum likelihood tree to determine the ancestral relationship between Ku70 and Ku80 in eukaryotes and Ku in prokaryotes. Out tree revealed a common node for some archaeal Ku, Ku70 and Ku80.

Conclusion

This led us to hypothesize that Ku from archaea transferred through horizontal gene transfer onto neozoa and then duplicated to form Ku70 and Ku80. Additionally, we analyzed the domains of the different eukaryotic species to demonstrate that fusion, fission, terminal addition, terminal deletion, single domain loss, single domain emergence events during evolution.

Background

Double stranded breaks (DSBs) are one of the most common forms of DNA lesions, which leads to the instability of genome followed by cell death. The cells produce DSBs by either pathologic means (chromosomal translocation, physical activity, reactive oxygen species, ionizing radiations and unusual action of nuclear enzyme on DNA) or physiological means ([VDJ] recombination). The cell has devised two major ways of tackling such breaks to maintain the integrity of genome: namely homologous recombination (HR) pathway and non-homologous end joining (NHEJ) pathway. Eukaryotes use the HR pathway and NHEJ to different extent, HR being dominant in lower eukaryotes like yeast and NHEJ being predominant in higher eukaryotes \(^1\). The NHEJ is suited to repair DSBs as the repair enzymes act in any order iteratively and in sequence-independent manner. In NHEJ, the enzymes function independent of each other accepting an array of DSBs as substrates (blunt ends, 5’ and 3’ overhangs, and DNA hairpins) \(^2\).

The repair of DSBs in eukaryotes via NHEJ pathway
NHEJ pathway depends on a multifunctional protein called Ku. Reports of Ku deficient cells showing inaccurate end-joining, dramatic radiosensitivity, and chromosomal breakage, translocations and aneuploidy demonstrate the importance of Ku in DSBs repair in eukaryotes. In humans, 500,000 Ku is present per cell, and Ku has a strong affinity for DNA ends allowing it to associate with DNA ends within 5 seconds of the damage. Ku is the DNA binding subunit of the DNA-dependent protein kinase (DNA-PK) complex, which comprises Ku and DNA-PK catalytic subunit (DNA-PKcs), a member of the phosphoinositide 3-kinase-related kinase (PIKK) family. The primary function of Ku is the recognition of DSB and the recruitment of DNA-PKcs for the successful repair of DNA. The DNA-PKcs carries out the repair by phosphorylating the proteins bound on DNA through its kinase activity and recruiting other proteins involved in repairing double stranded breaks like the DNA ligase.

Until twenty years ago, the NHEJ pathway had been thought to be restricted in eukaryota only. While most of the DNA repair pathways were first identified in prokaryotes and subsequently characterized in eukaryotes, NHEJ pathway was initially discovered in eukaryotes and an evolutionarily related prokaryote homolog of eukaryote NHEJ apparatus was identified and characterized later. Ku genes, one of the major markers for NHEJ pathway in many organisms, are absent in many prokaryotes, including the widely studied bacterial strain Escherichia coli K12. Recently, a large-scale search for ku-like genes suggests the presence of a potential NHEJ repair system in about 25% of the prokaryotes for which a genome was available in public databases. After discovery of prokaryotic Ku, NHEJ apparatus was discovered in many bacterial species. Bacterial NHEJ is performed by an array of proteins of growing complexity in bacteria such as Bacillus subtilis, Mycobacterium tuberculosis, Streptomyces and Sinorhizobium meliloti.

**Ku structure in Eukaryotes**

Ku is a highly abundant protein found in vivo as a stable heterodimer in eukaryotes comprising of two subunits, Ku70 (70kDa) and Ku80 (80kDa). Overall, the evidence suggests that the individual subunits are unstable and require heterodimerization to function. Both Ku70 and Ku80 subunits contain three domains: an N terminal alpha helix/beta barrel von Willebrand A (vWA) domain, a central core domain required for DNA binding and dimerization; and a helical C-terminal domain.

According to the crystal structure of Ku70/80, the two subunits dimerize through the central domain to form a ring to accommodate two turns of double-stranded DNA. The ring is lined with positively charged residues positioned to interact with the negatively charged sugar phosphate backbone of DNA in a sequence-independent manner.

The N-terminal vWA domain consists of a six-stranded beta-sheet in a Rossman fold. The amino side of beta sheet makes contact with the DNA groove while the carboxyl end of the fold is available for protein-protein interaction. Overall, the function of this domain is not well characterized but there is emerging evidence on its role in DSB repair and telomere regulation through interaction with NHEJ factors and the telomere complex components respectively.

Both Ku subunits also contain a helical C-terminal domain (CTD), which is the most divergent region of the two proteins. The Ku80 CTD is approximately 15kDa containing a helical and disordered region which is involved in recruiting DNA-PKCs to DNA in vertebrates while being missing in the lower eukaryotes like Saccharomyces and Arabidopsis. The Ku70 CTD on the other hand is composed of a highly flexible linker region followed by a
structured 5kDa helix-loop-helix region known as the SAP domain\(^{13}\). The SAP domain has putative DNA binding properties and has been shown to increase the overall DNA binding affinity of the heterodimers\(^{14\text{15}}\). Not only this, the C-terminal is also subjected to post translational modifications to regulate Ku interaction with pro-apoptotic proteins and recruitment of homeodomain proteins to DNA ends\(^{16\text{17}}\).

**Ku in Prokaryotes**

In silico approaches provided the first hint at the existence of a bacterial NHEJ leading to the identification of Ku 70/80 homologs and ATP-dependent DNA ligase homologs\(^{18\text{7}}\). Similarly, the first experimental validation of a bacterial NHEJ came from the study of Weller et al which showed that in vitro the *Mycobacterium tuberculosis* LigD (LigDMtb) protein is indeed an ATP-dependent DNA ligase which is stimulated by its cognate KuMtb partner, possibly via interaction between them\(^{19}\). The bacterial Ku proteins are small in size (30-40kDa) as compared to the much larger eukaryotic counterparts (70-80kDa). These smaller bacterial Ku homologues represent a conserved ‘Ku domain’ at the center of the eukaryotic Ku complexes but considerably lack other conserved domains present in the eukaryotic proteins. Among the bacterial species, which are predicted to possess a NHEJ repair system, some actually encode various putative Ku homologs. Examples are actinobacteria of the Streptomyces family, and proteobacteria of the alpha subdivision\(^{8}\). Some of the ku genes are even located on plasmids, implying their putative acquisition via horizontal gene transfer.

In contrast to the heterodimeric Ku complex of eukaryotes, the bacterial Ku complexes are predominantly homodimeric in structure, binding preferentially to the ends of double stranded DNA and possibly sequester other factors, including DNA ligases, forming a primordial DNA repair complex. Evidence for the association of the bacterial Ku proteins with other DNA repair proteins may be evident from the operon structure of the bacterial Ku genes. The Ku homologues are often organized into operons containing ATP-dependent DNA ligases. All of these bacteria also contain essential NAD\(^+\)-dependent DNA ligases; the presence of these additional ATP-dependent ligases suggests a more specific role for these enzymes in vivo roles such as DNA repair/recombination. Often Ku proteins form functional DNA repair complexes with these DNA ligases. To discover the role of Ku and ligase in DSBs repair in prokaryotes, an experiment was performed on mutant *Bacillus subtilis* bearing deactivating mutations in Ku (YkoU) and ligases (Ykovo). These mutants were found to be more sensitive to ionizing radiations in stationary phase and spores\(^{19\text{20}}\).

In this paper we have looked at the origin and evolution of Ku protein in eukaryotes and prokaryotes. We have thoroughly searched for Ku proteins in bacterial, archaeal and eukaryal domains of selected species. The phyla and kingdom where it is found and where it is missing has been clearly indicated. We have further used the beta barrel core domain of Ku from different eukaryotic and prokaryotic species to draw a phylogenetic tree. The phylogenetic relationship between Ku70, Ku80 and prokaryotic Ku has thus been predicted. Further we using the InterPro website, we have predicted the domains of Ku70 and Ku80 in eukaryotes and tried predicting a trend in evolution.

**Results**

**Search of Ku proteins in bacteria**
To study the evolution and history of Ku protein in all domains of life, a list of diverse eukaryotes, archaea and bacteria was compiled and Ku protein was searched in these organisms. The eukaryote genome was searched for the presence of homologs of Ku70 and Ku80 domains of humans, while the archaea and bacteria were searched for the presence of homologs of Ku protein in *Mycobacterium tuberculosis*. Hundred and sixteen genomes covering twenty-eight phyla of bacteria were searched for a presence of Ku. Ku was found only in twenty-five genomes. A few of the bacterial genomes contained multiple Ku proteins. In this list, Ku was found in actinobacteria, alpha proteobacteria, bacteroides/chlorobi, fibrobacteres/acidobacteria, firmicutes, themodesulfobacteria, beta proteobacteria, chlamydia/verrucomicrobia, and delta/epsilon proteobacteria (Table 2). This list is not exhaustive. It is possible that bacterium from any other phylum contain Ku. Those bacteria might just be absent in our list.

**Search of Ku proteins in archaea**

Similarly, Ku from *Mycobacterium tuberculosis* was blasted against archaea (taxid: 2157) in NCBI. It was found that only a number of defined archaenal species contained Ku. These species found in NCBI were *Archeoglobus fulgidus*, *Geoglobus ahangari*, *Archeoglobus venecicus*, *Methanosaeta harundinacea*, *Methanotrix soehngenii*, *Methanobacterium formicicum* and *Methanocella paludicola*.

**Search of Ku proteins in eukaryotes**

To search for Ku in eukaryotes, candidate approach was employed. To makes sure that not only the crown groups were included in the search, genomes of organisms of six different kingdoms were blasted. Ku domain of human Ku70 and Ku80 were used for blast searches. The reason behind including these domains instead of bacterial Ku domain is that search of bacterial Ku domain under normal homology search does not yield any results in eukaryotes (unpublished data). The six kingdoms included in this study are eozoa, animalia, amoebozoa, archaeplastida, fungi and ciliophora (belongs to the group chromista). The organisms are divided into three groups: group with both Ku70 and Ku80, group without Ku80 and group without Ku70 (Table 1).

Out of the twelve Eozoa species we looked into, not a single one had Ku70 or Ku80. Among the 7 amoebozoa species we looked into only *Dictyostelium discoideum* had both Ku70 and Ku80. *Acanthamoeba castellani* had Ku80 protein but lacked Ku70 protein. In the other five species—*Entamoeba invedens*, *Entamoeba histolytica*, *Physarum polycephalum*, *Hyperamoeba dachnya* and *Mastigamoeba balamuthi*—no Ku70 or Ku80 could be found.

Similarly, among the ciliophora, only *Paramecium tetraurelia* had both Ku70 and Ku80. *Oxytrichat心仪allax*, *Stylonychia lemnae*, and *Tetrahymena thermophila* had Ku70 only but no Ku80. The other four species *Euplotes octocarinatus*, *Euplotes crassus*, *Anophryoides haemophila* and *Paramecium caudatum* did not show the presence of Ku70 and Ku80. As of archaeplastida, Arabidopsis thaliana and Volvox carteri contain both Ku70 and Ku80. *Ostreococcus lucimarinus* contains Ku70 but no Ku80. The rest—*Griffithsia japonica*, *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii*, *Scenedesmus obliquus*, *Acetabularia acetabulum*, *Micromonas*, *Glaucocystis nostochinearum*, and *Cyanophora paradoxa*—do not show the presence of Ku70 and Ku80. Among the analyzed fungi, most species show the presence of either Ku70 or Ku80. *Saccharomyces cerevisiae*, *Ustilago maydis*, *Lacaria bicolor*, *Puccinia graminis*, *Schizosaccharomyces pombe*, *Aspergillus oryzae*, and *Talaromyces marneffei* contain both Ku70 and Ku80. Malassezia globosa shows only the presence of Ku70. *Mortierella alpine* and Neurospora crassa show the presence of Ku80, but do not show the presence of
Ku70. *Moniliophthora perniciosa*, *Candida albicans* and *Penicillium marneffei* do not have both Ku70 and Ku80. In animalia *Strongylocentrotus purpuratus, Nematostella vectensis, Hydra vulgaris* and *Monodelphis domestica* contain both Ku70 and Ku80. *Branchiostoma floridae* has Ku70 only. Finally, *Hydra magnipapillata* and *Buddenbrockia plumatellae* have neither Ku70 nor Ku80.

From this analysis, it can be shown that out of the six kingdoms analyzed only organisms in Eozoa contain no Ku70 or Ku80. All other 5 kingdoms contain organisms with either Ku80 or Ku70 or both. It is possible that organisms that do not currently show the presence of Ku domain in their genomes might have incompletely sequenced genome (Supp. Table 3). Upon completing sequencing, they might show the presence of Ku domain. Given that complete sequencing does not change the results, it can be surmised in four of the six domains there is presence of organisms that contain Ku70 only and no Ku80.

**Phylogenetic tree to analyze Ku evolution**

To analyze evolutionary relationship between Ku in bacteria and archaea and Ku70 and Ku80 in eukaryotes a maximum likelihood phylogenetic tree was drawn using full Ku sequences of bacteria and archaea and Ku domains of Ku70 and Ku80 in eukaryotes. The tree is shown in Figure 1. Overall inspection of the tree shows that Ku in bacteria and archaea, Ku domain of Ku70 and Ku domain of Ku80 largely separate into 3 clades. The tree also illustrates that Ku70 and Ku80 share a common node, which is not shared by prokaryotic clade. From this observation it can be comprehended that Ku70 and Ku80 arose from gene duplication event.

**Evolution and history of Ku proteins in prokaryotes**

The bottom 1/3rd of tree in Figure 1 covers single domain Ku protein in bacteria and archaea. A striking observation from this part of the tree is that Ku proteins of a single phylum of bacteria tend to cluster together and often form part of the same clade. This can be observed for alphaproteobacteria, betaproteobacteria, chlamydiae/verrucocicobia, bacteroides, firmicutes, actinobacteria and fibrobacteres. Delta/epsilon proteobacteria do not occur in clusters and are interspersed at different locations in the tree. One beta proteobacteria (*Cupraviridus necator* 2) is separated from the other betaproteobacteria. This might be because this protein arose from horizontal gene transfer from another species of bacteria and not from gene duplication from *Cupraviridus necator* 1. Other multiple Ku proteins present in the same species seem to have originated from gene duplication events since they tend to lie in the same clade (*Niastella koreensis, Saccharothrix espanasensis* and *Bradyrhizobium japonicum*). Seven archaeal Ku genes are found in three different clades. Two of the clades contain 2 archaeal proteins whereas 1 clade contains 3 archaeal proteins. *Archaeoglobus fulgidus, Geoglobus ahangari* and *Archaeoglobusvenenicus* are all related archaea and lie in the same clade. Similarly, *Methanosaeta harundinacea* and *Methanotrix soehngenii* lie in the same clade. Lastly, *Methanobacterium formicicum* and *Methanocella palludicola* lie in the same clade. It can be hypothesized that eukarya arose from archaea and thus eukaryal Ku70 and Ku80 share similarity to these archaeal Ku proteins. The first clade described above is deeply buried in the bacterial clades and thus most likely arose from bacterial horizontal gene transfer. *Methanobacterium* and *Methanocella* Ku proteins lies just below fungal Ku80 and thus it is possible that they gave rise to eukaryal Ku protein. Similarly, *Methanosaeta and Methanotrix* Ku protein lies below *Methanobacterium* and *Methanocella* Ku proteins clade. The lack of Ku protein in Asgard superphylum of archaea, the closest archaea to eukaryotes, and Eozoa, the earliest branching eukaryotes, makes the vertical inheritance of Ku from archaeal lineages to eukaryotes unlikely (unpublished data).
Evolution and history of Ku70 proteins

At the top of the tree, lies the entire cluster of Ku70 proteins included in the phylogenetic tree. All the Ku70 proteins included in the phylogenetic tree cluster together. Additionally, Ku70 proteins from the same kingdom tend to occur in the same clade with a few exceptions. One exception to this rule is the Ku70 protein from Ostreococcus lucimarinus. While Ku70 from other archaeplastida, Volvoxcarteri and Arabidopsis thaliana, cluster together O. lucimarinus Ku70 occurs separately. This might result from distinction of O. lucimarinus Ku70 from other species of archaeplastida.

Similarly, another exception is a Ku70 protein from Tetrahymena thermophilia. While the rest of the ciliophora Ku70 occur in the same clade, hinting at a common source of origin, one of the three Ku70s of T. thermophilia lies within the Ku80 cluster. Of the other two Ku70s of T. thermophilia, one lies in the same clade as P. tetraurelia and other branches out earlier. It can be hypothesized that earlier branching occurred in the common ancestor of the two species.

Ku70s of the rest of the fungi from S. cerevisiae to T. mameffei all cluster together, suggesting common ancestry. In animalia kingdom, Ku70 from H. sapiens and M. domestica, the two mammals, occur in the same clade and suggesting a common ancestry. Ku70 protein from other animalia species diverged earlier, but all the animalia Ku70 occur in the same clade. Ku70 in Monosiga breviskolis, a choanoflagellate, lies just below animalia clade hinting similarities between them. Ku70 from the only amoebozoa, D. discoideum, is sandwiched between the kingdoms ciliophora and archaeplastida.

Evolution and history of Ku80 proteins

In between the prokaryotic and Ku70 proteins cluster, lie the rest of the Ku80 proteins interspersed with one Ku70 protein. All the animalia and amoebozoa Ku80 proteins cluster along their kingdoms. The third T. thermophilia Ku70 lies next to P. tetraurelia Ku80 protein. T. thermophilia has no Ku80 protein, and this Ku70 protein can be considered its Ku80 protein. Seven of the 10 Ku80 proteins from fungi cluster together and the other three lies apart. A. castellani has one Ku80 protein that clusters with D. discoideum Ku80 and it is devoid of Ku70 protein.

On the other hand, two archaeplastidae Ku80s of A. thaliana and V. carteri are apart from each other. V. carteri Ku80 lies next to O. lucimarinus Ku70 suggesting its similarity to archaeplastidae Ku70 protein. Strikingly, the Ku80 protein of Arabidopsis thaliana comes up when Ku70 of Homo sapiens is blasted in the A. thaliana genome. This is the only time when blasting Ku70 pulls up Ku80. This instance suggests that Ku70 and Ku80 possibly arose through gene duplication of a common ancestor. Finally, all the animalia Ku80s lie in the same clade where H. sapiens and M. domestica lies in one nested clade and N. vectensis and H. vulgaris lies in the other hinting a common ancestry.

NJ tree and ML tree with Streptomyces species were drawn (Supplementary Figure 1 and Figure 2). The conclusions drawn earlier does not change with the trees.

Domain architecture of Ku70 protein

From Figure 2, we can perceive that the most common architecture of Ku70 protein across different kingdoms is the core Ku domain flanked by N and C terminus. The only amoebozoa, D. discoideum, which has this protein has this domain architecture in addition to APLF_PBZ domain at the end of C-terminus. Without a clear
evolutionary trend, a number of Ku70s across all kingdoms except Eozoa and Amoebozoa, contain an SAP domain next to the C-terminal domain. At the N-terminus of some Ku70 proteins lies the VWF domain.

**Domain architecture of Ku80 proteins:**

Figure 3 displays various Ku80 domain architecture found within different kingdoms. Ku80 protein of Homo sapiens and M. domesticus contains beta barrel Ku domain flanked by C-terminal and N-terminal domains. At the end of the N-terminal end lies the Ku PK binding domain. This domain architecture is conserved in all Ku80 proteins of the animalia kingdom except Nematostella vectensis, which has missing Ku PK binding domain and N-terminal domain. Since only this species has missing domains, it is possible that during speciation of this species these two domains were lost. All fungi Ku80 have missing C-terminal domain. This shows that C-terminal domain possibly evolved after Ku80 diverged in the common ancestor of fungi and animalia. Almost all fungi Ku80 possess Ku PK binding domain, suggesting that the common Ku80 ancestor of fungi and animalia had Ku PK binding domain. As for the N-terminal domain, half the fungi have it, while other half don't have it suggesting it was either lost in some fungi or some fungi and all animalia acquired it after the two kingdoms diverged. Some of the fungi also have VWF domain at the N-terminus of Ku80 protein, although this domain is mostly found in Ku70 proteins. Both amoebozoa Ku80s have an N-terminal domain with VWF domain attached. They both lack C-terminal domain and Ku PK binding domain. Another kingdom, archaeplastida has one A. thaliana Ku80 and one V. carteri Ku80. A. thaliana Ku80 has the same domain architecture as most animalia Ku80s. V. carteri Ku80 has Ku domain flanked Ku PK binding domain. From this domain architecture it has hard to reconcile what the ancestral archaeaplastida Ku80 looked like. Like in fungi, ciliophora Ku80s have missing C terminal domain. It can be hypothesized from domain architecture data that C-terminal domain was missing in the ancestral eukaryotic Ku80. It was later acquired by species in the animalia kingdom and A. thaliana.

**Discussion**

**Horizontal inheritance of Ku protein between prokaryotes and eukaryotes**

There are two possibilities for the inheritance of Ku protein between prokaryotes and eukaryotes. It is possible that Ku first arose in the prokaryotes and their vertically or horizontally got transferred to eukaryotes. Alternatively, Ku first arose in the eukaryotes and horizontally transferred to some prokaryotes. In eukaryotes, Ku domain is often flanked by other domains. If it indeed transferred from eukaryote to prokaryote, then only the Ku domain was transferred. Since Ku protein is often a simple, one domain protein in prokaryotes, it is more likely that this domain transferred onto later arising eukaryotes and diverged with multiple flanking domains. The first possibility of vertical transmission of Ku from prokaryote to eukaryote is negated by the findings of this paper. If there were indeed a vertical transmission of the protein than a lot more than 7 archaeal species would be expected to contain Ku protein. Especially the pronounced absence of Ku protein in Asgard superphylum of archaea (the closest prokaryote to eukaryote) and in the Eozoa kingdom of eukarya, indicates that this vertical inheritance from prokaryote to eukaryote is unlikely. It is possible that a number of archaea and eozoa lost Ku during evolution due to some selective pressure, but this hypothesis is very unlikely.

**Evidence that Ku in bacteria, and Ku70 and Ku80 in eukaryotes arise from a common ancestor**
Ku in prokaryotes was discovered from PSI blast to the second interaction of Ku70 and Ku80 Ku domains in eukaryotes. The sequence homology between Ku in bacteria and Ku in eukaryotes is very loose. Regular blast search of Ku in bacteria cannot find Ku70 and Ku80 Ku domains. This loose sequence homology raises questions about the common origin of bacterial Ku, Ku70 and Ku80 proteins. In this paper there are several evidences that indicate the common origin of these three proteins. Ku protein from *Methanobacterium formicicum* and *Methanocella paludicola*, although derived from blast search using *M. tuberculosis* Ku, share a common ancestral node with Ku70 and Ku80. This shows that these two Ku proteins possibly arose from bacterial Ku before giving rise to Ku70 and Ku80 in eukaryotes. To validate this point, a search for eukaryal Ku using these two archaeal Ku gives Ku70 and Ku80 protein in eukaryotes using regular homology search. This shows that bacterial and archaeal Ku proteins are closer to Ku70 and Ku80 proteins that originally demonstrated. It is possible that earliest eukaryotic Ku proteins evolved from horizontal transfer of archaeal Ku protein close to Ku from these two species. Although in the phylogenetic tree, Ku70, Ku80 and prokaryotic Ku largely separate into three clusters a few Ku proteins are found not in their cluster. An example of this is the Ku70 protein of *Tetrahymena thermophila*. This protein lies in the cluster of Ku80 proteins. This demonstrates that Ku70 and Ku80 proteins are closely related. Another such example comes from *A. thaliana* Ku80 protein. When Ku70 Ku domain of *Homo sapiens* is blasted in the *A. thaliana* genome it not only pulls Ku70 protein of *A. thaliana*, but also Ku80 protein of the species. This shows that Ku80 protein of *A. thaliana* is closely related to Ku70 protein.

**Evidence of vertical inheritance of Ku proteins:**

Across domains, Ku seems to have been inherited from horizontal gene transfer. But when inspected closer within a phylum in bacteria and archaea and within kingdoms in eukaryotes, there is a strong evidence for vertical gene transfer. In the phylogenetic tree, species in a phylum or kingdom largely cluster together. For example, all the members of animalia kingdom cluster together for both Ku70 and Ku80. This suggests that the ancestors of individual kingdom and phylum arose from common ancestors.

**A comprehensive model of Ku inheritance:**

A search for Ku protein in 116 genomes of bacteria yielded positive results in only 25 genomes. Our search yielded positive result for 1/4th of the bacterial species. This demonstrates that Ku protein is not a particularly rare protein in bacteria. In archaea, we found Ku protein in about 7 species among all sequenced archaeal genomes in NCBI. This demonstrates that Ku protein in rare in archaea. Similarly, when Ku was searched in 59 eukaryotic species, it was found in about 27 species. The rest 34 species lacked Ku protein. Some of these species may not show presence of Ku protein because their genome is not fully sequenced. Others may not have Ku protein. Given that no Eozoa had Ku protein it can be hypothesized that Ku protein got horizontally transferred from archaea to eukaryotes after Eozoa diverged. This single Ku protein duplicated and evolved into Ku70 and Ku80 proteins, which are inherited by most Neozoa. Some Neozoa, for some reason, lost Ku70 or Ku80 or both proteins. This model has been depicted in Fig. 4. The NHEJ machinery in bacteria and archaea are different from NHEJ machinery in eukarya. The polymerases, nucleases and ligases in prokaryotes and eukaryotes are distinct. It is possible that only Ku got horizontally transferred from prokaryote to eukaryote. The other proteins involved in NHEJ in eukaryotes then evolved independently from prokaryotic NHEJ machinery.

**Domain architecture of Ku70 and Ku80 proteins:**
Analysis of domain architecture of Ku70 among our 59 species reveals at least 9 different architectures. Often times the central Ku beta barrel domain is flanked by N- and C- terminal domains. At the end of the C terminus lies the SAP domain, whereas at the end of the N terminus lies the VWF domain. Similarly, an analysis of the Ku80 proteins among the 59 species reveals at least 9 different domain architectures. The central Ku beta barrel domain is flanked by N- and C- termini. At the end of the C terminus often lies the DNA PK binding domains, whereas at the end of the N terminus sometimes lies the VWF domain. To further broaden our perspective, we looked at different domain architectures of Ku beta barrel domain in InterPro website. About 100 different domain architectures were discovered. From this data, it can be discerned that a combination of different domains can arise in individual species during evolution. This combinatorial event is a common phenomenon during evolution of Ku. All six types of rearrangement events: fusion, fission, terminal addition, terminal deletion, single domain loss, single domain emergence is possibly active to bring about the evolutionary combinations of domains described above. Such frequent combinatorial events also explain the rationale for the existence of introns in eukaryotes. It is widely believed that introns exist so that exons, which often encode for complete protein domains, can be combined during evolutionary process to give rise to a wide range of proteins.

**Proposal of future experiments:**

Study of Ku proteins in eukaryotes raises several questions that can be solved through experiments. A number of eukaryotes only contain one of the two Ku proteins, Ku70 or Ku80. Does the presence of only one Ku protein suffice to make up an intact NHEJ machinery? When one of the two Ku proteins is knocked out in mice and humans, cells are deficient in NHEJ. Severe DNA repair phenotypes and immunological disorders arise. Does this happen organisms with only one copy of Ku? Another unanswered question is whether Ku from lower eukaryotes and eukaryotes with only one copy homodimerize in vitro. It is known that Ku in prokaryotes homodimerizes whereas Ku in higher eukaryotes heterodimerizes. It would be interesting to clone and express such Ku proteins and test them for homo-/hetero-dimerization.

Another interesting question that can be studied computationally or through wet laboratory experiments is when the NHEJ machinery started getting involved in adaptive immunity. The earliest eukaryotes that show adaptive immunity are jawed fish. It has been hypothesized that these organisms developed adaptive immunity when they acquired recombination-activating genes (RAGs) from transposons. Did NHEJ in these organisms immediately get involved in adaptive immunity, or were there changes in NHEJ that led them to their role in adaptive immunity in higher organism?

**Conclusion**

On contrary to the estimation that prokaryotic and eukaryotic Ku proteins belonged to different clades, our analysis shows a common node of origin. This led us to hypothesize that Ku from archaea transferred through horizontal gene transfer onto neozoa and then duplicated to form Ku70 and Ku80. Study of Ku proteins in prokaryotes and eukaryotes is both important. In certain eukaryotes NHEJ is important for survival, and study of Ku leads to an understanding of how this is possible. Additionally, since many human pathogens possess a NHEJ repair system, understanding bacterial NHEJ might be of therapeutic interest. Developing anti-NHEJ compounds might be a way to move forward since these compounds might hinder chances of pathogens to survive in genotoxic conditions in certain non-dividing states in different mammal hosts. However, many studies are yet to be conducted and many questions are yet to be answered in this particular field of study.
Methods

Sequence retrieval:

Ku protein sequence of *Mycobacterium tuberculosis* was retrieved from UniProt (https://www.uniprot.org/uniprot/P9WNV3). 116 prokaryotes from different phylum were selected and presence of Ku protein sequence was searched using BLASTP in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). For archaea, homologs of Ku sequence of *Mycobacterium tuberculosis* were searched using BLASTP against archaea (taxid: 2157) in NCBI. Similarly, for eukaryotes, *Homo sapiens* KU70(XRCC6) and KU80(XRCC5) protein sequences were retrieved from UniProt (https://www.uniprot.org/uniprot/P12956) (https://www.uniprot.org/uniprot/P13010). Using InterPro, KU70/80 beta barrel domain sequence length was determined and the retrieved sequences from UniProt were cleaved accordingly (https://www.ebi.ac.uk/interpro/search/sequence/). Presence of KU70/80 were searched in 59 eukaryotes selected using BLASTP in NCBI. Default blast parameters were used. During BLASTP of some organisms, repeated and highly similar sequences were displayed as a result. From these, only one sequence was retrieved for an organism unless the listed sequences were highly divergent.

Domain specification:

Using InterPro, KU70/80 beta barrel domain sequence length of all organisms were determined, retrieved sequences were cleaved and subjected to phylogenetic tree construction. KU N terminal, KU 70/80 Beta barrel domain, KU C terminal, and various other domain's sequence lengths were tabulated using Microsoft Excel-2010 (Supp. Tables 1 & 2). Various domain architectures of some primary organisms representing their respective phylum were drawn.

Phylogenetic tree construction:

All the sequences retrieved were named as Genus species | Accession number. Phylogenetic tree was constructed using MEGA-X desktop software downloaded from (https://www.megasoftware.net/). All the sequences were aligned using Muscle and Maximum Likelihood statistical method was employed. Various criteria were selected from available drop-down list. Test of phylogeny was done using 500 Bootstrap replications. Substitution model used was WAG model. Gamma Distributed with Invariant sites (G + I) rates was used with 5 number of discrete gamma categories. Partial deletion was used for Gaps/Missing data treatment with 95% site coverage cutoff. Nearest-Neighbor-Interchange (NNI) ML Heuristic method was used and Make initial tree automatically (Default – NJ/BioNJ) was chosen for Initial Tree for ML. Branch Swap filter was none and Number of threads selected was 3.

Abbreviations

DSBs: Double Stranded Breaks

NHEJ: Non-homologous End Joining

HR: Homologous recombination

DNA-PK: DNA-dependent protein kinase
PIKK: phosphoinositide 3-kinase-related kinase
vWA: von Willebrand A
CTD: C-terminal Domain
NCBI: National Centre for Biotechnology Information
UniProt: Universal Protein Resource
BLAST: Basic Local Alignment Search Tool
BLASTP: Protein-protein BLAST
MEGA: Molecular Evolutionary Genetics Analysis

Declarations

Competing interests
Authors have no conflict of interest to declare

Ethics approval and consent to participate
This paper does not require ethical approval and no human subjects are used

Consent for publication
No such consent required

Funding
No funding source available

Availability of data and material
The data for this paper has been derived from publicly available database

Author contributions
The project was conceived and designed by HKB. AM and SR looked up the databases, searched for sequences and drew family tree and domain organization. AM and SR wrote the introduction and methods section, HKB wrote the results and discussion section. All authors read the paper and made corrections.

Author contributions
We would like to thank Simon K. Shrestha for suggestions to the paper.

References
1. Bowater R, Doherty AJ. Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet. 2006. doi:10.1371/journal.pgen.0020008.

2. Lieber MR. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annu Rev Biochem. 2010. doi:10.1146/annurev.biochem.052308.093131.

3. Boulton SJ, Jackson SP. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. EMBO J. 1996. doi:10.1002/j.1460-2075.1996.tb00890.x.

4. Gu Y, Jin S, Gao Y, Weaver DT, Alt FW. Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. Proc Natl Acad Sci U S A. 1997. doi:10.1073/pnas.94.15.8076.

5. Mimori T, Hardin JA, Steitz JA. Characterization of the DNA-binding protein antigen Ku recognized by autoantibodies from patients with rheumatic disorders. J Biol Chem. 1986.

6. Mari PO, Florea BI, Persengiev SP, et al. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. Proc Natl Acad Sci U S A. 2006. doi:10.1073/pnas.0609061103.

7. Doherty AJ, Jackson SP, Weller GR. Identification of bacterial homologues of the Ku DNA repair proteins [2]. FEBS Lett. 2001. doi:10.1016/S0014-5793(01)02589-3.

8. McGovern S, Baconnais S, Roblin P, et al. C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. Nucleic Acids Res. 2016. doi:10.1093/nar/gkw149.

9. Fell VL, Schild-Poulter C. The Ku heterodimer: Function in DNA repair and beyond. Mutat Res - Rev Mutat Res. 2015. doi:10.1016/j.mrrev.2014.06.002.

10. Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair. Nature. 2001. doi:10.1038/35088000.

11. Gell D, Jackson SP. Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. Nucleic Acids Res. 1999. doi:10.1093/nar/27.17.3494.

12. Singleton BK, Torres-Arzayus MI, Rottinghaus ST, Taccioli GE, Jeggo PA. The C Terminus of Ku80 Activates the DNA-Dependent Protein Kinase Catalytic Subunit. Mol Cell Biol. 1999. doi:10.1128/mcb.19.5.3267.

13. Aravind L, Koonin EV. SAP - A putative DNA-binding motif involved in chromosomal organization. Trends Biochem Sci. 2000. doi:10.1016/S0968-0004(99)01537-6.

14. Hu S, Pluth JM, Cucinotta FA. Putative binding modes of Ku70-SAP domain with double strand DNA: A molecular modeling study. J Mol Model. 2012. doi:10.1007/s00894-011-1234-x.

15. Wang J, Dong X, Reeves WH. A model for Ku heterodimer assembly and interaction with DNA. Implications for the function of Ku antigen. J Biol Chem. 1998. doi:10.1074/jbc.273.47.31068.

16. Cohen HY, Lavu S, Bitterman KJ, et al. Acetylation of the C terminus of Ku70 by CBP and PCAF controls Bax-mediated apoptosis. Mol Cell. 2004. doi:10.1016/S1097-2765(04)00094-2.

17. Kim KB, Kim DW, Park JW, et al. Inhibition of Ku70 acetylation by INHAT subunit SET/TAF-Iβ regulates Ku70-mediated DNA damage response. Cell Mol Life Sci. 2014. doi:10.1007/s00018-013-1525-8.

18. Aravind L, Koonin EV. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res. 2001. doi:10.1101/gr.181001.
19. Weller GR, Kysela B, Roy R, et al. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science (80-)*. 2002. doi:10.1126/science.1074584.

20. Bhattarai H, Gupta R, Glickman MS. DNA ligase C1 mediates the LigD-independent nonhomologous end-joining pathway of Mycobacterium smegmatis. *J Bacteriol*. 2014;196(19):3366–76. doi:10.1128/JB.01832-14.

21. Petitjean C, Deschamps P, López-García P, Moreira D. Rooting the domain archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol Evol*. 2014. doi:10.1093/gbe/evu274.

22. Postberg J, Forcob S, Chang WJ, Lipps HJ. The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol*. 2010. doi:10.1186/1471-2148-10-259.

23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018. doi:10.1093/molbev/msy096.

24. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001. doi:10.1093/oxfordjournals.molbev.a003851.

25. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol*. 2013. doi:10.1093/molbev/mst012.

26. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 1985. doi:10.2307/2408678.

27. Cavalier-Smith T. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett*. 2010. doi:10.1098/rsbl.2009.0948.

28. Dohmen E, Klasberg S, Bornberg-Bauer E, Perrey S, Kemena C. The modular nature of protein evolution: Domain rearrangement rates across eukaryotic life. *BMC Evol Biol*. 2020. doi:10.1186/s12862-020-1591-0.

29. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. *Molecular Biology of the Gene*. 7th ed. Pearson; 2013.

30. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: Genetic events and selective pressures. *Nat Rev Genet*. 2010. doi:10.1038/nrg2703.

**Tables**

Table 1: List of eukaryotes against which NCBI BLASTP was performed using *Homo sapiens* Ku70 and Ku80 proteins as query. *Homo sapiens* Ku70(XRCC6) and Ku80(XRCC5) protein sequences were retrieved from UniProt and Ku 70/80 beta barrel domain sequence length was determined using InterPro and the sequence was cleaved accordingly. Presence of Ku70/80 were searched in 59 eukaryotes selected using BLASTP in NCBI. Default blast parameters were used. (a) Eukaryotes consisting of both Ku70 and Ku80 (b) Eukaryotes consisting of Ku70 but devoid of Ku80 (c) Eukaryotes consisting of Ku80 but devoid of Ku70 were listed. The entire list of Eozoa, which we had taken, were found to be devoid of both Ku80 and Ku70, against which BLASTP was performed.
| Animalia               | Fungi                  | Amoebozoa         | Archaeplastidae | Ciliophora          |
|-----------------------|------------------------|-------------------|-----------------|---------------------|
| *Strongylocentrotus purpuratus* | *Ustilago maydis*       | *Dictyostelium discoideum* | *Arabidopsis thaliana* | *Paramecium tetraurelia* |
| *Nematostella vectensis* | *Lacaria bicolori*      |                   |                 | *Volvox carteri*    |
| *Hydra vulgaris*        | *Puccinia graminis*     |                   |                 |                     |
| *Monodelphus domestica* | *Schizosaccharomyces pombe* |                   |                 |                     |
| *Homo sapiens*          | *Aspergillus oryzae*    |                   |                 |                     |
| *Monosiga brevicolis*   | *Talaromyces marneffei* |                   |                 |                     |
|                       | *Saccharomyces cerevisiae* |                   |                 |                     |

a.
| Animalia | Fungi | Amoebozoa | Eozoa | Ciliophora | Archaeplastidae |
|----------|-------|-----------|-------|------------|----------------|
| Hydra magnipapillata | Malassezia globosa | Entamoeba invedens | Trypanosoma cruzi | Oxytricha trifallax | Griffithsia japonica |
| Buddenbrockia plumatellae | Candida albicans | Entamoeba histolytica | Trypanosoma brucei | Euplotes crassus | Ostreococcus lucimarinus |
| Branchiostoma floridanae | Moniliophthora perniciosa | Physarum polycephalum | Leishmania infantum | Euplotes octocarinatus | Cyanidioschyzon merolae |
| | Penicillum marneffei | Hyperamoeba dachnya | Leishmania major | Anophryoides haemophila | Chlamydomonas reinhardtii |
| Mastigamoeba balamuthi | Diplonema papillatum | Stylonychia lemnae | Micromonas | | |
| Giardia lamblia | Tetrahymena thermophila | Scenedesmus obliquus | | | |
| Spironucleus vortens | Acetabularia acutabulum | | | | |
| Streblomastix strix | Glaucocystis nostochinearum | | | | |
| Euglena gracilis | Cyanophora paradoxa | | | | |
| Sawyeria marylandensis | | | | | |
| Trichomonas vaginalis | | | | | |
| Reclinomonas americana | | | | | |

b.
| Animalia        | Fungi                      | Eozoa             | Amoebozoa          | Ciliophora          | Archaeplastidae     |
|-----------------|----------------------------|-------------------|--------------------|--------------------|---------------------|
| Buddenbrookia plumatellae | Mortierella alpina         | Trypanosoma cruzi | Entamoeba histolytica | Euploites octocarinatus | Griffithsia japonica |
| Hydra magnipapillata | Moniliophthora perniciosa | Trypanosoma brucei | Physarum polycephalum | Euploites crassus   | Cyanidioschyzon merolae |
| Neurospora crassa | Leishmania major           | Mastigamoeba balamuthi | Anophryoides haemophila | Chlamydomonas reinhardti |
| Candida albicans | Diplonema papillatum       | Hyperamaeoba dachnya | Paramecium caudatum | Scenedesmus obliquus |
| Penicillium marneffei | Giardia lamblia           | Acanthamoeba castellanii | Micromonas         |
| Spironucleus vortens | Entamoeba invedens        |                   |                    |                    |
| Trichomonas vaginalis |                           |                   |                    |                    |
| Streblomastix strix |                             |                   |                    |                    |
| Sawyeria marylandensis |                            |                   |                    |                    |
| Euglena gracilis |                              |                   |                    |                    |
| Reclinomonas americana |                                |                   |                    |                    |
| Leishmania infantum |                                       |                   |                    |                    |

Table 2: List of ten phyla consisting of prokaryotes with prokaryotic Ku protein. These prokaryotes were obtained using NCBI BLASTP using *Mycobacterium tuberculosis* Ku protein sequence as input query. Seven archaea shown within the dark box in the table were found to contain Ku protein sequence. Ku protein sequence of *Mycobacterium tuberculosis* retrieved from UniProt was used as query sequence for BLASTP against 116 prokaryotes from different phylum selected\(^\text{21}\). For archaea, homologs of Ku sequence of *Mycobacterium tuberculosis* were searched using BLASTP against archaea (taxid: 2157) in NCBI.
| Actinoplanes | Alphaproteobacteria | Euryarchaeota | Bacteroidetes/chlorobi |
|-------------|---------------------|--------------|------------------------|
| Actinoplanes WP_014687497.1 | Bradyrhizobium japonicum BAL07922.1 | Archaeoglobus fulgidus 028548.2 | Niastella koreensis AEV97944.1 |
| Nocardia brasiliensis AFU03477.1 | Bradyrhizobium japonicum BAL13505.1 | Archaeoglobus veneficus AEA46245.1 | Niastella koreensis AEW00550.1 |
| Rhodococcus jostii ABG98966.1 | Bradyrhizobium japonicum BAL07271.1 | Methanocella paludicola WP_012899333.1 | Chitinophaga pinensis ACU58495.1 |
| Saccharothrix espanaensis CCH315712.1 | Rhizobium leguminosarum bv. viciae CAK07635.1 | Methanothrix soehngenii WP_013720546 |
| Streptomyces sviceus EDY54145.1 | | Methanoseta harundinacea KUK05950.1 |
| Saccharothrix espanaensis CCH31573.1 | | Geoglobus ahangari WP_048095863.1 |
| | | Methanobacterium formicicum CEL24363.1 |

| Fibobacteres/acidobacteria | Firmicutes | Thermodesulfobacteria |
|---------------------------|------------|----------------------|
| Candidatus Solibacter | Paenibacillus mucilaginosusAFH65824.1 | Thermodesulfatator indicus AEH45291.1 |
| Terriglobus roseus AFL88371.1 | Desulfovosporosinus orientis AFH65824.1 |
| Acidobacterium capsulatum C1F735.1 | Bacillus thuringiensis AFU11505.1 |
| Granulicella mallensis AEU35712.1 | | |
Table 3: List of prokaryotic phyla without Ku protein. No any prokaryotic Ku homologs were found in prokaryotes within these 22 phyla during BLASTP search against *Mycobacterium tuberculosis* Ku protein sequence as input query. The list of prokaryotes selected for this research purpose was not extensive and this result does not assure that every prokaryote within these phyla are devoid of prokaryotic Ku protein.

| Chlamydiae/verucomicrobia                  | Delta-epsilon proteobacteria               | Betaproteobacteria                       |
|-------------------------------------------|-------------------------------------------|------------------------------------------|
| *Parachlamydia acanthamoebae* CCB86018.1  | *Desulfomonile tiedjei* AFM25046.1         | *Cupriavidus necator* AEI81269.1         |
|                                           |                                           |                                          |
| *Opitutus terrae* B1ZWL2.1                | *Haliangium ochraceum* ACY15831.1         | *Cupriavidus necator* AEI82225.1         |
|                                           |                                           |                                          |
|                                           | *Paraburkholderia xenovorans* ABE30641.1  |                                          |
|                                           | *Achromobacter xylosoxidans* ADP13795.1    |                                          |
|                                           |                                           | *Delftia acidovorans* ABX36971.1         |
| Phylums                      |
|-----------------------------|
| Aquiflora                   |
| Chloroflexi                 |
| Chrysiogenetes              |
| Cyanobacteria               |
| Deferribacteres             |
| Deinococcus-thermus         |
| Dictyoglomi                 |
| Elusimicrobia               |
| Fusobacteria                |
| Gemmatimonadetes            |
| Nitrospirae                 |
| Plantomycetes               |
| Proteobacteria              |
| Spirochaetes                |
| Synergistetes               |
| Tenericutes                 |
| Thermotogae                 |
| Caldiserica                 |
| Aigarchaeota                |
| Thaumarchaeota              |
| Crenarchaeota               |
| Korarchaeota                |

**Figures**
Figure 1

Phylogenetic tree inferred by using the Maximum Likelihood method and Whelan And Goldman model\textsuperscript{24} in MEGA-X. The bootstrap consensus tree inferred from 500 replicates \textsuperscript{26} is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Particular color is assigned to each phyla and kingdom and organism labels are colored accordingly. In addition, corresponding phyla/kingdoms are labelled at left of the tree. From this tree it can be inferred that the Ku70, Ku80 and prokaryotic Ku largely separate into 3 clades. Some of the archaeal Ku share node with Ku70 and Ku80 hinting at the common origin.
Figure 2

KU N terminal, KU 70/80 Beta barrel domain, KU C terminal, and various other domain's architecture determined using InterPro. Ku 70 domain architecture of (a) Branchiostoma florigae and Homo sapiens, both sharing similar architecture and Strongylocentrotus purpurat representing Animalia. Nematostella vectensis and Monosiga brevicolis also share similar architecture to Homo sapiens. Likewise, Hydra Vulgaris and Monodelphis domestica has domain architecture similar to Strongylocentrotus purpurat. (b) Malassezia globosa, Ustilago maydis, Puccinia graminis, Saccharomyces cerevisiae and Talaromyces marneffei representing Fungi. Laccaria bicolor, Schizosaccharomyces pombe and Aspergillus oryzae has similar architecture to Ustilago maydis. Two
Ku70 sequences of Puccinia graminis were found to exhibit two different domain architecture one of which was similar to that of Saccharomyces cerevisiae. (c) Oxytricha trifallax, Stylocydia leunae and Tetrahymena thermophila representing Ciliophora. Three sequences of T. thermophila obtained during BLASTP were found to exhibit three different domain architecture. Paramecium tetraurelia was found to exhibit similar domain architecture as that of Oxytricha trifallax. (d) Dictyostelium discoideum representing amoeboza and (e) Arabidopsis thaliana and Ostreococcus lucimarinus representing archaeplastidae. Volvox carteri shares similar architecture to that of Ostreococcus lucimarinus.

Figure 3

KU N terminal, KU 70/80 Beta barrel domain, KU C terminal, and various other domain's architecture determined using InterPro. Ku 80 domain architecture of (a) Stronglyocentrotus purpuratus, Homo sapiens and Nematostella vecentisis representing Animalia. Other animalia Hydra vulgaris and Monodelphus domestica share similar architecture to that of Stronglyocentrotus purpuratus. (b) Puccinia graminis, Mortierlla alpine, Ustilago maydis and Lacaria bicoloris representing domain structures of Fungi. Ustilago maydis and Neurospora crassa share common domains. Also, Mortierlla alpine, Aspergillus oryzae, Sphaerofoama artica and Taleromyces marneffei share common domains. Similarly, Puccinia graminis and Saccharomyces cerevisiae share common domains. Lastly, Lacaria bicoloris share common domains with Schizosaccharomyces pombe. (c) Acanthamoeba castellanii representing amoeboza which also share common domain with Dictyostelium.
discoideum. (d) Arabidopsis thaliana and Volvox carteri representing archaeplastidae and (e) Paramecium tetraurelia representing Ciliophora. Domain specifications were obtained using Interpro and the architectures were drawn using MS WORD-2010.

Figure 4

A model of inheritance of Ku across domains. The figure above describes phylogenetic classification of the domain eukarya into kingdoms. The earliest branching Eozoa, divided into Euglenozoa and Excavata, do not contain Ku proteins. The eukaryotic lineage that evolved after Eozoa branched out possibly received a copy of Ku protein from horizontal gene transfer. This copy possibly duplicated into Ku70 and Ku80 proteins, which were then vertically inherited to kingdoms animalia, fungi, amoebozoa, chromista and archaeplastida. The Ku70 and Ku80 proteins then added and lost domains and sometimes got entirely eliminated from the genome as Neozoa evolved.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementBMCKupaper.docx