A Study of Phylogenetic Relationships and Homology of Cytochrome C using Bioinformatics

Kulkarni Keya and Sundarrajan Priya

St. Xavier’s College (Autonomous), Mumbai
*Corresponding author Email: priya.s@xaviers.edu

Manuscript Details

Received: 17.02.2016
Revised: 22.03.2016
Accepted: 26.07.2016
Published: 18.09.2016

ISSN: 2322-0015

ABSTRACT

Cytochrome C is an essential and ubiquitous protein. It is highly conserved across the spectrum of species and hence is used to study cladistics. In this paper, analysis of cytochrome C using bioinformatics tools has been carried out, in order to find the evolutionary relationship between species. The amino acid sequences of cytochrome C were used to establish the phylogenetic relationships of species. The phylogeny inferred from the sequence alignment of cytochrome C was confirmed by comparing its structure and function. The protein was found to be functionally redundant. From the analysis cytochrome C, it appears that two different organisms which have the same or even similar protein sequences are genealogically related. The homologous similarities of cytochrome C are suggestive of common ancestry.

Keywords: Cytochrome C, phylogenetic relationships, homology, molecular clock.

INTRODUCTION

The Cytochrome complex (cyt c) is a small hemeprotein loosely associated with the inner mitochondrial membrane. It belongs to the cytochrome C family of proteins and is found in all organisms. Cyt c is a highly soluble protein, unlike other cytochromes. In humans, cyt c is encoded by the CYCS gene [UniProtKB - P99999]. Cyt c is an essential and ubiquitous protein. It has a crucial role in the electron transport chain as well as in apoptosis as shown in Figure 1).

Cyt c is highly conserved across a wide spectrum of species. This implies that it has changed little in millions of years of evolution. This, along with its small size (molecular weight of 12,000 Daltons), makes it useful in studying cladistics [UniProtKB - P99999].

The cyt c molecule has been studied for the glimpse it gives into evolutionary biology. The idea of a “molecular clock” is attributed to Emile Zuckerkanzl and Linus Pauling who noticed that the number of amino acid differences in hemoglobin and cyt c between different lineages changes linearly with time, as estimated from fossil evidence.
They suggested that the rate of evolutionary change of any specified protein was approximately constant over time and over different lineages [Zuckerkandl and Pauling, 1962]. Knowledge of the rate of molecular evolution in certain sets of lineages also facilitates establishing the dates of phylogenetic events, including those not documented by fossils, such as the divergence of living taxa and formation of the phylogenetic tree. [Wilson et al, 1977].

Phylogeny consists of the evolutionary relationships between any set of species. Phylogeny is determined using a ubiquitous protein or gene. This is done to ensure that the comparisons are independent of the overall species phenotype. For example, both humans and chimpanzees have many similar anatomical characters and functions, so we might expect their proteins to be similar, regardless of whether they are genealogically related or not. However, we can compare the sequences of basic genes that are used by all living organisms, such as the cyt c gene, which have no influence over specific chimpanzee or human characteristics.

In this paper, cyt c of different species is compared in order to determine an evolutionary relationship between them. Cyt c is an ancient molecule, and it has evolved very slowly. Only one-third of its amino acids are unchanged. This conservatism is a great help in working out the evolutionary relationships between distantly-related creatures. [Kumar S, 2005] Cyt c is a protein involved in using energy in the cell. It is found in most species. Over time, random mutations in the DNA sequence occur. As a result, its amino acid sequence also changes. Cells without usable cyt c are unlikely to survive. [Margoliash E, 1963]. Hence, the relationship between organisms can be compared by examining the amino acid sequence of cyt c.

A protein can be assessed comprehensively by comparing its structure, function and sequence. The similarities or differences in these three criteria are inter-linked. Hence, the analysis involves comparing the above factors and thereby determining the phylogeny of certain species. Homologs are proteins that share a common ancestry. They are coded for by genes that have been derived from a shared ancestry. A subdivision of homologous proteins is called Paralogs. They are genes related through gene duplication. Another subdivision of homologs are Orthologs. Orthologs are essentially the “same” proteins in different species, synthesized due to passing of “same” gene from a common ancestor when new species diverge from their shared ancestor. [Wilson et al, 1977] More closely related organisms will have more similar cyt c because they have more recent common ancestor. The more recent the common ancestor, the less time for DNA mutations to occur. Changes in DNA lead to changes in mRNA leads to differences in protein amino acid sequences.
MATERIAL AND METHODS

Databases used:
1. Primary nucleotide sequence databases: EMBL (European Molecular Biology Laboratory) [Stoesser G et al, 2002; embl.org], GenBank (maintained by the National Centre for Biotechnology Information) [Benson DA et al, 2005; ncbi.nlm.nih.gov/genbank], and DDBJ (DNA Data Bank of Japan) [Tateno Y et al, 2002; ddbj.nig.ac.jp] have been used.
2. Protein sequence databases: SWISS-PROT [Bairoch and Apweiler, 2000; ebi.ac.uk/swissprot] and its computer-annotated supplement, TrEMBL, as well as the Protein Information Resource (PIR) [Wu CH et al, 2003; pir.georgetown.edu] have been used.
3. Sequence motif databases: Pfam [Punta M et al, 2012; pfam.xfam.org] and PROSITE [Hulo N et al, 2006; prosite.expasy.org] are databases of protein families and domains that have been used.
4. Macromolecular 3-D structure databases: Protein Data Bank (PDB) [Parasraman S, 2012; rcsb.org] is the primary database for 3D structures that has been used along with SCOP (Structural Classification of Proteins) which classifies protein 3D structures in a hierarchical scheme of structural classes.

Tools used:
1. Single sequence alignment: BLAST (Basic Local Alignment Search Tool) [Altschul, S.F. et al, 1990; blast.ncbi.nlm.nih.gov] and FASTA [Lipman, D] et al, 1985; ebi.ac.uk/Tools/sss/fasta]. Each has its own algorithm for comparing sequences and measuring similarity. BLAST results are in the form of bit scores and E-values. The bit score gives an indication of how good the alignment is; the E-value is a parameter that describes the number of hits one can “expect” to see by chance when searching a database of a particular size. LALIGN [Huang and Miller, 1991; expasy.org/genomics/sequence_alignment] is a tool that gives the percent identity of a match. LALIGN also shows the actual alignment of the two sequences.
2. Multiple sequence alignment: CLUSTALW [Thompson et al, 1994; clustal.org] is a multiple sequence global alignment tool. Boxshade [ch.embnet.org/software/BOX_form.html] is a tool used to display conserved domains.
3. Homology determining tools: CLUSTALW_DIST [Thompson et al, 1994; clustal.org] quantifies the evolutionary distance between sequence pairs, infers a phylogenetic tree from an alignment, and calculates the evolutionary distances between species based on the differences between the aligned sequences. ConSurf-DB [consurfdb.tau.ac.il] provides evolutionary conservation profiles for proteins of known structure in the PDB. The evolutionary conservation of each amino acid position in the alignment is calculated.
4. Structure viewing tool: Sirius has been used to visualise a protein or nucleic acid structure in three dimensions [sirius.sdsc.edu].

All tools used are from the Next Generation Biology Workbench. [Subramaniam S, 1998]

RESULTS

1. Structural Comparison
Domains are distinct functional and/or structural units of a protein. Domains are often identified as recurring (sequence or structure) units. In molecular evolution such domains may have been utilized as building blocks, and may have been recombined in different arrangements to modulate protein function. Conserved domains are defined as recurring units in molecular evolution, the extents of which can be determined by sequence and structure analysis. Conserved domains contain conserved sequence patterns or motifs, which allow for their detection in polypeptide sequences. The comparative analysis of the structures of related proteins can reveal the effects of the amino acid sequence changes that have occurred during evolution. Previous work on individual protein families has shown that mutations, insertions and deletions produce changes in three-dimensional structure.
Database: PDB
Tool: Sirius

Structural Analysis:
Database: PDB
Tool: ConSurf-DB
Input: The PDB id of cytochrome C (3cyt) was entered as a query in ConSurf-DB. This tool analyses the structure of the query protein in all the species in PDB.
Output: The result is the structure of cytochrome C with variable and conserved regions clearly marked.

Multiple cytochrome C sequences are lined up against human cytochrome C to determine which residues are conserved across all species. These conserved residues are then located on the human cytochrome C structure.
2. **COMPARING FUNCTION**

The function of a protein changes if the temperature of highest efficiency for the protein is shifted one way or the other, or if it has enzymatic function, that a change in the activation energy needed for a reaction changes, or other such changes occur in how well or under what conditions the protein works best. Changes in function tend to be either selected for or selected against in each species, so one can expect that such changes either spread to become commonplace or are eliminated from the population of that species. Genes and their encoded proteins that perform the same metabolic functions in different organisms are similar but in most cases not identical with respect to their DNA and protein sequences. For example, the cyt c protein that performs the same electron transport function in horse and cow mitochondria is very similar, but not identical.

The results show that the regions involved in the functioning of the protein are highly conserved while the relatively unimportant sites are variable. Thus, the evolutionary profile obtained makes sense in view of the protein function: Highly conserved residues delineate the Heme binding site. This is in correlation with the results of previous studies [Fitch WM, Margoliash E, 1967].
Output:

3. **SEQUENCE SIMILARITY**

DNA is the primary means of passing inherited information from generation to generation. Because proteins are constructed from the information in DNA, they also are reliable indicators of inherited information. The amino acid sequences of proteins are often used to establish the phylogenetic relationships of species. The variance of cytochrome C of different organisms is measured in the number of differing amino acids, each differing amino acid being a result of a base pair substitution, a mutation. If each differing amino acid is assumed to be the result of one base pair substitution, it can be calculated how long ago the two species diverged by multiplying the number of base pair substitutions by the estimated time it takes for a substituted base pair of the cytochrome C gene to be successfully passed on.

**I. SINGLE SEQUENCE ALIGNMENT**

Database: SWISSPROT.
Tool: BLASTP

Input: Here, only one query value was filled – with the accession number of human cytochrome C. Hence, the amino acid sequence of human cytochrome C is compared with the amino acid sequences of cytochrome C of all the organisms on the database.

Output: The query resulted in 181 hits, implying that there are 181 organisms in this database with cytochrome C protein. Here is the result of a select few:

Table 1: Sequences producing significant alignments

<table>
<thead>
<tr>
<th>ACCESSION NUMBER</th>
<th>COMMON NAME</th>
<th>BIT SCORE</th>
<th>E-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8MY23</td>
<td>Human</td>
<td>212</td>
<td>5e-55</td>
</tr>
<tr>
<td>P00008</td>
<td>Rabbit</td>
<td>180</td>
<td>2e-45</td>
</tr>
<tr>
<td>Q6IQM2</td>
<td>Zebrafish</td>
<td>167</td>
<td>1e-41</td>
</tr>
<tr>
<td>P00038</td>
<td>Honeybee</td>
<td>151</td>
<td>9e-37</td>
</tr>
<tr>
<td>P00076</td>
<td>Euglena</td>
<td>100</td>
<td>3e-21</td>
</tr>
</tbody>
</table>

In this result, the bit score of rabbit cytochrome C is the highest (after human cytochrome C itself) while the lowest bit score belongs to euglena. This implies that human cytochrome C matches the best with rabbit cytochrome C, while human and euglena cytochrome C are the most different. Here, the lowest E-value is that of zebrafish while the highest E-value belongs to honeybee. Hence, the bit score of zebrafish cytochrome C is statistically the most significant while the bit score of honeybee cytochrome C has a high probability of occurring by chance.
II. PAIRWISE SEQUENCE ALIGNMENT

Table 2: Pairwise alignments of some species

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Chimpanzee</th>
<th>Horse</th>
<th>Donkey</th>
<th>Mouse</th>
<th>Lamprey</th>
<th>Maize</th>
<th>Euglena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>--</td>
<td>100</td>
<td>88.5</td>
<td>91.3</td>
<td>80.8</td>
<td>66.7</td>
<td>56.6</td>
<td></td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>100</td>
<td>--</td>
<td>88.5</td>
<td>91.3</td>
<td>80.8</td>
<td>66.7</td>
<td>56.6</td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td>88.5</td>
<td>88.5</td>
<td>--</td>
<td>99.0</td>
<td>84.6</td>
<td>63.7</td>
<td>58.6</td>
<td></td>
</tr>
<tr>
<td>Donkey</td>
<td>89.4</td>
<td>89.4</td>
<td>99.0</td>
<td>--</td>
<td>84.6</td>
<td>64.7</td>
<td>58.6</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>91.3</td>
<td>100</td>
<td>94.2</td>
<td>95.2</td>
<td>85.6</td>
<td>66.7</td>
<td>56.6</td>
<td></td>
</tr>
<tr>
<td>Lamprey</td>
<td>80.8</td>
<td>80.8</td>
<td>100</td>
<td>--</td>
<td>59.2</td>
<td>55.6</td>
<td>51.5</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>66.7</td>
<td>66.7</td>
<td>63.7</td>
<td>64.7</td>
<td>59.2</td>
<td>--</td>
<td>51.5</td>
<td></td>
</tr>
<tr>
<td>Euglena</td>
<td>56.6</td>
<td>56.6</td>
<td>58.6</td>
<td>58.6</td>
<td>55.6</td>
<td>51.5</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

In pairwise alignments of cyt c sequences of organisms with varied bit scores in the single sequence alignment, some had a high bit score (chimpanzee), while others had a low bit score (euglena).

Database: NBRF
Tool: BLAST
Input: To compare only two sequences, both query boxes must be filled with the accession numbers of the sequences. Hence, the amino acid sequence of the first query was aligned with that of the second query only.

Identities are the number and fraction of identical residues in the sequence. The more identities there are between two molecules, the more recently they have evolved from a common ancestral molecule and thus the closer the kinship of their owners. Thus the cytochrome c of the chimpanzee is identical to that of humans except for one amino acid, whereas yeast cytochrome c differs from that of humans at 44 positions. In this case, 92% of the amino acids (leucine and isoleucine) of human and mouse cytochrome C are similar.

The cladogram drawn from cytochrome C analysis matches the taxonomical data. It was noted that the chimpanzee and human sequences taken from SWISSPROT are 97% identical, while the same sequences are 100% identical when the database is NBRF. This discrepancy is due to the difference in the algorithms of the databases. The distance of a branch indicates the number of changes that have taken place along the branch.

III. MULTIPLE SEQUENCE ALIGNMENT

In multiple sequence alignment, all similar sequences can be compared in one single figure or table. The basic idea is that the sequences are aligned on top of each other in a common coordinate system. Cytochrome C sequences of human, chimpanzee, dog and horse were compared.

Data from: UniProtKB
Search by: FASTA

Tool used for aligning sequences: CLUSTALW
Input: The sequences of the organisms were saved in FASTA format. Entries that said “partial sequences” in their description were avoided. No sequence was selected twice. CLUSTALW tool was selected to run the task.

Output:
Sequence 1: Canis_familiaris 105 aa [P00011 CYC_CANLF]
Sequence 2: Equus_burchelli 105 aa [P00004 CYC_HORSE]
Sequence 3: Pan_troglodytes 105 aa [P99998 CYC_PANTR]
Sequence 4: Homo_sapiens 105 aa [P99999 CYC_HUMAN]

Start of Pairwise alignments
Sequences (1:2) Aligned. Score: 95
Sequences (1:3) Aligned. Score: 89
Sequences (1:4) Aligned. Score: 89
Sequences (2:3) Aligned. Score: 89
Sequences (2:4) Aligned. Score: 89
Sequences (3:4) Aligned. Score: 100
Phylogenetic Relationships and Homology of Cytochrome C using Bioinformatics

Checking for conserved domains

Tool used: Boxshad

The portions of the sequence that are conserved between all species are highlighted in black. Amino acids that are similar, but not identical are shown with a grey background. Those that have a different character are shown with a white background. Most of the sequence of cytochrome C is conserved across these four species. There are also many amino acids which are identical, if not similar. The non-conserved regions are mostly found in the dog and horse cytochrome C sequences. Hence, human and chimpanzee have the maximum conserved domains.

Determining the distance between sequence pairs

Tool used: CLUSTALW_DIST

Output:

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan_troglodytes</td>
<td>MGDVEK GKKI F IMKGS QC HTV EXGKKHT GPN LHGLF GRTK GQA PG PSYTAAN KN KGI TWD</td>
</tr>
<tr>
<td>Homo_sapiens</td>
<td>MGDVEK GKKI F IMKGS QC HTV EXGKKHT GPN LHGLF GRTK GQA PG PSYTAAN KN KGI TWD</td>
</tr>
<tr>
<td>Canis_familiaris</td>
<td>MGDVEK GKKI F VQKGA QC HTV EXGKKHT GPN LHGLF GRTK GQA PG PSYTAAN KN KGI TWD</td>
</tr>
<tr>
<td>Equus_burchelli</td>
<td>MGDVEK GKKI F VQKGA QC HTV EXGKKHT GPN LHGLF GRTK GQA PG PSYTAAN KN KGI TWD</td>
</tr>
</tbody>
</table>

Cladogram 2: Distance between sequence pairs

Black and white dashed lines indicate that the sequences are identical. [Eernisse D], 1998] Hence, dog and horse are more closely related to each other than to humans or chimpanzees. Meanwhile, chimpanzees and humans share their cytochrome C sequence, and are very closely related. This implies that the cytochrome C sequence of an organism is similar to another species belonging to the same order.

IV. CLADOGRAM OF CYTOCHROME C SEQUENCES OF SOME SPECIES

Multiple sequence alignments were done between different species across the three kingdoms – Plantae, Fungi and Animalia.

Database: UniProt

Tool: CLUSTALW_DIST

Input: The amino acid sequences from Uniprot were saved in FASTA format and aligned using CLUSTALW.

The alignment data was used to run the task using CLUSTALW_DIST tool.

Output: CLUSTALW_DIST program produced a phylogenetic tree inferred from this data.

The black and white dotted line implies 100% identity. Hence, the cytochrome C residues are completely conserved in human and chimpanzee. Comparing cytochrome C among the different species revealed that only 27 amino acid residues are identical. More than 60 residues differ among them (this number is not exact because there are slight differences in the length of the molecule in some species). However, the degree of similarity among amino acid sequences in cytochrome C corresponds closely to the phylogenetic relationships. [8] That is, mammalian sequences are more similar to each other than to any reptilian sequence and vice versa.
V. DISTANCE MATRIX

A distance matrix is a table that shows all the pairwise comparisons between species. Database: UniProtKB  
Tool: LALIGN (FASTA)  
Input: Selection of organisms was based on their bit scores in the single sequence alignment. Some organisms selected had a high bit score (example, chimpanzee), while others had a low bit score (example, yeast). This was done to get a highly branched cladogram, with as many clades as possible. Output:

Overall, it was found that the cytochrome C sequences of most species compared had little no change in their biochemical properties. There are also quite a lot of amino acids having similar properties and charge, indicating that the change in amino acid is conservative. A pairwise score is calculated for every pair of sequences that are to be aligned. Pairwise scores are calculated as the number of identities in the best alignment divided by the number of residues compared (gap positions are excluded). Pairwise scores are initially calculated as percent identity scores and are converted to distances by dividing by 100 and subtracting from 1.0 to give number of differences per site. Hence, the cladogram is constructed by the algorithm.

The cladogram below is constructed using the pairwise alignments of table 3.

The tree corresponds quite well to the evolutionary relationships among the species. But there are some anomalies. It indicates, for example, that the bat is more closely related to birds than mammals. This is certainly wrong [Tsagkogeorga, 2013]. But sequence analysis of other proteins can resolve such discrepancies.

VI. COMPARISON OF CYTOCHROME C FAMILY

Cytochromes c are electron-transfer proteins having one or several heme C groups, bound to the protein by thioether bonds involving sulphydryl groups of cysteine residues. The fifth heme iron ligand is always
### Table 3: Approximate percent identity in the overlap of 105 amino acids for different species

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Rhesus Monkey</th>
<th>Dog</th>
<th>Horse</th>
<th>Donkey</th>
<th>Pig</th>
<th>Rabbit</th>
<th>Duck</th>
<th>Pigeon</th>
<th>Chicken</th>
<th>Bat</th>
<th>Turtle</th>
<th>Tuna</th>
<th>Fruit fly</th>
<th>Moth</th>
<th>Yeast</th>
<th>Candida</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus monkey</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>87</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td>83</td>
<td>84</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donkey</td>
<td>81</td>
<td>83</td>
<td>88</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>87</td>
<td>88</td>
<td>96</td>
<td>95</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>88</td>
<td>89</td>
<td>94</td>
<td>89</td>
<td>91</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duck</td>
<td>83</td>
<td>84</td>
<td>88</td>
<td>84</td>
<td>83</td>
<td>87</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pigeon</td>
<td>84</td>
<td>85</td>
<td>88</td>
<td>84</td>
<td>85</td>
<td>87</td>
<td>92</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicken</td>
<td>82</td>
<td>83</td>
<td>86</td>
<td>84</td>
<td>82</td>
<td>87</td>
<td>89</td>
<td>97</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bat</td>
<td>82</td>
<td>83</td>
<td>86</td>
<td>83</td>
<td>84</td>
<td>88</td>
<td>86</td>
<td>92</td>
<td>95</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turtle</td>
<td>81</td>
<td>82</td>
<td>87</td>
<td>84</td>
<td>81</td>
<td>87</td>
<td>89</td>
<td>93</td>
<td>92</td>
<td>92</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuna</td>
<td>69</td>
<td>68</td>
<td>71</td>
<td>73</td>
<td>77</td>
<td>85</td>
<td>84</td>
<td>74</td>
<td>73</td>
<td>74</td>
<td>76</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit fly</td>
<td>67</td>
<td>68</td>
<td>76</td>
<td>76</td>
<td>72</td>
<td>74</td>
<td>77</td>
<td>74</td>
<td>75</td>
<td>74</td>
<td>68</td>
<td>74</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moth</td>
<td>64</td>
<td>65</td>
<td>72</td>
<td>67</td>
<td>66</td>
<td>69</td>
<td>71</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>71</td>
<td>67</td>
<td>59</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast</td>
<td>44</td>
<td>43</td>
<td>39</td>
<td>40</td>
<td>38</td>
<td>41</td>
<td>41</td>
<td>42</td>
<td>38</td>
<td>38</td>
<td>41</td>
<td>36</td>
<td>34</td>
<td>37</td>
<td>36</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Candida</td>
<td>34</td>
<td>35</td>
<td>34</td>
<td>32</td>
<td>31</td>
<td>33</td>
<td>33</td>
<td>32</td>
<td>34</td>
<td>34</td>
<td>30</td>
<td>33</td>
<td>31</td>
<td>35</td>
<td>35</td>
<td>39</td>
<td>59</td>
</tr>
</tbody>
</table>
provided by a histidine residue. Cytochromes c possess a wide range of properties and function in a large number of different redox processes. This family has four classes - I to IV. Class I includes the low-spin soluble cyt c of mitochondria and bacteria, with the haem-attachment site towards the N-terminus, and the sixth ligand provided by a methionine residue about 40 residues further on towards the C-terminus. On the basis of sequence similarity, class I cyt c were further subdivided into five classes, IA to IE. [Margoliash E, 1963]

Database: Pfam
Tool: CLUSTAL W and CLUSTAL_DIST
Input: Members of the cytochrome C family were compared in order to assess their relationship.

The Cladogram 5 shows that Cytochrome C and c1 have maximum number of same residues; Cytochrome b and b245 have nearly the same amino acid sequence. Hence, Cytochrome C and c1 belong to the same class while Cytochrome b and b245 belong to the same class.

**CONCLUSION**

Cytochrome C is a protein that is evolutionary conserved across a wide spectrum of species. We have taken advantage of this property of the molecule to establish homology and phylogenetic relationships between species. In this paper, a cluster of homologs was identified, a multiple sequence alignment was done and a phylogenetic tree was constructed. The amino acid sequences of cytochrome C were used to establish the phylogenetic relationships of species.

As cytochrome C is ubiquitous, it ensures that the comparisons are independent of the overall species phenotype. [Brown TA, 2002] The phylogeny inferred from the sequence alignment of cytochrome C was confirmed by comparing its structure and function. The protein was found to be functionally redundant i.e., many dissimilar cytochrome C sequences form the same general structure and perform the same general biological role. However, functional redundancy need not be exact in terms of performance; some functional cytochrome C sequences may be slightly better at electron transport than others. Hence, from the analysis cytochrome C, it appears that two different organisms which have the same or even similar protein sequences are genealogically related.

From the phylogenetic trees, it can be predicted that human and chimpanzee cytochrome C sequences should be much more similar than, say, human and yeast cytochrome C - simply due to inheritance. The homologous similarities of cytochrome C are especially suggestive of common ancestry. The high degree of functional redundancy of the cytochrome C molecule and the fact that the phylogenies derived from them generally match other phylogenies very well indicate that cytochrome C is molecular evidence of evolution. These results match those of Fitch and Margoliash [Fitch WM, Margoliash E, 1967]. The results also are in accordance with the theory of common descent, proposed by Charles Darwin, which says that a group of organisms have common descent if they have a common ancestor [Darwin C, 1859].

Cytochrome C is a protein that is evolutionary conserved across a wide spectrum of species. We have taken advantage of this property of the molecule to establish homology and phylogenetic relationships between species. In this paper, a cluster of homologs was identified, a multiple sequence alignment was done and a phylogenetic tree was constructed. The amino acid sequences of cytochrome C were used to establish the phylogenetic relationships of species.

As cytochrome C is ubiquitous, it ensures that the comparisons are independent of the overall species phenotype. [Brown TA, 2002] The phylogeny inferred from the sequence alignment of cytochrome C was confirmed by comparing its structure and function. The protein was found to be functionally redundant i.e., many dissimilar cytochrome C sequences form the same general structure and perform the same general biological role. However, functional redundancy need not be exact in terms of performance; some functional cytochrome C sequences may be slightly better at electron transport than others. Hence, from the analysis cytochrome C, it appears that two different organisms which have the same or even similar protein sequences are genealogically related.

From the phylogenetic trees, it can be predicted that human and chimpanzee cytochrome C sequences should be much more similar than, say, human and yeast cytochrome C - simply due to inheritance. The homologous similarities of cytochrome C are especially suggestive of common ancestry. The high degree of functional redundancy of the cytochrome C molecule and the fact that the phylogenies derived from them generally match other phylogenies very well indicate that cytochrome C is molecular evidence of evolution. These results match those of Fitch and Margoliash [Fitch WM, Margoliash E, 1967]. The results also are in accordance with the theory of common descent, proposed by Charles Darwin, which says that a group of
organisms have common descent if they have a common ancestor [Darwin C, 1859].

REFERENCES

11. Lipman, DJ; Pearson, WR. "Rapid and sensitive protein similarity searches". Science, 1985; 227 (4693): 1435–41
25. UniProtKB - P99999 (CYC_HUMAN)

Web References

[1] embl.org
[3] ddbj.nig.ac.jp
[4] ebi.ac.uk/swissprot
[5] pigr.georgetown.edu
[8] rcsb.org
[10] ebi.ac.uk/Tools/sss/fasta
[12] clustal.org
[14] consurfdb.tau.ac.il
[15] sirius.sdsc.edu

© 2015| Published by IRJSE