**THEORETICAL ANALYSIS OF MUTATION PATTERN OF THE CYTOCHROME P450 SUPERFAMILY.**

MATUSHKIN Yu.G.+, MOROZOVA I.N., MOROZOV P.S.

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 10 Lavrentiev Ave., Novosibirsk, 630090, Russia;
e-mail: mat@bionet.nsc.ru;

+Corresponding author

Keywords: nucleotide sequence, cytochrome p450, evolution, mutational analysis, regression analysis

Species-specificity of mutational process in the cytochrome P450 superfamily (CYP2 family of the human, mouse, and rat) was demonstrated. Significant distinction between the mutational patterns of different families (CYP2 and CYP11) within the same species (rat) was also shown, indicating a marked specificity of mutational process in different families of the cytochrome genes. Analysis of three mutational patterns of the CYP21 family-evolutionary pattern, patterns of pseudogenes, and damaged (of congenital adrenal hyperplasia patients) alleles-refutes the widely accepted hypothesis on gene conversion-assisted transfer of the mutations that originated in pseudogenes into normal alleles of CYP21 gene.

**1. Introduction**

Cytochromes p450 is a large superfamily of hem-containing membrane-binding proteins with the molecular weight of 50,000 (400-530 amino acids). Currently, the superfamily comprises over 500 members, belonging to 85 eukaryotic and 20 prokaryotic species together with about 30 pseudogenes [8]. p450 are involved in oxidative metabolism of a wide range of exo- and endogenous substances (totally, over 80 substances) in bacteria, fungi, plants, and animals [1]. The superfamily includes microsomal enzymes and the enzymes that are active in mitochondria; they are coded for by nuclear p450 genes, located on different chromosomes. About 100 p450 families have been so far defined; 14 of them belong to mammals [8]. Homology of individual p450 families is rather low (10-30%), and the structure-function relation has not yet been clarified sufficiently. The evolution of p450 occurred through gene duplications, gene fusions, gene conversions, and intron losses [2].

Certain cytochrome families (11, 17, and 21) are involved in cortisol and aldosterone biosyntheses; disturbance of these processes combined with accumulation of androgens [6] results in congenital adrenal hyperplasia (CAH), which is one of the most frequently occurring human diseases and is transmitted by autosomal recessive inheritance [5].

Four enzymes of the five necessary for cortisol biosynthesis from cholesterol are cytochromes (families 11, 17, and 21). Their genes are located on different chromosomes. CYP11A and CYP11B are the enzymes active in mitochondria; 17 and 21, in microsomes. Over 90% cases of congenital adrenal hyperplasia is the result of the CYP21 (EC 1.14.99.10) deficiency of various degrees [9]; the rest cases are caused by the damage of CYP11 function.

CYP21 gene was described in 1984-1986. In humans, two copies of the gene occur: the active gene (CYP21 or CYP21B) and the inactive pseudogene (CYP21P or CYP21A). Homology of these copies is 98%; they are located on chromosome 6 and form a tandem repeat [4]. The pseudogene displays several small deletions and amino acid substitutions, resulting in its inactivation. About 20% of CYP21 deficiency cases are connected with deletion of the entire gene or its essential part. The rest of mutant alleles contain small deletions or insertions [9].

**2. Materials and Methods**

A set of genes of the p450 cytochrome superfamily was created using the EMBL databank, the Nelson's database on cytochromes (<http://drnelson.utmem.edu/nelsonhomepage.html>), and relevant publications. The software package VOSTORG [10] was used to carry out the multiple alignment of these sequences and their phylogenetic analysis applying bootstrap test to estimate the reliability of the tree topology obtained. The software package AMS [7] was used to study the evolutionary mutational patterns in the subtrees that were confirmed reliable by bootstrap test.

The mutational pattern is considered to mean the list of all mutation types with indication of their frequencies for each position of the aligned sequence set along with the description of the expected consequences of these mutations. Mutational pattern can be estimated at different stages of the evolutionary process. First, it is the pattern of spontaneous mutations. We can estimate it through analysis of pseudogenes or direct experimental procedures (which is sometimes difficult). Second, it is the pattern of the mutations yet unfixed, which also contribute to the overall population polymorphism of the locus under study. Third, it is the pattern revealed phylogenetically and tested by selection, which we can reveal through analysis of actual genes of different species. And finally, it is the somatic mutational pattern, revealed experimentally (in this work, in primary sequences of genes of the CAH patients).

Similarity of evolutionary mutational patterns of the p450 cytochrome superfamily was studied in the sets of 23 sequences of CYP1 family; 18, CYP11; 41, CYP2; 10, CYP6; 7, CYP7; 5, CYP21; 34, prion proteins; and 55, globins.

Sequences of human, mouse, and rat C2 (9, 9, and 16, respectively) were used to estimate the species-specificity of the evolutionary mutational patterns.

Specificity of the evolutionary mutational patterns relative to the families of the p450 cytochromes was studied using 4 rat C11 sequences, 16 rat C2 sequences, and 7 C7 sequences (rat, hamster, mouse, rabbit, and human).

Mutational patterns of cytochrome CYP21 was studied using five CYP21 sequences, seven human mutant alleles of CAH patients, and three sequences of pseudogenes CYP21 (two human and one mouse).

**3. Results and Discussion**

Multifactorial analysis has demonstrated that mutational patterns of CYP1, CYP2, CYP11, and CYP6 families are similar, whereas the patterns of CYP7 and CYP21 differ from the patterns these four families. On the whole, their patterns are similar to those of globins and differ significantly from the pattern of prion proteins.

Correlation coefficients of the mutational patterns of the CYP450 protein family

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | С1 | С2 | С11 | С6 | С7 | Globins | Prions |
| C1 | 1.00 | 0.93 | 0.91 | 0.93 | 0.70 | 0.84 | 0.49 |
| C2 | 0.93 | 1.00 | 0.84 | 0.94 | 0.56 | 0.86 | 0.42 |
| C11 | 0.91 | 0.84 | 1.00 | 0.85 | 0.79 | 0.78 | 0.55 |
| C6 | 0.93 | 0.94 | 0.85 | 1.00 | 0.62 | 0.84 | 0.41 |
| C7 | 0.70 | 0.56 | 0.79 | 0.62 | 1.00 | 0.55 | 0.51 |
| Globins | 0.84 | 0.86 | 0.78 | 0.84 | 0.55 | 1.00 | 0.43 |
| Prions | 0.49 | 0.42 | 0.55 | 0.41 | 0.51 | 0.43 | 1.00 |

All correlations are significant at p < 0.05, N=380 (Casewise deletion of missing data).

Mutational patterns of CYP2 family were built for three different species: human, rat, and mouse. Regression analysis has demonstrated that, despite the overall similarity of these patterns, they display significant species-specific differences.

Linear regression coefficient (standard error)

|  |  |  |  |
| --- | --- | --- | --- |
| CYP2 (p450) | Human | Mouse | Rat |
| Human | 1.0 | 0.942593 (0.017176) | 0.877328 (0.024683) |
| Mouse | 0.942593 (0.017176) | 1.0 | 0.907174 (0.021641) |
| Rat | 0.877328 (0.024683) | 0.907174 (0.021641) | 1.0 |

It was shown that the similarity of evolutionary mutational patterns of different families of one species (rat CYP2 and CYP11) is lower than the similarity of different families belonging to different species (rat CYP11 and CYP7). This indicates a specificity of the mutational process in different families of the p450 cytochrome genes within one species.

Linear regression coefficient (standard error)

|  |  |  |  |
| --- | --- | --- | --- |
|   | CYP2\_rat | CYP11\_rat | CYP7 |
| CYP2\_rat | 1.0 | 0.629808 (0.039952) | 0.659338 (0.038671) |
| CYP11\_rat | 0.629808 (0.039952) | 1.0 | 0.872651 (0.025118) |
| CYP7 | 0.659338 (0.038671) | 0.872651 (0.025118) | 1.0 |

It was demonstrated that the evolutionary mutation pattern of CYP21 gene (damages of this gene cause adrenal hyperplasia in humans) is under a strong negative selection pressure (that is, the mutations accumulated are clustered in the region with the peak around 0.7: the distance between amino acids according to Miyata matrix, Fig. 1). In CYP21 pseudogenes, the majority of mutations accumulated change drastically the physico-chemical properties of the amino acid in the corresponding position (the peak of the distribution is at 2.8, Fig. 3). The accumulated mutations found in the damaged (for example, by frameshift) alleles of the normal CYP21 gene, obtained from adrenal hyperplasia patients, are distributed according to the evolutionary pattern, indicating the occurrence of stabilizing selection pressure (Fig. 2). The results obtained refute the widely accepted hypothesis on gene conversion-assisted transfer of the mutations that originated in pseudogenes to the normal alleles of CYP21 gene3, since the mutational patterns of abnormal alleles and pseudogenes are essentially different, as is evident from Figures.



Figure 1. Normal CYP21 genes of P450 cytochromes (5 species). Dependence of the number of substitutions on the physico-chemical distance between amino acids before and after the substitution. Abscissa, physico-chemical distance between amino acids (according to Miyata matrix); ordinate, % substitutions.



Figure 2. Abnormal C21 alleles of human P450 cytochromes. Dependence of the number of substitutions on the physico-chemical distance between amino acids before and after the substitution. The axis designations are as in Fig. 1.

****

Figure .3. Pseudogenes C21 of human P450 cytochromes. Dependence of the number of substitutions on the physico-chemical distance between amino acids before and after the substitution. The axis designations are as in Fig. 1.

**Acknowledgements**

This work was supported by the Russian Programs on Human Genome and Priority Directions in Genetics.

**References**

1. Coon M.J., Ding X., Pernecky S.J., Vaz A.D.N.(1992) FASEB J. 6, 669-673.
2. Degtyarenko K.N., Archakov A.I. (1993) FEBS Lett. 332:1-8.
3. Helmberg A, Tusie-Luna MT, Tabarelli M, Kofler R, White,P.C., Mol.Endocrinol. (1992) 6,1318-1322.
4. Higashi,Y.,Yoshioka,H.,Yamane,M.,Gotoh,O., and Fujii-Kuriyama, Y. (1986). Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: A pseudogene and a genuine gene. Proc.Natl.Sci.USA 83,2841-2845.
5. Kuttenn,F., Couillin,P., Girard,F., Billaud,L., Vincens,M., Boucekkine,C., Thalabard,J.-C., Maudelonde,T., Spritzer,P., Mowszowicz,I., Boue,A., & Mauvais-Jarvis,P. (1985) N.Engl.J.Med. 313,224-231.
6. Miller,W.L.&Morel,Y. (1989) Annu.Rev.Genet. 23,371-393.
7. Morozov P.S., Matushkin Yu.G. (1996) AMS: a software package for inferring and analysing of mutational spectra.; Proceedings of the International Conference "Intelligent Systems for Molecular Biology'96", Washington University, St.Louis, Missouri, USA
8. Nelson,D.R., Koymans,L., Kamataki,T., Stegeman,J.J., Feyereisen,R., Waxman,D.J., Waterman,M.R., Gotoh,O., Coon,M.J., Estabrook,R.W., Gunsalus,I.C., Nebert,D.W. (1996) Pharmacogenetics 6,1-42.
9. White,P.C.,Tusie-Luna,M.T.,New,M.I., and Speiser,P.W. (1994). Mutation in steriod 21-hydroxylase (CYP21). Hum.Mutat. 3,373-378.
10. A.A. Zharkikh, A.Yu. Rzhetsky, P.S. Morosov, T.L. Sitnikova, J.S. Krushkal. (1991), VOSTORG: a package of microcomputer programs for sequence analysis and construction of phylogenetic trees. Gene, 101: 251-254.