Genomics 103 (2014) 40–47

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

# Purifying selection against gene conversions in the folate receptor genes of primates

Nicholas Petronella, Guy Drouin *

*Département de biologie et Centre de recherche avancée en génomique environnementale, Université d'Ottawa, Ottawa, Ontario K1N 6N5, Canada*

ARTICLE INFO

*Article history:*
Received 19 June 2013
Accepted 22 October 2013
Available online 30 October 2013

*Keywords:*
Folate receptor genes
*FOLR1*
*FOLR2*
*FOLR3*
Gene conversion
Purifying selection
GC-content

ABSTRACT

We characterized the gene conversions between the human folate receptor (*FOLR*) genes and those of five other primate species. We found 26 gene conversions having an average length of 534 nucleotides. The length of these conversions is correlated with sequence similarity, converted regions have a higher GC-content and the average size of converted regions from a functional donor to another functional donor is significantly smaller than the average size from a functional donor to a pseudogene. Furthermore, the few conversions observed in the *FOLR1* and *FOLR2* genes did not change any amino acids in their coding regions and did not affect their promoter regions. In contrast, the promoter and coding regions of the *FOLR3* gene are frequently converted and these conversions changed many amino acids in marmoset. These results suggest that purifying selection is limiting the functional impact that frequent gene conversions have on functional folate receptor genes.



## 1. Introduction

The human folate receptor gene family is located at chromosome 11q13–14. Folate receptors participate in the binding and transport of folate (a variety of vitamin B9) and the naturally occurring form of folic acid [1,2]. Folic acid derivatives have been found to be critical for purine synthesis, cell division, tissue growth and DNA methylation [3].

In humans, the folate receptor gene family consists of three functional genes *FOLR1*, *FOLR2* and *FOLR3* in addition to two nonfunctional pseudogenes *FOLR1P1* and *FOLR3P1* [4,5]. *FOLR1* is the most widely expressed folate receptor gene in adult tissues and produces a variety of transcripts through simple or complex alternative splicing [6]. *FOLR2* is mainly expressed in placental tissue while *FOLR3* is a secretory molecule expressed predominantly in hematopoietic cells [5,7]. Of the three functional *FOLR* genes, *FOLR1* is likely the most important because mutations in this gene have been shown to result in neural tube malformations and brain-specific folate transport deficiencies leading to neurodegeneration early in childhood [8,9]. Furthermore, mice gene knockout studies have shown that *FOLR1* is essential for neural tube development whereas knockouts of *FOLR2* have a normal phenotype [6]. There is still little information on the role of the FOLR3 receptor [10].

Much is known about the expression and organization of these folate receptor genes due to their potential affiliation with disease, specifically neural tube defects (NTDs). Embryonic neural tubes form the brain and spinal cord during the development of the embryo. If the neural tubes fail to close properly, the results are congenital malformations known as neural tube defects. Though well studied, causes for NTDs still remain largely unknown. Many studies, however, highlight the little understood protective effects of folic acid supplementation during pregnancy [3,4,10,11]. Due to this phenomenon, the folate receptor genes have been recognized as a means to better understand NTDs and they have been the focus of numerous studies (e.g., ref. [2,6,8]).

Gene conversions are a type of homologous recombination that involves the unidirectional movement of genetic material from a donor sequence to an acceptor sequence. Conversion events have been found to be connected to a variety of human inherited diseases, especially when mutations that accumulate in nonfunctional pseudogenes overwrite essential components of a highly similar functional gene [12,13]. Given that FOLR genes are located near one another on human chromosome 11, that they encode proteins having relatively high levels of sequence identity and that small gene conversion events have been shown to introduce deleterious mutations in some patients, it is therefore of interest to study the gene conversions occurring between FOLR genes as well as between them and their pseudogenes [1,6,8].

The recent surge in available sequence data has facilitated the opportunity to further our understanding of the folate receptor gene family in humans as well as other primate species. Here, a comparative genomics study was performed on the effects of gene conversion in the folate receptor gene family of *Homo sapiens* (human), *Pan troglodytes*

* Corresponding author at: Département de biologie, Université d'Ottawa, 30 Marie Curie, Ottawa, Ontario K1N 6N5, Canada. Fax: +1 613 562 5486.
*E-mail address:* gdrouin@science.uottawa.ca (G. Drouin).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.10.004

(chimpanzee), *Callithrix jacchus* (marmoset), *Nomascus leucogenys* (gibbon), *Pongo abelii* (orangutan), and *Macaca mulatta* (rhesus monkey). In particular, we wanted to determine whether gene conversions were less frequent between functionally more important genes or gene regions. Since one would expect functionally important genes to be evolving under stronger purifying selection, one would expect that the impact of gene conversions would be lesser between functionally more important genes and gene regions. Our results do show that gene conversions are less frequent, do not lead to amino acid changes and do not affect the promoter regions of the functionally more important *FOLR1* and *FOLR2* genes. This suggests that stronger purifying selection is acting on functionally more important genes and gene regions.

## 2. Results

### 2.1. Gene conversions

We did not detect any gene conversion events between the three gibbon *FOLR* genes (results not shown). A total of 26 conversion events, spanning both exons and introns, were detected between the *FOLR* genes of the other five primate species (Table 1). Functional genes are 17 times donor sequences and 12 times acceptor sequences, whereas pseudogenes are 9 times donor sequences and 14 times acceptor sequences (Table 1). Compared with an average of 14.5 times, the functional genes are not more often donor or acceptor ($\chi^2$-test; $p = 0.35$). Similarly, compared with an average of 11.5 times, the pseudogenes are not more often donor or acceptor ($\chi^2$-test; $p = 0.30$). The smallest of these gene conversions is 27 nucleotides long and the largest is 1727 nucleotides long. The average size (±standard error) of these 26 conversions is 534 ± 97 nucleotides. The average size of conversions involving functional genes as acceptors (±standard error) is 482 ± 130 bp whereas those involving pseudogenes as acceptors is 545 ± 146 bp. These length differences are not statistically significant ($t$-test, $p = 0.62$). Interestingly, the average size (±standard error) of converted regions from a functional donor to another functional donor (141 ± 34 bp) is significantly smaller than the average size from a functional donor to a pseudogene (665 ± 157 bp; $t$-test, $p = 0.007$). Fig. 1 shows the distribution of each conversion event along with its approximate length.

Converted regions are more GC-rich than non-converted regions. When considering the 20 conversions larger than 100 nucleotides long, the average GC-content (±standard error) of the converted regions is 58.61% ± 0.89 and is significantly larger than that of the non-converted regions (48.36% ± 0.5, $t$-test, $p = 9.2 \times 10^{-13}$). For this analysis, we only considered conversions larger than 100 bp to minimize the effect of stochastic variation.

### 2.2. Correlation between sequence similarity and length of gene conversions

A correlation test was performed in order to examine the effect of overall sequence similarity on the length of conversion events; similarities between converted sequences are listed in Table 1. Similarities between both the coding and full genomic sequences were tested for their potential effects on conversion lengths. Coding sequences contain only coding sequences (exons) whereas genomic sequences contain exons and introns. Positive correlations are present between sequence similarity and gene conversion length (Spearman rank correlation tests, coding $\rho = 0.61, 0.81, 0.93, 0.87, 0.61$; genomic $\rho = 0.86, 0.77, 0.93, 0.87, 0.65$ for human, chimp, orangutan, rhesus and marmoset, respectively) but only the rhesus correlation is significant (coding, $p = 0.005$; genomic $p = 0.005$). The absence of significant correlations for most species is likely due to the low number of data points. For example, orangutan, human and chimpanzee only have three or four data points (Table 1). Performing correlation analyses on all primate species yields significant correlations for both coding and genomic sequences (Spearman rank

**Table 1**
Gene conversions between the folate receptor genes of five primate species.

| | Length (bp) | Donor | Location (from–to) | Acceptor | Location (from–to) | % coding similarity | % genomic similarity |
|---|---|---|---|---|---|---|---|
| *Human* | | | | | | | |
| 1 | 1092 | *FOLR3P1* | 927 bp before exon1–bp3 intron1 | *FOLR3* | 920 bp before exon1–bp3 intron1 | 96.9 | 95 |
| 2 | 101 | *FOLR3* | bp2815 intron1–bp74 exon2 | *FOLR2* | bp2106 intron1–bp74 exon2 | 88.4 | 79.2 |
| 3 | 36 | *FOLR2* | bp 100–136 of exon 4 | *FOLR1P1* | bp 100–136 of exon 3 | 70.4 | 69.1 |
| 4 | 1727 | *FOLR1* | bp2600 of intron1–540 bp after exon4 | *FOLR1P1* | 307 bp before exon1–379 bp after exon3 | 89.6 | 91.6 |
| *Chimpanzee* | | | | | | | |
| 1 | 908 | *FOLR3* | 751 bp before exon1–bp155 exon1 | *FOLR3P1* | 768 bp before exon1–bp155 exon1 | 95.5 | 92.2 |
| 2 | 274 | *FOLR1* | bp102 intron2–bp222 exon3 | *FOLR1P1* | bp100 intron2–bp87 intron3 | 90.7 | 89 |
| 3 | 78 | *FOLR3P1* | bp111–189 exon4 | *FOLR1P1* | bp105–183 exon4 | 71.7 | 57.8 |
| 4 | 738 | *FOLR1P1* | bp44 exon4–502 bp after exon4 | *FOLR1* | bp43 exon4–556 bp after exon4 | 90.7 | 89 |
| *Orangutan* | | | | | | | |
| 1 | 1116 | *FOLR3P1* | 992 bp before–bp155 | *FOLR3* | 959 bp before exon1–bp155 exon1 | 93.5 | 92.8 |
| 2 | 719 | *FOLR3* | bp87 intron2–123 bp after exon4 | *FOLR3P1* | bp4092–4829 | 93.5 | 92.8 |
| 3 | 49 | *FOLR3P1* | bp4545–4599 | *FOLR1P1* | bp101–149 exon4 | 63.3 | 59.3 |
| *Rhesus* | | | | | | | |
| 1 | 93 | *FOLR2* | bp91–184 exon2 | *FOLR3* | bp3631–3724 | 87.2 | 65.5 |
| 2 | 165 | *FOLR2* | bp2080–2267 intron1 | *FOLR3* | bp3516–3682 | 87.2 | 65.5 |
| 3 | 76 | *FOLR3* | bp3991–4066 | *FOLR1* | bp65 exon3–bp5 intron3 | 77.5 | 58.2 |
| 4 | 740 | *FOLR3* | bp3771–4510 | *FOLR3P1* | bp372 intron4–99 bp after exon6 | 97.9 | 93.9 |
| 5 | 1275 | *FOLR3* | bp459–1734 | *FOLR3P1* | bp527–1797 intron3 | 97.9 | 93.9 |
| 6 | 296 | *FOLR1* | bp39 intron2–bp44 intron3 | *FOLR1P1* | bp38 intron2–bp45 intron3 | 87.5 | 86.7 |
| 7 | 52 | *FOLR1* | bp64–116 exon3 | *FOLR3P1* | bp64–116 exon5 | 76.2 | 55.1 |
| 8 | 1201 | *FOLR3* | 827 bp before–bp372 | *FOLR3P1* | 777 bp before exon1–bp3 intron2 | 97.9 | 93.9 |
| *Marmoset* | | | | | | | |
| 1 | 124 | *FOLR3P1* | bp44–168 exon4 | *FOLR2* | bp92–216 exon5 | 87.4 | 79.5 |
| 2 | 263 | *FOLR2* | bp20 intron3–bp2065 intron2 | *FOLR3* | bp18 intron2–bp3063 intron1 | 89.1 | 82.5 |
| 3 | 99 | *FOLR3* | bp57–156 exon4 | *FOLR1* | bp87–186 exon4 | 75.3 | 69.9 |
| 4 | 354 | *FOLR3P1* | bp63 intron3–bp333 intron2 | *FOLR3* | bp62 intron3–bp43 intron2 | 97.8 | 96.4 |
| 5 | 371 | *FOLR3P1* | bp27 intron2–bp2938 intron1 | *FOLR3* | bp28 intron2–bp2938 intron1 | 97.8 | 96.4 |
| 6 | 657 | *FOLR3* | 230 bp after exon4–bp76 exon3 | *FOLR3P1* | 230 bp after exon4–bp512 intron3 | 97.8 | 96.4 |
| 7 | 1282 | *FOLR3P1* | bp1023 intron1–90 bp before exon1 | *FOLR3* | bp1017 intron1–90 bp before exon1 | 97.8 | 96.4 |