Genomics 105 (2015) 17–22

Contents lists available at ScienceDirect

Genomics

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

Adaptive evolution at immune system genes and deep pregnancy implantation in primates

Alberto Civetta

Department of Biology, University of Winnipeg, Winnipeg, Manitoba, Canada

ARTICLE INFO

Article history:
Received 29 April 2014
Accepted 7 November 2014
Available online 15 November 2014

Keywords:
Positive selection
Pregnancy
KIR2DL4
HLA-E
Pleiotropy

ABSTRACT

A major evolutionary change in the lineage ancestral to humans, chimpanzee and gorilla (HCG) has been the embedding of the embryo into maternal tissue. Thus, the first layer of cells (trophoblast) to differentiate after fertilization must adapt to invade the uterus. Such event would likely leave signatures of positive selection at genes with roles in embryo implantation. Here, 163 pregnancy implantation genes are tested for evidence of adaptive diversification in the ancestral lineage to HCG. Two immune system genes, HLA-E and KIR2DL4 showed evidence of positive selection. Some of the positive selected sites involve amino acid substitution with predicted damaging effects on protein function, thus highlighting the possibility of antagonistic pleiotropic effects. Selection at a gene coding for a receptor expressed in uterine cells (KIR) that interacts with trophoblast human leukocyte antigen (HLA) genes suggests a main role for immunological adaptations in embryo deep invasion of the maternal endometrium.

1. Introduction

Embryo development in primates relies on major cell differentiation of the uterine epithelium. Such differentiation allows the establishment of proper blood circulation and thus secures nutrients and oxygen exchange with the fetus [1]. Data on pregnancy loss in humans suggests that the process of implantation and embryo development is fragile [2,3] and thus embryos are likely under high selective pressure to gain maternal resources for proper implantation and to maximize their chances for survival [4]. The gain of maternal resources by the embryo might come at a cost, but the origin and evolution of an embryo implantation system might have been boosted by fitness gains to both mother and offspring.

In humans, an endocrine feedback mechanism prepares the endometrium for implantation and placentation [5] and mother–embryo interactions that lead to the formation of the placenta are a first crucial step to secure a pregnancy's chance of successful progression. While all primates establish an early connection between the maternal blood supply and the embryo, there is a major difference in embryo implantation and pregnancy progression among primates. For example, in lorises and lemurs (Strepsirrhini) there is no invasion of the uterus (epitheliochorial placentation) whereas in New and Old World monkeys there is restricted invasion, in some cases with clear separation between the placenta and the decidua by a cytotrophoblastic shell [6]. Among apes, gibbons show the presence of a cytotrophoblastic shell and orangutans also show indications of restricted invasion as in gibbons [6]. In chimpanzees and gorillas the embryo becomes embedded into maternal tissue (interstitial implantation) as it does in humans, with artery invasion extending deep into the myometrium [1,6,7]. One hypothesis is that this differentiation in the lineage leading to human, chimpanzee and gorilla (HCG lineage) might have been driven by positive selection through the establishment of a tighter maternal-embryo connection. Such fetus–maternal connection allowed increases in acquisition of resources and success of pregnancy progression. If so, such selective pressure would be evident on genes that have been shown to associate with proper embryo implantation and implicated in disorders leading to pregnancy loss.

Sequence data was retrieved for genes associated with embryo implantation and pregnancy from different primates to perform phylogenetic analysis of selection. The histocompatibility antigen alpha chain E (HLA-E) and the killer cell immunoglobulin-like receptor two domains long cytoplasmic tail 4 (KIR2DL4) genes showed evidence of adaptive diversification in the HCG lineage. The signal of selection at a gene coding for a receptor expressed in the uterine natural killer cells (KIR) that interacts with trophoblast human leukocyte antigen (HLA) genes suggests a main role for immunological adaptations linked to embryo deep invasion of the maternal endometrium during pregnancy. Moreover, the fact that some positively selected sites are predicted to be deleterious in humans suggests that ancestral reproductive adaptations can have antagonistic pleiotropic effects in derived species.

2. Material and methods

2.1. Data collection and analysis

Three different sources were used for the identification of genes linked to pregnancy. The AmiGO 1.x browser (http://amigo1.geneontology.org/cgi-bin/amigo/go.cgi) was used to search for "female pregnancy" GeneOntology and filter by *H. sapiens* to identify genes linked to

E-mail address: a.civetta@uwinnipeg.ca.

http://dx.doi.org/10.1016/j.ygeno.2014.11.003
0888-7543/

decidualization, embryo implantation, or more broadly embryonic or maternal processes involved in female pregnancy. Genes linked to pre-eclampsia and pregnancy loss were retrieved using term searches from the Genetic Association Database (http://geneticassociationdb.nih.gov/). Finally, a list of genes associated with early stages of pregnancy such as uterine responsiveness/receptivity, embryo implantation and decidualization were retrieved from a recent review on mechanisms of implantation [8]. The outputs from different sources were manually curated to remove multiple entries for the same gene linked to different reference sources. Protein coding DNA sequence data was retrieved for all genes and species from ENSEMBL (http://uswest.ensembl.org/index.html) and the HomoloGene database within NCBI (http://www.ncbi.nlm.nih.gov/genbank/). Fasta files were created for all gene sequences for human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo abelii*), gibbon (*Nomascus leucogenys*), macaque (*Macaca mulatta*), marmoset (*Callithrix jacchus*) and bushbaby (*Otolemur garnettii*).

For each gene, protein sequences from different species were aligned using the MUSCLE program within MEGA 5.2, and the alignments were used to generate nucleotide sequence alignments in Phylogenetic Analysis by Maximum Likelihood (PAML) format [9]. In cases when more than one isoform was available per species, analysis was conducted using the longest isoform available, if differences were only in terms of isoform length, or by including the different isoforms if there were differences in sequence information. Positive selection along the branch ancestral to primates (Human, Chimpanzee and Gorilla) with deep embryo implantation (foreground) was tested using the species tree topology and the mixed branch-site model A (model = 2; NSsites = 2) within codeml in PAML [10]. The log-likelihood of the branch-site model was compared to the same model but fixing the ω value of the foreground branch to ω = 1 so that any significant variation in ω between foreground and background branches could be attributed to positive selection as opposed to differences in selective constraints [11]. The site models (M1a, M2a, M7, M8 and M8a) within PAML were used to test whether any signal of selection along the HCG ancestral branch could be driven by a general pattern of gene selection in primates. For proteins showing evidence of positively selected sites, the presence and location of signal peptide cleavage sites was identified using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/). Potentially functionally important motifs were identified using MotifScan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) as those with strong matches to PROSITE motif patterns and Pfam motif models and thus unlikely to be false positives.

A number of computational methods have been developed to estimate the potential deleterious effect of amino acid substitutions on protein function. Most methods use the principle that amino acid substitutions at evolutionary conserved sites across multiple and distant species would have deleterious effect on protein function. As well, the methods usually incorporate information on the potential physic–chemical severity of a change. The potential deleterious effect of positively selected amino acid substitutions at the branch ancestral to humans, chimpanzees and gorillas on human protein sequence was tested using different amino acid substitution program predictors. SIFT (http://sift.bii.a-star.edu.sg/) uses the distribution of amino acid residues observed at different positions in sequence alignments and computes a score based on sequence homology and the physic-chemical similarity to the inputted amino acid substitution. The scores range from 0 (deleterious) to 1 (neutral) with qualitative prediction of scores <0.05 as affecting protein function. [12]. PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) uses a naïve Bayes classifier to derive information from sequence alignments and protein structural properties and predicts the effect of an amino acid substitution on the function of a protein. PolyPhen-2 scores near 1 are more likely predicted to be deleterious [13]. PANTHER (http://www.pantherdb.org/) calculates position specific evolutionary conservation (PSEC) scores for each site using Hidden Markov Model families and assigns probabilities of changes being deleterious based on the severity of the amino acid change [14]. The tests were run by inputting the human amino acid sequence and the amino acid substitutions flagged by PAML as under positive selection.

3. Results

The search for genes associated with terms related to pregnancy and embryo implantation produced 163 single gene entries with sequence data for a minimum of four different species that included at least one species of the HCG trio and one outgroup (Supplementary Table 1). Among the 163 genes found, there were several genes found to associate with more than one process during embryo implantation and pregnancy, with 60 genes associated with the terms pre-eclampsia and pregnancy loss, 22 linked to uterine processes related to embryo receptivity, 51 with the attachment of the embryo to the uterine lining (implantation), 42 with decidualization (the changes occurring in the endometrium of the pregnant uterus just after the onset of implantation) and another 20 linked to reproductive processes occurring in the embryo or the mother that allow for the proper development of the embryo (Supplementary Table 1). Using various phylogenetic approaches, four genes were identified as exhibiting evidence of positive selection and became the focus of this study (Table 1).

Two genes, *Apolipoprotein L2 (APOL2)* and *forkhead box A2* (*FOXA2*), which showed evidence of positive selection (*APOL2*: $2\Delta\ell = 9.00$; $P = 0.003$ and *FoxA2*: $2\Delta\ell = 8.94$; $P = 0.003$) (Table 1 and Table S1) could not be fully assessed as true cases of adaptive diversification along the branch ancestral to human, chimpanzee and gorilla. First, an estimate of infinity for the foreground branch (HCG) ω of *APOL2* does not reflect the action of selection, as it is influenced by the lack of synonymous changes. Second, both genes lacked enough species sequences to truly establish that the bout of selection is specific to the HCG lineage. For *APOL2*, the lack of an orangutan sequence implies that the bout could be ancestral to great apes rather than HCG and for *FOXA2* the lack of orangutan and gibbon sequences makes it difficult to determine the ancestral lineage undergoing positive selection (i.e. HCG, great apes or apes). Finally, positive selection at both *APOL2* and *FOXA2* was not supported by the identification of positively selected sites with Bayes Empirical Bayes (BEB) posterior probabilities of being under selection higher than 50% (Table 1). The result suggests that while codon sites might be under positive selection, the pressure has likely being weak for the method to identify them.

Two other genes showing evidence of positive selection were *HLA-E* and *KIR2DL4* (*HLA-E*: $2\Delta\ell = 5.58$; $P = 0.018$ and *KIR2DL4*: $2\Delta\ell = 15.05$; $P < 0.001$) (Table 1and Table S1). It is possible that the bouts of selection detected at these genes along the branch ancestral to HCG is a common characteristic among members of the gene family and linked to aspects of their evolution not related to embryo implantation. Thus, other gene paralogs with and without roles in pregnancy were tested. No evidence of positive selection was found along the branch leading to primates with deep embryo implantation for any other Human Leukocyte Antigens (HLAs) or Killer cell Immunoglobulin-like Receptors (KIRs) (Table S2).

For *HLA-E* seven codon sites had BEB posterior probabilities higher than 50% of being under positive selection, with four sites (238Q, 242G, 243H and 251E) within the immunoglobin domain of the protein (Table 1 and Fig. 1). Four of the seven nonsynonymous changes were flagged by at least one of the amino acid substitution predictive model as causing deleterious changes in the human protein sequence composition. Three such substitutions (Q240W, G244D and E253D) were within the immunoglobin domain (Table 2 and Fig. 1). Nine codon sites were detected with BEB posterior probabilities higher than 50% as under positive selection for the KIR2DL4 protein, with four (29 N, 32 T, 55I and 69 F) within or adjacent to the immunoglobin domain (Table 1 and Fig. 2). Five of the codon sites flagged as under positive selection were found to cause potentially deleterious amino acid substitution to the