Evidence for Positive Selection in the C-terminal Domain of the Cholesterol Metabolism Gene *PCSK9* Based on Phylogenetic Analysis in 14 Primate Species

Keyue Ding, Samantha J. McDonough, Iftikhar J. Kullo*

Division of Cardiovascular Diseases, Mayo Clinic and Foundation, Rochester, Minnesota, United States of America

Background. Cholesterol homeostasis is maintained through finely tuned mechanisms regulating intestinal absorption, hepatic biosynthesis and secretion as well as plasma clearance. Proprotein convertase subtilisin/kexin type 9 (PCSK9) is a secreted enzyme of the serine protease family that reduces cellular uptake of plasma low-density lipoprotein (LDL) cholesterol by promoting LDL receptor (LDL-R) degradation. Species-specific positive selection has been noted in the *LDLR* promoter, leading to differential expression of *LDLR* among primates. Whether *PCSK9* experienced significant selective pressure to maintain a functional relationship with its target protein, LDL-R, is unknown. ***Methodology/Principal Findings***. We compiled the sequences of the coding regions of *PCSK9* from 14 primate species in the clade of Hominoids, Old World monkeys and New World monkeys. To detect selective pressure at the protein level, the ratios of nonsynonymous/synonymous substitution rate (d_N/d_S) under different evolutionary models were calculated across the phylogeny of *PCSK9*. Maximum likelihood analyses of d_N/d_S ratios for the aligned coding region sequences among 14 primate species indicated that *PCSK9* was subject to a strong functional constraint (i.e., purifying selection). However, positive selection was noted in the functional carboxyl-terminal (C-terminal) domain in many branches across the phylogeny, especially in the lineage leading to the orangutan. Furthermore, at least five positively selected amino acids were detected in this lineage using the branch-site model A. In a sliding-window analysis, several d_N/d_S peaks in the C-terminal domain in both the human and the orangutan branches were noted. ***Conclusions***. These results suggest that among primates, differential selective pressure has shaped evolutionary patterns in the functional domains of *PCSK9*, an important regulator of cholesterol homeostasis.

INTRODUCTION

The low-density lipoprotein (LDL) receptor gene (*LDLR*) plays a key role in cholesterol homeostasis by receptor-mediated endocytosis of LDL cholesterol. Proprotein convertase subtilisin/kexin type 9 (*PCSK9*, MIM 607786) – a secreted enzyme of the serine protease family – is a newly discovered regulator of *LDLR* [1–3]. The human *PCSK9* locus spanning 25 kb and containing 12 exons, resides on chromosome 1p32. The *PCSK9* protein contains five functional domains, a signal-peptide and prodomain at its N-terminus, followed by a catalytic domain, a putative domain and a cysteine-rich carboxy-terminal domain [4–6] (Figure 1). *PCSK9* induces LDL-R breakdown [7], internalization and recycling [8,9] and thereby reduces LDL clearance, and increases plasma levels of LDL cholesterol. Overexpression of wild-type *Pcsk9* gene in mice results in reduced number of LDL-R and hypercholesterolemia [4,8,9].

Mutations in *PCSK9* can cause severe autosomal dominant hypercholesterolemia [4,8–14] (i.e., 'gain-of-function' mutations), and also low circulating levels of LDL cholesterol [15–19] (i.e., 'loss-of-function' mutations). Kotowski et al. [17] described a spectrum of nonsense/missense mutations in *PCSK9* that were associated with low or elevated LDL cholesterol levels, in both black and white subjects. Relatively common sequence variants in *PCSK9* also contribute significantly to inter-individual variation in plasma levels of LDL cholesterol in the general population [17]. Cohen et al. [20] showed that two nonsense mutations (Y142X and C679X) in blacks, and one missense mutation (R46L) in whites, were associated with reduced plasma levels of LDL cholesterol and lower incidence of coronary heart disease events. We have summarized these non-synonymous variations in *PCSK9* in Figure 1.

Extant primates show a wide range of phenotypic adaptations to diverse environmental conditions, including substrate and diet [21]. Significant differences in lipid profiles occur among primates; for example, New World and Old World monkeys have significantly lower serum total cholesterol, triglycerides, and LDL cholesterol levels than Hominoids [22]. Among Hominoids, gorillas have the highest circulating total cholesterol, triglycerides, and high-density lipoprotein (HDL) cholesterol levels [22]. Caceres et al. [23] found that several genes related to lipid metabolism were differentially expressed in humans and non-human primates. *LDLR* has also been shown to be differentially expressed among mammals [24].

Activation of the sterol regulatory element binding protein-2 (SREBP-2), a key transcription factor of *LDLR*, not only leads to increased expression of *LDLR*, but also of *PCSK9* [25], and the mRNA expression of *LDLR* and *PCSK9* are coordinately up-regulated in absence of sterols [26]. The dual regulation of *LDLR* [26] suggests that *PCSK9* might be involved in a co-evolutionary

· ·
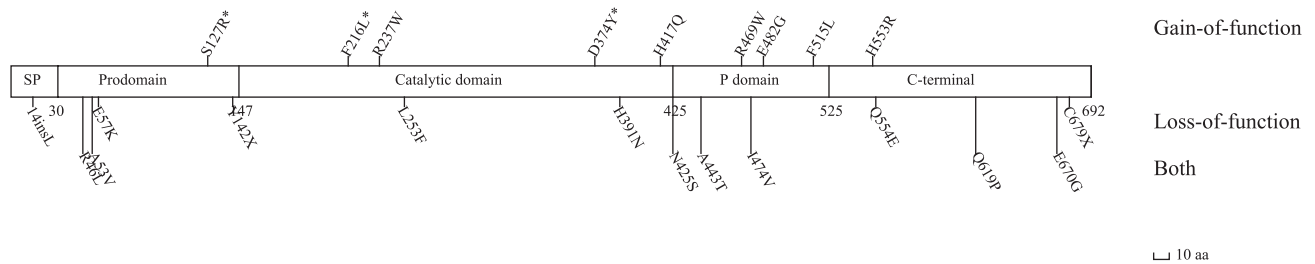
Figure 1. Distribution of non-synonymous variations along *PCSK9*. There are five functional domains in the *PCSK9* protein [4–6]: 1) a signal peptide (SP) (1∼30 aa), 2) a prodomain (31∼147 aa), 3) a catalytic domain (148∼425 aa), 4) a putative P domain (426∼525 aa), and 5) a C-terminal domain (526∼691 aa). Gain-of-function mutations are only identified in families with hypercholesterolemia [10] or subjects with high LDL cholesterol levels [17], and loss-of-function mutations in subjects with low levels of LDL cholesterol [15–19]. Some non-synonymous mutations have been identified in subjects with either high or low plasma LDL cholesterol [4,17], and are labeled 'both' in the figure. Rare mutations found in families with autosomal dominant hypercholesterolemia are labeled with an asterisk. The gain-of-function mutations are: S127R, F216L, R237W, D374Y, H417Q, R469W, E482G, F515L and H553R. The loss-of-function mutations are: 14insL, E57K, Y142X, L253F, H391N, Q554E, and C679X. Mutations associated with either high- or low- plasma levels of LDL cholesterol subjects are: R46L, A53V, N425S, A443T, I474V, Q619P, and E670G.
doi:10.1371/journal.pone.0001098.g001

network of LDL cholesterol metabolism, i.e., variation at one gene evolving with variation at the other. *PCSK9* is only found in vertebrates, suggesting it is the product of recent evolution [27]. However, it is unclear whether natural selection has driven the evolution of *PCSK9* in vertebrates, especially in primates.

The ratio of nonsynonymous/synonymous substitution rate (i.e., $\omega = d_N/d_S$) provides a sensitive measure to detect selective pressure at the protein level [28]. A significantly higher non-synonymous substitution rate than synonymous substitution rate (i.e., $d_N/d_S > 1$) is evidence for adaptive evolution at the molecular level, whereas $d_N/d_S < 1$ suggests purifying selection (i.e., selective constraint) [29]. This criterion has been used to identify several cases of positive selection, such as the primate stomach lysozyme (*LYZ*) [30,31], and *BRCA1* in humans and chimpanzees [32] (for a list of genes under positive selection in the human lineage, see review by Sabeti et al. [33]).

In the present study, we used a phylogeny-based maximum likelihood method to analyze nonsynonymous and synonymous substitution rate of *PCSK9* sequences across a range of primates, including Hominoids, Old World monkeys, and New World monkeys. We aimed to address three questions: 1) whether the ratio of nonsynonymous and synonymous evolutionary rate of *PCSK9* varied significantly among various primate lineages, 2) whether there is an episodic adaptive evolution of *PCSK9* in primates, and 3) which, of any, amino acids of *PCSK9* were under positive selection.

RESULTS

Comparative analysis of coding regions of PCSK9

We characterized the coding regions of *PCSK9* in human and non-human primates (Table 1). There was considerable evidence for interspecies genomic alterations within *PCSK9* (Figure 2). First, in the signal peptide domain, a variable number of CTG codons [encoding Leucine (Leu, L)] were identified (Figure 2). A species-specific nine Leu repeat (L9) was noted only in humans and chimpanzee; whereas in other species, the number of Leu repeats varied from L6 to L8. In the CTG repeat region, interspecies nucleotide differences were synonymous except nonsynonymous changes in the Old World monkeys (CTG→CCA, 46→48) and in spider monkey (CTG→CGG, 67→69). Second, in the C-terminal domain, species-specific loss-of-function mutation was seen in the clade of New World monkeys (via a premature stop codon in tamarin and dusky titi). We speculate that this nonsense mutation

leads to a loss-of-function since the adjacent C679X mutation in humans is a loss-of-function mutation.

Variable d_N/d_S ratios for the C-terminal domain of PCSK9 among primate lineages

A neighbor-joining (nj) phylogenetic tree of *PCSK9* from 14 primate species was reconstructed from the coding sequence alignment, and the maximum likelihood estimate of the tree topology was acquired using the 'hyphy' package [34]. We used this gene tree in subsequent analyses to detect whether non-neutral evolution might have operated on *PCSK9*.

Under a one-ratio model, which assumes the same d_N/d_S for the entire tree, the cumulative d_N/d_S for the coding regions of *PCSK9* was 0.186 (the log likelihood value $l_0 = -5109.09$). We tested variable d_N/d_S ratios for *PCSK9* among lineages, using the free-ratio model, which assumes a different d_N/d_S ratio for each branch in the phylogeny (Figure S1). The free-ratio model led to a log likelihood $l_1 = -5097.59$. The free-ratio model was not found to

Table 1. A listing of the primates in this study.

Species	Scientific name	Lineage	Accession number
Human	*Homo sapiens*	Hominoid	EF692496
Chimpanzee	*Pan troglodytes*	Hominoid	EF692497
Bonobo	*Pan paniscus*	Hominoid	EF692498
Gorilla	*Gorilla gorilla*	Hominoid	EF692499
Orangutan	*Pongo pygmaeus*	Hominoid	EF692500
Pigtailed macaque	*Macaca nemestrina*	Old World Monkey	EF692501
Rhesus macaque	*Macaca mulatta*	Old World Monkey	EF692502
Colobus	*Colobus guereza*	Old World Monkey	EF692503
Dusky titi	*Callicebus moloch*	New World Monkey	EF692504
Tamarin	*Saguinus labiatus*	New World Monkey	EF692505
Marmoset	*Callithrix jacchus*	New World Monkey	EF692506
Squirrel monkey	*Saimiri boliviensis*	New World Monkey	EF692507
Spider monkey	*Ateles geoffroyi*	New World Monkey	EF692508
Woolly monkey	*Lagothrix lagotricha*	New World Monkey	EF692509

doi:10.1371/journal.pone.0001098.t001

```
                    **** *** ***** ****** * .*:****::*.***********:* .:* ::*:*********:.********** .::: *.**************** *:***** ******
          Human MGTVSSRRSWWPLPLLLLLLLLLGPAGARAQEDEDGDYEELVLALRSEEDGLAEAPEHGTTATFHRCAKDPWRLPGTYVVVLK-EETHLSQSERTARRLQAQAARRGYLTKILHVFHGLLPGFLV
     Chimpanzee .............................................................-..............-.......................................
        Gorilla ..............--...........................................................................................
         Bonobo ............................................................................................................
      Orangutan ..............--...................................A.......-...-..R....................................D....
Pigtailed_Macaque .............P................................D....A.......-..-..R.....................................H....
 Rhesus_Macaque .............P................................D....A.......-..-..R.....................................H....
        Colobus .............P.....L..........................D....A.......-..-..R.....................................H....
     Dusky_Titi ........L.....--.......P......D...........P.D.LQ..A.......DSDA.R..P................I.L....H....
        Tamarin ......L....--.......T.....D.D.............E...D.LQN.A.......-..-..R..P................I.L....D....
       Marmoset ......L....--.....T.T....ED.D.............VD.LQ..A.......-..-..R..P................I.L....D....
 Squirrel_Monkey ......L....--.......D..................D.LQ..A.........E.......-..-..R..P................I.L....D....
   Spider_Monkey ....R...L....--.........R.......D...................Q..A..........-..-..QR..P................I.L....D....
   Woolly_Monkey ....R...L....--.........D...................LQ..A..........-..-..QRL.P................I.L....D....
          ruler 1........10.........20.........30.........40.........50.........60.........70.........80.........90.......100.......110.......120.....

                    *** *******:*********** *************** .**:******* .***************** .************* .**:**************************************.::*
          Human KMSGDLLELALKLPHVDYIEEDSSVFAQSIPWNLERITPPRYRADEYQPPDGGSLVEVYLLDTSIQSDHREIEGRVMVTDFENVPEEDGTRFHRQASKCDSHGTHLAGVVSGRDAGVAKGASMRS
     Chimpanzee ............................................................................................................
        Gorilla ............................................................................................................
         Bonobo ............................................................................................................
      Orangutan ............................................................................................................
Pigtailed_Macaque ...........................A...........K....................S..........................GL..
 Rhesus_Macaque ...........................A...........K....................S..........................GL..
        Colobus ...........................A...........K....................S..........................GL..
     Dusky_Titi ..R......R...............A.............N................S.............S...........L..
        Tamarin ..R......R...............A.Q...........N...........G.......GS..K...............LH.
       Marmoset ..R......R...............A.............N...........G.......GS................L..
 Squirrel_Monkey ..R......R...............A.............N...........G.......GS................L..
   Spider_Monkey ..R......R.........Y.....A.............N...........G.......S................L..
   Woolly_Monkey ..R......R.........Y.....A.............N...........G.......S................L..
          ruler ..130.......140.......150.......160.......170.......180.......190.......200.......210.......220.......230.......240.......250

                    *:***********.:******* *.***************.:*********:****:*** * ****:***********************:***** ****************************:.*** ********** **:****:
          Human LRVLNCQGKGTVSGTLIGLEFIRKSQLVQPVGPLVVLLPLAGGYSRVLNAACQRLARAGVVLVTAAGNFRDDACLYSPASAPEVITVGATNAQDQPVTLGTLGTNFGRCVDLFAPGEDIIGASSD
     Chimpanzee ............................................................................................................
        Gorilla ............................................................................................................
         Bonobo ............................................................................................................
      Orangutan .................................M.......................................................................
Pigtailed_Macaque .....................................F...............................................
 Rhesus_Macaque .....................................F...............................................
        Colobus .....................................F...............................................
     Dusky_Titi .H...........SA.......................R...G.......A.................L...............
        Tamarin ..............S.......C.........................R....A.............................
       Marmoset ..............S.................................A.............................
 Squirrel_Monkey ..............S.................................................................
   Spider_Monkey ..............S.......N.......................K....A.............................
   Woolly_Monkey ..............S.......N.......................A.............................
          ruler .......260.......270.......280.......290.......300.......310.......320.......330.......340.......350.......360.......370.....

                    *******:*************:***..*.:***********************.***************:***** **********************:.*** ********** **:****:
          Human CSTCFVSQSGTSQAAAHVAGIAAMMLSAEPELTLAELRQRLIHFSAKDVINEAWFPEDQRVLTPNLVAALPPSTHGAGWQLFCRTVWSAHSGPTRMATAIARCAPDEELLSCSSFSRSGKRRGER
     Chimpanzee ...................................................................V.............
        Gorilla ...................................................................V.............H
         Bonobo ...................................................................V.............
      Orangutan ..................V................V.......................................
Pigtailed_Macaque ......R................................R..........V...Q.....
 Rhesus_Macaque ......R................................R..........V...Q.....
        Colobus ......R........................................R..........VV.....
     Dusky_Titi ......R..............................T..........M.....
        Tamarin ......R.......................E..........M.....
       Marmoset ......R.......K........................M.............R....
 Squirrel_Monkey ......R.......V........................M.............S.....
   Spider_Monkey ......R........................M.....
   Woolly_Monkey ......R........................M.....
          ruler ..380.......390.......400.......410.......420.......430.......440.......450.......460.......470.......480.......490.......500

                    ::****: ** * *****:** *:********.***:****** :.****. **:* *:*****:***::;****** .:*:* * .**:** ** **:** ***********:.******* *
          Human MEAQGGKLVCRAHNAFGGEGVYAIARCCLLPQANCSVHTAPPAEASMGTRVHCHQQGHVLTGCSSHWEVEDLGTHKPPVLRPRGQPNQCVGHREASIHASCCHAPGLECKVKEHGIPAPQEQVTV
     Chimpanzee ....................................I.......G.....................M.........R.............
        Gorilla ....................................G.....................
         Bonobo ....................................G.....................
      Orangutan ..............C..............GSG....L....V...............R.......
Pigtailed_Macaque I.....R..............V........G.................I.
 Rhesus_Macaque I.....R..............V........G.................I.
        Colobus I.....R..............V........G.................I.
     Dusky_Titi I.....RR..L.P........V........G.G....A...........MK......K.....D..M.SG..T......Y.......L.....
        Tamarin I.....RR..L..................G.G....A....I.........E..H...M...G..T...........L.....
       Marmoset I.....RR..L..................G.G....A....I.........G..HD...M...G..T...........L.....
 Squirrel_Monkey I.....RR..L......K...........I.....G..........A....A......E........G...S..M..SG..T..T.......L.....
   Spider_Monkey IK....RR..L..................G.G....H...........S...V..D..M.SG..T......L.....
   Woolly_Monkey I...........................GTG..........V..D..M.SG..T...........L.....
          ruler .......510.......520.......530.......540.......550.......560.......570.......580.......590.......600.......610.......620.....

                    :**:*******..******:****.:*:*****:*::*:. *::::::;***** **.** ** :**
          Human ACEEGWTLTGCSALPGTSHVLGAYAVDNTCVVRSRDVSTTGSTSEEAVTAVAICCRSRHLAQASQELQ
     Chimpanzee ..........................A.........A...........
        Gorilla ...............GI.............R.....LA..........V.......
         Bonobo ...............GI.............R.....LA..........V.......
      Orangutan .......................I.....MA......R...........
Pigtailed_Macaque ...D....................A...........V......
 Rhesus_Macaque ...D......P...................K....A...........V......
        Colobus ...D......N...................MA...........V......
     Dusky_Titi T..................D.................A...........X...D..
        Tamarin T..................D...........T.A...........Q.....X...
       Marmoset T..........I.....D...........S.....T.AT.........Q.....
 Squirrel_Monkey ...........I.....D.....Q...........A................
   Spider_Monkey ...................D...........N...Q.A...............
   Woolly_Monkey ...................D.........G..NI...........W........
          ruler ..630.......640.......650.......660.......670.......680.......690...
```
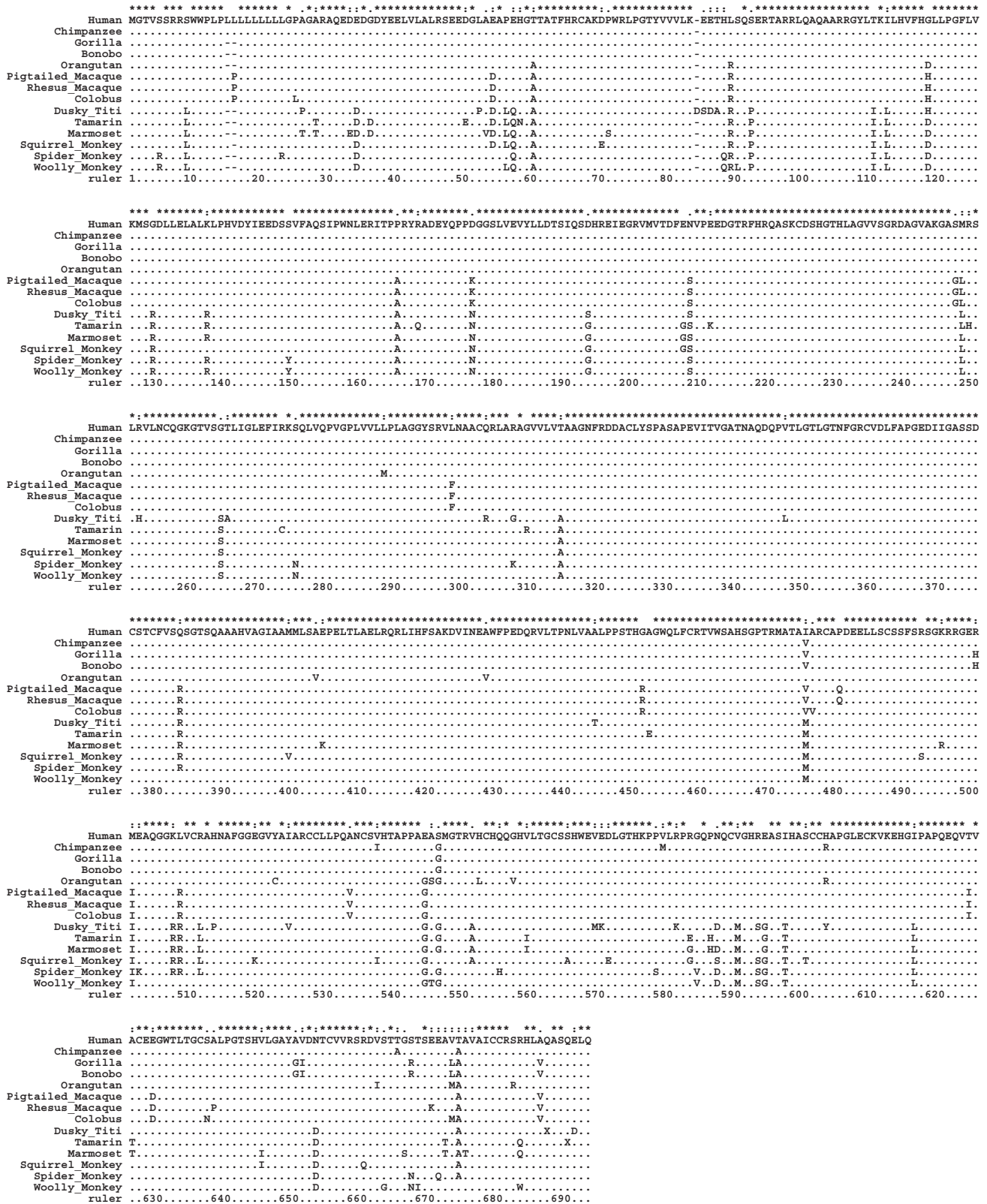
Figure 2. Protein sequence alignment of *PCSK9* in 14 primates. "." Indicates identity to the first sequence (i.e., human) in each alignment. "-" indicates an alignment gap, and "X" indicates a stop codon. The coordinates after 84 should be minus one to be consistent with that in Human reference sequence (NP_777596), since an insertion at position 84 was present in the dusky titi. The signal peptide (SP) domain (1–90) shows the evolution of Leucine (Leu) repeats (15–23) in *PCSK9*, and the C-terminal domain shows the premature stop codon (X) in the tamarin (686), and dusky titi (689).

be significantly better than the one-ratio model (the statistic $2\Delta l = 2(l_1 - l_0)$, $P = 0.480$, degrees of freedom (df) = 24).

In addition to testing the entire coding regions, it is important to test the structural and functional domains of proteins separately [35]. We evaluated selective pressures in different structural and functional domains by repeating the d_N/d_S analyses within these domains (the domain structure of PCSK9 is shown in Figure 1). For the C-terminal domain of PCSK9, the free-ratio model ($l_1 = -1351.35$) was found to be significantly better than the one-ratio model ($l_0 = -1369.71$, the cumulative $d_N/d_S = 0.386$) ($2\Delta l = 36.72$, $P = 0.047$, df = 24), suggesting variable selective constraint across the phylogeny for the C-terminal domain (Figure 3). We did not observe a significant difference between these two models for each of the other four domains and when the four domains were combined (data not shown).

Non-neutral evolution of the PCSK9 C-terminal domain

Comparison of the rates of nonsynonymous and synonymous DNA changes (i.e., the ratios of d_N/d_S) between species can be used to assess the types of selective pressures that may have acted on a gene [36]. In the entire PCSK9 sequence, d_S exceeded d_N in most of the branches in the primate phylogeny ($d_N/d_S < 1.0$) (Figure S1), indicating that functional constraint (i.e., purifying selection) might have acted on PCSK9 throughout primate evolution. In the branch of bonobo and gorilla, the d_N/d_S ratio was = 1 (i.e., neutral evolution).

We then compared the values of d_N and d_S between species in the C-terminal domain of PCSK9 since non-homogeneity in d_N/d_S ratio was noted among the primate lineages. We found that many branches of the primate phylogeny, including internal branches, showed evidence of evolution under relaxed selective constraint or
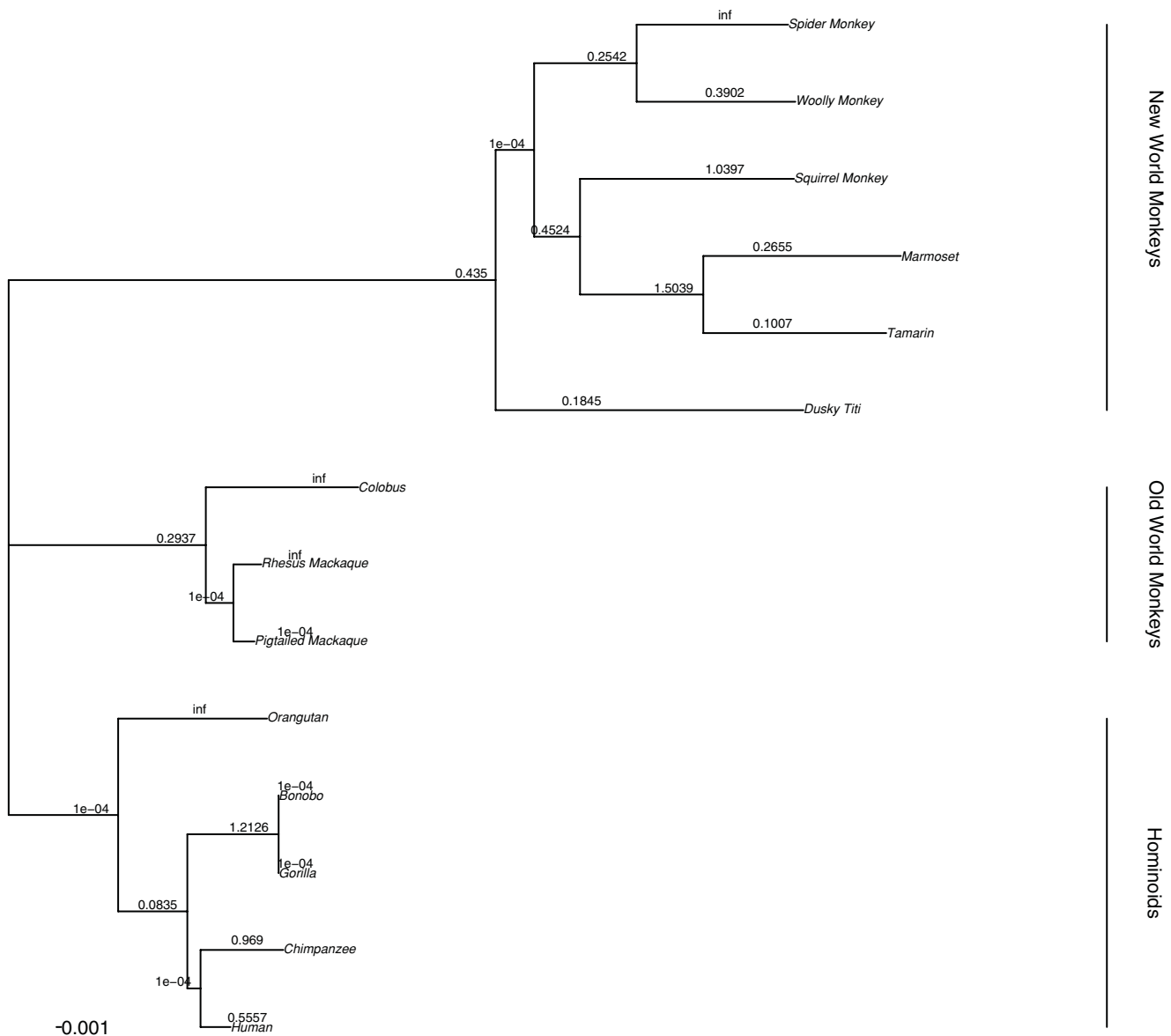
Figure 3. Ratios of d_N/d_S estimated for the C-terminal domain of PCSK9 in indicated branches of the primate phylogeny. Values of d_N/d_S along each branch were calculated by using the free-ratio model using the CODEML program in 'PAML' [72]. Branch lengths were estimated by maximum likelihood under this model. A d_N/d_S value of >1 suggests that positive selection has acted along that lineage. 'Inf' indicates cases where $d_S = 0$. The phylogenetic tree was deduced from the entire coding sequence of PCSK9.
doi:10.1371/journal.pone.0001098.g003

positive selection (i.e., $d_N/d_S > 1.0$) (Figure 3). In the Hominoid clade, d_N/d_S values was infinity ($\frac{N \times \tilde{d}_N}{S \times d_S}$: 7.1/0.0) in the lineage leading to orangutan, and 1.218 (5.1/1.1) to the common ancestors of bonobo and gorrila. The d_N/d_S in the chimpanzee and human lineages approximated one [0.9890 (4.0/1.1) and 0.5657 (2.0/1.0), respectively], indicating relaxed selective constraint in these two lineages. In addition, we noted $d_N/d_S > 1$ in lineages leading to colobus and rhesus macaque in the clade of Old World monkeys, as well as spider monkey, squirrel monkey, and the common ancestors of marmoset and tamarin (Figure 3). Thus, the C-terminal domain has been subject to positive selection for at least 33 million years (i.e., the primate divergence time) [37].

Then, by two-ratio likelihood tests using PAML, we tested for the presence of positive selection in the C-terminal domain of PCSK9 in the branch of orangutan and the lineage leading to the common ancestors of bonobo and gorilla in the Hominoid clade. Log likelihood values and d_N/d_S estimates from each maximum likelihood model were considered, and the likelihood ratio test results are presented in Table 2. The null hypotheses 1–3 were rejected, indicating that the d_N/d_S ratio in the branch of orangutan is significantly higher than the background ratio of d_N/d_S (i.e., the null hypothesis of d_N/d_S ratio homogeneity among lineages was rejected). Although the alternative hypotheses of positive selection ($d_N/d_S > 1$) in the branch of orangutan (null hypotheses 7 and 8 in Table 2) were not accepted at the level of 0.05, the statistical significance was marginal ($P = 0.087$ and 0.090, respectively). However, we did not observe the d_N/d_S ratio in the lineage leading to bonobo and gorilla to be significantly different from the background d_N/d_S ratio.

To find out whether the C-terminal domain was under relaxed selective constraint or positive selection, we also plotted d_N/d_S ratios estimated by Nei-Gojobori method for pairwise comparisons within the primates (Figure 4). In the whole gene sequence and non C-terminal domains region, d_N/d_S was <1 in all pairwise comparisons, indicating a history of functional selective constraint. However, a higher d_N/d_S value was noted in the C-terminal domain, including $d_N/d_S > 1$ in five out of 10 pairwise comparisons

within Hominoids. For example, a d_N/d_S ratio of 1.122 was noted in the human vs. orangutan comparison.

Amino acids sites under positive selection

Finally, we identified the particular codon sites that have been subject to positive selection using the site-models [38,39] and the branch-site models [40]. Table 3 lists the log likelihood values and parameter estimates for the C-terminal domain of PCSK9 under several site models and branch-site models. We used two likelihood ratio tests (LRTs) to test for positive selection. In the site models, the first test compared M1a (neutral) against M2a (selection) [38,40,41], in which $2\Delta l$ is 0.68 (df = 2, $P > 0.05$), and no amino acid sites were under positive selection. The second test compared M7 against M8, in which no sites were shown under positive selection ($2\Delta l = 1.66$, $P > 0.05$). We also used the branch-site model A to detect the codon sites under positive selection by the Bayes Emprical Bayes (BEB) approaches in the lineage of orangutan (i.e., the foreground lineage) [40]. The $2\Delta l$ between the null model (neutral, $l = -1366.56$) and the alternative model (selection, $l = -1363.55$) is 6.02. The critical values at the 5% and 1% levels for the LRT are 2.71 and 5.41, respectively [42]. Thus, the test for the branch-site model A is significant (df = 2, $P < 0.01$), indicating presence of codon sites under positive selection (544A, 551H, 556G, 661V, and 681S, $P_b > 95\%$) in the C-terminal domain of PCSK9.

Sliding window analysis

The d_N/d_S profiles in the sliding window analysis across PCSK9 sequence are shown in Figure 5. As expected, the cumulated d_N/d_S ratio in primates in the sliding window analysis appear quite stochastic and bears weak correlation to the domain structure, although the d_N/d_S is slightly higher in the C-terminal domain. However, in the lineages leading to humans and orangutan, we observed that nonsynonymous substitutions were significantly more concentrated within C-terminal domain of PCSK9 (i.e., three d_N/d_S peaks). The d_N/d_S peaks were consistent between the lineage leading to humans and to orangutan.

Table 2. Log likelihood values, parameter estimates under different models, and likelihood ratio statistics ($2\Delta\ell$) for d_N/d_S hypotheses testing.

Model	Para[a]	ln l[b]	d_N/d_S (ω)			Model compared	Null hypothesis	$2\Delta l$	P (χ^2, df = 1)[c]
			ω_0	ω_h	ω_g				
A. One ratio: $\omega_0 = \omega_h = \omega_g$	27	−1369.71	0.3856	$=\omega_0$	$=\omega_0$	A vs. D	1. ($\omega_h = \omega_g$) = ω_0	6.24	0.012*
B. Two ratios: $\omega_0 = \omega_h$, ω_g	28	−1366.30	0.3503	$=\omega_0$	Inf	A vs. B	2. $\omega_g = \omega_0$	6.82	0.009*
C. Two ratios: $\omega_0 = \omega_g$, ω_h	28	−1369.21	0.3686	1.0106	$=\omega_0$	C vs. E	3. $\omega_g = \omega_0$	7.02	0.008*
D. Two ratios: ω_0, $\omega_h = \omega_g$	28	−1366.59	0.3314	2.6347	$=\omega_h$	A vs. C	4. $\omega_h = \omega_0$	1.00	0.317
E. Three ratios: ω_0, ω_h, ω_g	29	−1365.70	0.3326	1.0027	Inf	B vs. E	5. $\omega_h = \omega_0$	1.20	0.273
F. Two ratios: $\omega_0 = \omega_h$, $\omega_g = 1$	27	−1367.76	0.3502	$=\omega_0$	1.0000	D vs. H	6. ($\omega_h = \omega_g$)≤ 1	1.10	0.394
G. Two ratios: $\omega_0 = \omega_g$, $\omega_h = 1$	27	−1369.21	0.3686	1.0000	$=\omega_0$	B vs. F	7. $\omega_g \leq 1$	2.92	0.087†
H. Two ratios: ω_0, $\omega_h = \omega_g = 1$	27	−1367.14	0.3324	1.0000	1.0000	E vs. I	8. $\omega_g \leq 1$	2.88	0.090†
I. Three ratios: ω_0, ω_h, $\omega_g = 1$	28	−1367.14	0.3324	1.0140	1.0000	C vs. G	9. $\omega_h \leq 1$	0.00	1.000
J. Three ratios: ω_0, $\omega_h = 1$, ω_g	28	−1365.70	0.3326	1.0000	Inf	E vs. J	10. $\omega_h \leq 1$	0.00	1.000

Note. – ω_h is d_N/d_S ratio in lineage leading to the common ancestor of gorilla and bonobo; ω_g is d_N/d_S ratio in lineage leading to orangutan; ω_0 is the background d_N/d_S ratio.
[a]The number of parameters estimated in the model.
[b]Log likelihood values.
[c]P values were obtained from the difference in the two log likelihood values of two models ($2\Delta l$) with χ^2 distribution (df = 1).
*Significant ($P < 0.05$); † Significance is marginal ($P < 0.10$).
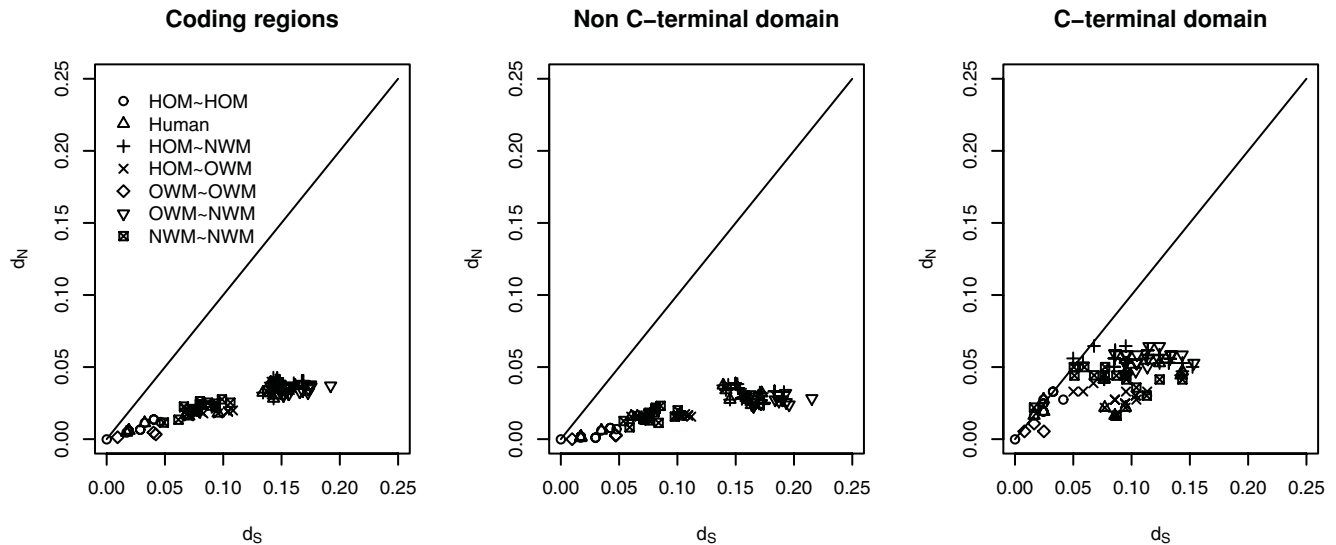doi:10.1371/journal.pone.0001098.t002

Figure 4. Positive selection or relaxed selective constraint of the C-terminal domain of *PCSK9*. d_N is plotted versus d_S for all pairwise combinations of primate sequences. The pairwise ratios of d_N/d_S were calculated using the Nei-Gojobori method implemented in the package 'PAML' [72]. Pairwise combinations of Hominoids (HOM), Old World monkeys (OWM), and New World monkeys (NWM) are plotted; for example, 'Human' represents the points that are making comparisons between human and another primate. We plotted the entire sequence, non C-terminal domain, and C-terminal domain separately. The higher pairwise d_N/d_S ratio in the C-terminal domain suggests that this domain is evolving in a non-neutral model, which maybe due to positive selection or relaxed selective constraint in some lineages. The entire sequence and non C-terminal domain of *PCSK9* showed a net signature of purifying selection.
doi:10.1371/journal.pone.0001098.g004

DISCUSSION

The main finding of our study is that there is evidence for functional constraint (i.e., purifying selection) in the coding sequences of *PCSK9* through primate evolution. We noted that a functional domain of *PCSK9* (i.e., C-terminal domain) was less conserved at the amino acid level than other gene regions, and likelihood ratio tests (LRTs) revealed evidence of positive selection in the lineage leading to orangutan of the Hominoid clade on this domain. Furthermore, we identified the particular codon sites that have been subject to positive selection in this lineage. We discuss the implications of these comparative sequence data for un-

derstanding the evolutionary history of primate *PCSK9*, hypotheses concerning their role in primate phenotypic evolution, and insights into *PCSK9*-associated human diseases.

Evolutionary history of Leucine (L) repeats in the signal peptide (SP) domain and premature stop codon in the C-terminal domain

Comparative sequence analysis revealed a dynamic evolutionary history of leucine (Leu, L) repeats in the signal peptide domain. The number of Leu repeats varied from L9 to L6 among different

Table 3. Log likelihood values and parameter estimates under the site models and branch-site models.

Model	Para[a]	ln *l*[b]	Estimates of parameters	2Δ*l*[c]	Positively selected sites[d]
M0 (one-ratio)	1	−1369.71	$\omega_0 = 0.3856$		not allowed
Site models					
M1a (Nearly Neutral)	2	−1366.56	$p_0 = 0.7704$ ($p_1 = 0.2296$); $\omega_0 = 0.2110$ ($\omega_1 = 1$)		not allowed
M2a (Positive Selection)	4	−1366.22	$p_0 = 0.9032$; $p_1 = 0$ ($p_2 = 0.0968$); $\omega_0 = 0.2736$ ($\omega_1 = 1$), $\omega_2 = 1.7797$	(M1a vs. M2a) 0.68	none
M7 (beta)	2	−1367.06	$p = 0.6432$; $q = 0.9806$		not allowed
M8 (beta & ω)	5	−1366.23	$p_0 = 0.9054$; $p = 37.7260$; $q = 99.0000$ ($p_1 = 0.0947$); $\omega = 1.7952$	(M7 vs. M8) 1.66	none
Branch-site models					
M1a	2	−1366.56	$p_0 = 0.7704$ ($p_1 = 0.2296$); $\omega_0 = 0.2110$ ($\omega_1 = 1$)		
Model A	4	−1363.55	$p_0 = 0$; $p_1 = 0.7632$ ($p_2 = 0.2368$); $\omega_0 = 0.1827$ ($\omega_1 = 1$), $\omega_2 = $ Inf	(M1a vs. Model A) 6.02[**]	544A, 551H, 556G, 661V, 681S

[a]The number of parameters estimated in the model.
[b]Log likelihood values.
[c]LRT to detect positive selection [**] *P* < 0.01.
[d]Amino acid sites under positive selection based on a Bayes Empirical Bayes (BEB) probability > 95%. No amino acids sites under the model M2a and M8 were shown to be under positive selection.
doi:10.1371/journal.pone.0001098.t003
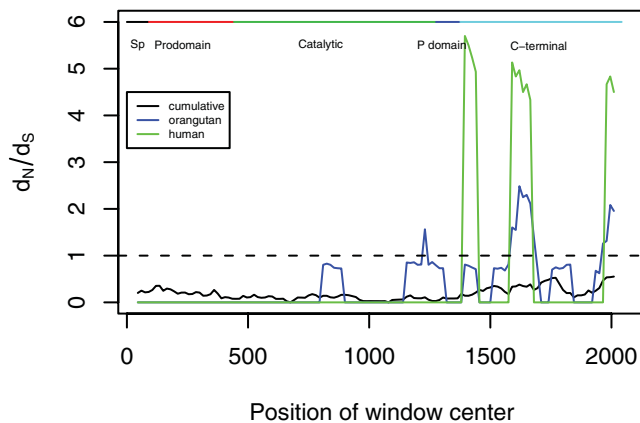
Figure 5. Sliding-window analysis of the cumulative d_N/d_S across primates (black), the lineage leading to human (green), and orangutan (blue). The gene average d_S across primates is 0.5680, in the lineage leading to human is 0.0030, and orangutan is 0.0204.
doi:10.1371/journal.pone.0001098.g005

clades (Figure 2). An additional in-frame insertion (CTG) leading to a L9→L10 polymorphism in African-Americans and Caucasians [19] is associated with hypocholesterolemia. We speculate that the number of Leu repeat may influence levels of the *PCSK9* protein and thereby levels of LDL cholesterol, although this needs confirmation.

In the C-terminal domain of *PCSK9*, a premature stop codon was seen in the New World monkeys – tamarin and dusky titi (Figure 2) – but not in the Hominoids. As mentioned before, monkeys have significantly lower LDL cholesterol levels than Hominoids [22]. It is unknown whether the loss-of-function mutation in the C-terminal domain is a random phenomenon or a common feature that influences cholesterol metabolism. A premature stop codon mutation in human *PCSK9* (C679X) is considered to be under positive selection, and it is speculated that loss of *PCSK9* function interferes with the life cycle of the malaria parasite through cholesterol restriction [43,44]. The "less-is-more" hypothesis of Olson [45] posits that loss of gene function during hominoid evolution may in some cases have conferred a fitness benefit and led to adaptive evolution that may help explain differences among primates [46].

Amino acid substitution patterns
The d_N/d_S ratio for the coding regions of *PCSK9* across the primate species was <1 (cumulative d_N/d_S = 0.186). There were no lineages with d_N/d_S>1, and d_N/d_S ratios did not vary among branches (P = 0.480, Figure S1). This is not unexpected, as averaging d_N/d_S across all sites is not a powerful test of adaptive evolution [39]. However, the nonsynonymous substitution rate in the C-terminal domain was significantly higher than in other domains. The cumulative d_N/d_S = 0.386 is higher than that in the entire coding region (d_N/d_S = 0.186), suggesting that different selection pressures have acted on amino acid changes across different functional regions of this gene.

The hypothesis of d_N/d_S homogeneity among branches was rejected for the C-terminal domain (P = 0.047, Figure 3), which could reflect either relaxed selective constraint or positive selection for amino acid substitution along one or more lineages. We were particularly interested in two lineages in the Hominoid clade leading to orangutan (d_N/d_S = infinity), as well as the common ancestor of bonobo and gorilla (d_N/d_S = 1.213), and tested whether the d_N/d_S ratio was significantly >1 on these two branches. Likelihood ratio tests (LRTs) from a two-ratio model revealed that

positive selection (d_N/d_S>1) had acted on the orangutan lineage although the statistical evidence was marginal (P = 0.087 and P = 0.090) (Table 2). However, the pattern of d_N/d_S heterogeneity across lineages is consistent with a relaxed selective constraint. The recently developed branch-site model A is powered to detect the particular amino acid sites that have been subject to positive selection in a given lineage (i.e., a foreground branch) [40,47], and at least five positively selected amino acids of the C-terminal domain existed in the lineage leading to orangutan (Table 3). No amino acids under positive selection were detected using site models (Table 3). It should be noted that the power of the LRTs is dependent on the number of coding sequences [48]. We sampled 14 primate species in Hominoid, Old World monkey, and New World monkey clades in our phylogenetic analyses (Table 1). A greater number of the species might have permitted more robust inferences of positive selection on the C-terminal domain of *PCSK9*.

The sliding-window analysis ratio further characterized the non-random nonsynonymous substitution along *PCSK9* and d_N/d_S peaks were obvious in the C-terminal domain (Figure 5). Although the d_N/d_S ratio in the lineage of humans is <1 (d_N/d_S = 0.566), three striking peaks of d_N/d_S (> 4) in human lineage were noted in the C-terminal domain. However, these peaks could be partly explained on the basis that human *PCSK9* shows very little synonymous divergence (d_S = 0.003).

In the present study, we calculated the d_N/d_S ratios across the phylogenetic tree using the 'gene' tree instead of the 'species' tree. We also performed analyses using the 'species' tree [49], in which bonobo is most closely related to chimpanzee and gorilla is sister to the human/chimpanzee clade. Although the log likelihood values (l_0 for one-ratio model and l_1 for free-ratio model) under the 'species' tree were different from that under the 'gene' tree, we did not find a significant difference in the d_N/d_S ratio between the 'species' tree and 'gene' tree for the coding regions and C-terminal domain. Although we did not observe d_N/d_S ratio >1 in the lineage leading to bonobo or gorilla in the C-terminal domain, we did note d_N/d_S = infinity in the lineage leading to orangutan. In addition, we used the branch-site A model to test for positive selection in the lineage leading to orangutan based on the species tree. The $2\Delta l$ between the null model (neutral, l = −1408.2) and the alternative model (selection, l = −1406.7) is 3.0 (df = 2, P<0.05) (critical value is 2.71 at 5% significance level [42]). We detected three codon sites under positive selection (544A, 551H, and 681S, P_b>95%) in the C-terminal domain of *PCSK9* (using BEB analysis). Zhang et al. [47] suggested a critical value of 3.84 (for P<0.05), however, such a threshold may be too conservative for a sequence length of 200 codons [47].

Structural and functional implication of PCSK9
The correct folding of the C-terminal domain is crucial for *PCSK9* function but catalytic activity is not required for *PCSK9* to bind and degrade LDL-R in cultured human hepatoma cells [50]. The C-terminal domain of a proprotein convertase contains unique sequences regulating their cellular localization and trafficking [51]. For example, *PCSK9* exhibits a Cys-His-rich domain that is required for cell surface binding in an LDL-R-dependent fashion [6] and plays a role in the regulation of auto-processing of *PCSK9*. The structural characteristic of C-terminal domain may determine the colocalization of *PCSK9* with LDLR at the cell surface [52] or lead to other novel functional properties. Hence, positive selection operating on the C-terminal domain was most likely directed at creating novel biochemical properties.

Species-specific differences in *PCSK9* expression patterns have been noted in brain and liver among humans, chimpanzee, and orangutan [53]. *PCSK9* is transiently expressed during embryonic

development in telencephalon and cerebellum where *LDLR* expression is not prominent [6,54]. Specific knockdown of *Pcsk9* mRNA led to embryonic death at 4 days after fertilization in zebrafish [54], and complete knockout of *Pcsk9* in mouse led to a ~50% reduction in circulating levels of LDL cholesterol, but did not result in a lethal phenotype [55]. Over-expression of *PCSK9* induces apoptosis in neural development [6,56], which results in a higher percentage of differentiated neurons and promotes cortical neurogenesis. These results indicate a novel function of *PCSK9* in central nervous system development, distinct from that in cholesterogenic organs such as liver [54]. One could hypothesize that relaxed selective constraint or positive selection has operated on the C-terminal domain of *PCSK9* due to the key role of *PCSK9* in early brain development.

Implications for human diseases

There is increasing interest in identifying gene loci affected by natural selection since they are medically important [57–63]. Loss-of-function or gain-of-function mutations in *PCSK9* have been reported to be associated with significant alterations in plasma levels of LDL cholesterol (Figure 1). Both evolutionary conservation indicating negative purifying selection and accelerated evolution driven by positive selection signify functionally significant regions of the genome [64]. To assess the potential severity of human *PCSK9* mutations, we assessed the levels of conservation or divergence of non-synonymous mutations listed in Figure 1 by aligning the amino acids among 14 species (Figure 2). We expected that the amino acids known to be important for *PCSK9* function (i.e., residues at which disease-causing mutations occur) would be highly conserved. All gain-of-function mutations in *PCSK9* leading to hypercholesterolemia in humans are 100% conserved at the amino acid level across all the primates we sampled. In case of loss-of-function mutations leading to hypocholesterolemia, all but two (E57K and Q554E) are 100% conserved across the primates. Mutations leading to both hypercholesterolemia and hypocholesterolemia appear to be less conserved, since four out of seven such mutations are not all conserved, including A53V, I474V, Q554E, and E670G. We noted a striking pattern of I474V variation (SNP rs562556) across the primates. The ancestral state of the 474th amino acid (M or V) in New World monkeys is not clear given the lack of an outgroup. The 'V' allele diverged to 'I' or 'V' in the Hominoid clade, suggesting a dynamic evolutionary history of the 474th amino acid. The human mutation I to V replicates the ancestral state, and the recurrence of this ancestral state has functional consequences [65].

To survey polymorphisms within human populations, we analyzed *PCSK9* SNPs using resequenced data from 24 African-Americans and 23 European-Americans (i.e., 47 individuals) in SeattleSNPs database (pga.gs.washington.edu/). A total of 229 polymorphic sites in African-Americans and 125 polymorphic sites in European-Americans were found in the human panel, eight of which resulted in amino acid changes. In addition, an in-frame insertion/deletion (CTG) in the signal-peptide domain was noted in both populations. Six of the eight non-synonymous sites are located in the putative domain and C-terminal domain (some investigators combine these two domains as 'C-terminal' domain), corresponding to the regions that have been predicted to be under positive selection. None of the non-synonymous sites was found in the catalytic domain. We used SIFT [66] and PolyPhen [67] to predict the effect of the amino acid changes (Table S2). In case of amino acid 474, the nonsynonymous substitution was predicted to be damaging (i.e., cause functional alteration), but the derived allele frequency of 0.79 in African-Americans and 0.87 in European-Americans suggests that positive selection acted to increase the frequency of this polymorphism. In humans,

a signature of recent positive selection was noted on this common variation using long-range haplotype (LRH) test; that is, positive selection had acted on the derived allele 'I' in African-Americans and the ancestral allele 'V' in European-Americans (Ding and Kullo, manuscript in revision). In addition, a signature of positive selection on the derived allele of E670G (rs505151), which resides in the C-terminal domain, was also noted in African-Americans. We speculate that non-conserved mutations across the primates might be the substrate for non-neutral evolution and responsible for the phenotypic variation in the general population.

In conclusion, phylogenetic analysis of the cholesterol metabolism gene *PCSK9* across a range of primates reveals lineage-specific patterns of variation. Although the gain-of-function mutations at *PCSK9* reflect strong functional constraint and a history of purifying selection, a signature of relaxed selective constraint or positive selection was noted in the C-terminal domain of *PCSK9*. It is possible that different modes of selection have operated on different functional domains of *PCSK9*.

MATERIALS AND METHODS

Primate Genomic DNA Sources

The comparative sequences of the *PCSK9* coding regions were obtained in 14 species from three sources. First, the human (accession no.: NM_174936) and chimpanzee (XM_427085) mRNA sequences of *PCSK9* were downloaded from NCBI (www.ncbi.nlm.nih.gov). Next, we acquired the BAC (bacterial artificial chromosome) clone sequence including *PCSK9* from Programs for Genomic Application (PGA) at Berkeley (pga.lbl.gov/seq), including colobus (AC188217), dusky titi (AC188268), squirrel monkey (AC188233), and marmoset (AC188221). Coding regions of *PCSK9* for these species were extracted by aligning the human mRNA sequence to the BAC sequence using the 'sim4' program [68]. Finally, DNA samples for a primate panel, including rhesus macaque, pigtailed macaque, bonobo, gorilla, chimpanzee, orangutan, tamarin, spider monkey, woolly monkey, and lemur, were obtained from Coriell Cell Repositories (Camden, NJ). The species name and scientific name for each species are listed in Table 1.

Sequencing of PCSK9 exons from Primate Genomic DNA

In the primate panel, *PCSK9* was amplified and sequenced exon by exon from genomic DNA with high fidelity polymerase chain reaction (PCR). Primers and PCR conditions are listed in Table S1. PCR products were sequenced directly in both forward and reverse directions. Exon reads were assembled together to create virtual transcript for each primate using the Sequencer® program (version 4.5, www.genecodes.com) and visually checked for accuracy. The lemur *PCSK9* sequence could not be obtained due to difficulty in PCR amplification. Sequences of coding regions for eight species in this primate panel were obtained. A total of 2072 bp of *PCSK9* coding sequence (the length is based on the human sequence and excludes the stop codon) in 14 species was compiled. All sequences have been submitted to the GenBank database under the accession nos. EF692496–EF692509 (Table 1).

Detecting lineage-specific episodes of positive selection

Sequences were aligned using ClustalW [69], followed by manual inspection and analysis. We used the 'HYPHY' package to estimate the topology of phylogenetic tree using the maximum likelihood method [34]. Since the gene tree was different from the species tree, analyses were done based on gene tree as well as the species tree.

We used the maximum likelihood method based on codon-substitution model by Yang [28,31,70] to test whether there was a significant difference in d_N/d_S ratio (i.e., ω) among lineages and whether d_N/d_S was significantly > 1 (i.e., positive selection) in a given lineage. The 'one-ratio' model assumes the same ratio for all branches in the phylogeny. The most general model – 'free-ratio' model – assumes an independent d_N/d_S ratio for each branch in the phylogeny. If there is a phylogenetic tree of many species, this model involves as many d_N/d_S parameters as the number of branches in the tree. The models used in the phylogenetic analysis can be compared using the likelihood-ratio test to examine interesting hypotheses [31]. The null hypothesis is the 'one-ratio' model, and can be used to test whether there is a differential d_N/d_S ratio among lineages. Positive selection or relaxed selective constraint in some lineages could contribute to the heterogeneity in the d_N/d_S ratio.

Detecting amino acid sites under positive selection

The above methods for lineage-specific selection assume that all amino acid sites have the same d_N/d_S ratio, i.e., averages the d_N/d_S ratio across all sites. Since many amino sites might be under strong purifying selection due to functional constraint ($d_N/d_S \approx$ zero) and positive selection often operates episodically on a few amino acid sites [47], it seems likely that this is a more conservative test and amino acid sites under positive selection cannot be detected.

Several methods have been developed to address this problem, such as the site models which allow d_N/d_S to vary among codons [38,71]. In the present study, we also used an improved branch-site likelihood method to detect positive selection at the amino acid sites [28,40,47]. This branch-site model [28,40,47] assumed that the branches on the phylogeny are divided *a priori* into foreground (i.e, may have experienced positive selection) and background lineages. We used the likelihood-ratio test 2 (i.e., the branch-site test of positive selection) constructed from this branch-site model [47]. The null hypothesis of this LRT is the branch-site model A list above but with ω_2 fixed $= 1$, which can be used to directly test for positive selection on the foreground lineages [47]. The Bayes empirical Bayes (BEB) approach was used to calculate the posterior probabilities that a codon belongs to the site class of positive selection on the foreground lineages [40]. The test should be compared with the 50:50 mixture of point mass 0 and χ^2_1 (with critical values to be 2.71 and 5.41, at the 5% and 1% significance levels, respectively) [42]. Zhang et al [47] also suggested the use of χ^2_1 distribution for assessing the significance of the test (3.84 and 5.99 at the 5% and 1% significance levels, respectively). This LRT

test seemed conservative overall, but exhibited better power in detecting positive selection than the branch-based test [47].

We used the 'CODEML' program in PAML version 3.15 [72] to calculate the d_N/d_S ratio and perform the maximum likelihood phylogenetic analysis. To calculate the d_N/d_S ratio at lineages (defined as all branches in the phylogeny, both terminal species nodes and internodes), sequences associated with species-specific premature stop codons were removed.

Sliding-window analysis of d_N/d_S

Sliding-window analysis of d_N/d_S was performed with a window size of 90 bp (30 codons) and a sliding increment of 15 bp (5 codons). We used the approach by Choi and Lahn [73] to calculate the d_N/d_S of each window as the ratio between window-specific d_N and gene-average d_S, since noise in window-specific d_S can sometimes hamper the analysis. In addition, the use of gene-average instead of window-specific d_S should not introduce any systematic bias [73].

SUPPORTING INFORMATION

Figure S1 Phylogeny of coding regions of *PCSK9*. *PCSK9* was resequenced from a panel of primates including Hominoids, Old World monkeys, and New World monkeys. Branch lengths were estimated by maximum likelihood under the free-ratio model, which assumes an independent d_N/d_S ratio for each branch.
Found at: doi:10.1371/journal.pone.0001098.s001 (0.01 MB EPS)

Table S1 Polymerase chain reaction (PCR) primers and conditions for *PCSK9* exon analyses
Found at: doi:10.1371/journal.pone.0001098.s002 (0.10 MB DOC)

Table S2 SIFT and Polyphen prediction of amino acid polymorphisms
Found at: doi:10.1371/journal.pone.0001098.s003 (0.04 MB DOC)

Author Contributions

Conceived and designed the experiments: KD IK. Performed the experiments: SM. Analyzed the data: KD IK. Wrote the paper: KD IK.

REFERENCES

1. Espenshade PJ, Cheng D, Goldstein JL, Brown MS (1999) Autocatalytic processing of site-1 protease removes propeptide and permits cleavage of sterol regulatory element-binding proteins. J Biol Chem 274: 22795–22804.
2. Brown MS, Goldstein JL (1999) A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. Proc Natl Acad Sci U S A 96: 11041–11048.
3. Elagoz A, Benjannet S, Mammarbassi A, Wickham L, Seidah NG (2002) Biosynthesis and cellular trafficking of the convertase SKI-1/S1P: ectodomain shedding requires SKI-1 activity. J Biol Chem 277: 11265–11275.
4. Benjannet S, Rhainds D, Essalmani R, Mayne J, Wickham L, et al. (2004) NARC-1/PCSK9 and its natural mutants: zymogen cleavage and effects on the low density lipoprotein (LDL) receptor and LDL cholesterol. J Biol Chem 279: 48865–48875.
5. Naureckiene S, Ma L, Sreekumar K, Purandare U, Lo CF, et al. (2003) Functional characterization of Narc 1, a novel proteinase related to proteinase K. Arch Biochem Biophys 420: 55–67.
6. Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, et al. (2003) The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. Proc Natl Acad Sci U S A 100: 928–933.

7. Lalanne F, Lambert G, Amar MJ, Chetiveaux M, Zair Y, et al. (2005) Wild-type PCSK9 inhibits LDL clearance but does not affect apoB-containing lipoprotein production in mouse and cultured cells. J Lipid Res 46: 1312–1319.
8. Maxwell KN, Breslow JL (2004) Adenoviral-mediated expression of Pcsk9 in mice results in a low-density lipoprotein receptor knockout phenotype. Proc Natl Acad Sci U S A 101: 7100–7105.
9. Park SW, Moon YA, Horton JD (2004) Post-transcriptional regulation of low density lipoprotein receptor protein by proprotein convertase subtilisin/kexin type 9a in mouse liver. J Biol Chem 279: 50630–50638.
10. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 34: 154–156.
11. Leren TP (2004) Mutations in the PCSK9 gene in Norwegian subjects with autosomal dominant hypercholesterolemia. Clin Genet 65: 419–422.
12. Timms KM, Wagner S, Samuels ME, Forbey K, Goldfine H, et al. (2004) A mutation in PCSK9 causing autosomal-dominant hypercholesterolemia in a Utah pedigree. Hum Genet 114: 349–353.
13. Sun XM, Eden ER, Tosi I, Neuwirth CK, Wile D, et al. (2005) Evidence for effect of mutant PCSK9 on apolipoprotein B secretion as the cause of unusually severe dominant hypercholesterolaemia. Hum Mol Genet 14: 1161–1169.

14. Naoumova RP, Tosi I, Patel D, Neuwirth C, Horswell SD, et al. (2005) Severe hypercholesterolemia in four British families with the D374Y mutation in the PCSK9 gene: long-term follow-up and treatment response. Arterioscler Thromb Vasc Biol 25: 2654–2660.
15. Shioji K, Mannami T, Kokubo Y, Inamoto N, Takagi S, et al. (2004) Genetic variants in PCSK9 affect the cholesterol level in Japanese. J Hum Genet 49: 109–114.
16. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, et al. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37: 161–165.
17. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, et al. (2006) A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am J Hum Genet 78: 410–422.
18. Berge KE, Ose L, Leren TP (2006) Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy. Arterioscler Thromb Vasc Biol 26: 1094–1100.
19. Yue P, Averna M, Lin X, Schonfeld G (2006) The c.43_44insCTG variation in PCSK9 is associated with low plasma LDL-cholesterol in a Caucasian population. Hum Mutat 27: 460–466.
20. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 354: 1264–1272.
21. Fleagle JG, McGraw WS (1999) Skeletal and dental morphology supports diphyletic origin of baboons and mandrills. Proc Natl Acad Sci U S A 96: 1157–1161.
22. Crissey S, Barr J, Slifka K, Bowen P, Stacewicz-Sapuntzakis M, et al. (1999) Serum concentrations of lipids, vitamins A and E, vitamin D metabolites, and carotenoids in nine primate species at four zoos. Zoo biology 18: 551–564.
23. Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, et al. (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci U S A 100: 13030–13035.
24. Horton JD, Goldstein JL, Brown MS (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest 109: 1125–1131.
25. Tall AR (2006) Protease variants, LDL, and coronary heart disease. N Engl J Med 354: 1310–1312.
26. Attie AD, Seidah NG (2005) Dual regulation of the LDL receptor–some clarity and new questions. Cell Metab 1: 290–292.
27. Seidah NG, Khatib AM, Prat A (2006) The proprotein convertases and their implication in sterol and/or lipid metabolism. Biol Chem 387: 871–877.
28. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19: 908–917.
29. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford: Oxford University Press.
30. Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. Nature 385: 151–154.
31. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15: 568–573.
32. Huttley GA, Easteal S, Southey MC, Tesoriero A, Giles GG, et al. (2000) Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Australian Breast Cancer Family Study. Nat Genet 25: 410–413.
33. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, et al. (2006) Positive natural selection in the human lineage. Science 312: 1614–1620.
34. Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676–679.
35. Lynn DJ, Freeman AR, Murray C, Bradley DG (2005) A genomics approach to the detection of positive selection in cattle: adaptive evolution of the T-cell and natural killer cell-surface protein CD2. Genetics 170: 1189–1196.
36. Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18: 486.
37. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, et al. (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9: 585–598.
38. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929–936.
39. Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155: 431–449.
40. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107–1118.
41. Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168: 1041–1051.
42. Self S, Liang K (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J Am Stat Assoc 82: 605–610.
43. Horton JD, Cohen JC, Hobbs HH (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 32: 71–77.
44. Mbikay M, Mayne J, Seidah NG, Chretien M (2007) Of PCSK9, cholesterol homeostasis and parasitic infections: Possible survival benefits of loss-of-function PCSK9 genetic polymorphisms. Med Hypotheses.
45. Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64: 18–23.
46. Perry GH, Tito RY, Verrelli BC (2007) The evolutionary history of human and chimpanzee Y-chromosome gene loss. Mol Biol Evol 24: 853–859.
47. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22: 2472–2479.
48. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18: 1585–1592.
49. Purvis A (1995) A composite estimate of primate phylogeny. Philos Trans R Soc Lond B Biol Sci 348: 405–421.
50. McNutt MC, Lagace TA, Horton JD (2007) Catalytic activity is not required for secreted PCSK9 to reduce LDL receptors in HepG2 cells. J Biol Chem 282: 20799–20803.
51. Seidah NG, Prat A (2007) The proprotein convertases are potential targets in the treatment of dyslipidemia. J Mol Med 85: 685–696.
52. Nassoury N, Blasiole DA, Tebon Oler A, Benjannet S, Hamelin J, et al. (2007) The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependence on the LDLR. Traffic 8: 718–732.
53. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. (2002) Intra- and interspecific variation in primate gene expression patterns. Science 296: 340–343.
54. Poirier S, Prat A, Marcinkiewicz E, Paquin J, Chitramuthu BP, et al. (2006) Implication of the proprotein convertase NARC-1/PCSK9 in the development of the nervous system. J Neurochem 98: 838–850.
55. Rashid S, Curtis DE, Garuti R, Anderson NN, Bashmakov Y, et al. (2005) Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. Proc Natl Acad Sci U S A 102: 5374–5379.
56. Bingham B, Shen R, Kotnis S, Lo CF, Ozenberger BA, et al. (2006) Proapoptotic effects of NARC 1 (= PCSK9), the gene encoding a novel serine proteinase. Cytometry A 69: 1123–1131.
57. Hahn MW, Rockman MV, Soranzo N, Goldstein DB, Wray GA (2004) Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the factor VII locus in humans. Genetics 167: 867–877.
58. Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, et al. (2004) Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. Am J Hum Genet 74: 898–916.
59. Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, et al. (2004) CYP3A variation and the evolution of salt-sensitivity variants. Am J Hum Genet 75: 1059–1069.
60. Rockman MV, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, et al. (2004) Positive selection on MMP3 regulation has shaped heart disease risk. Curr Biol 14: 1531–1539.
61. Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, et al. (2005) Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. PLoS Genet 1: e82.
62. Ding K, Kullo IJ (2006) Molecular evolution of 5′ flanking regions of 87 candidate genes for atherosclerotic cardiovascular disease. Genet Epidemiol 30: 557–569.
63. Kullo IJ, Ding K (2007) Patterns of population differentiation of candidate genes for cardiovascular disease. BMC genetics 8: 48.
64. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424: 788–793.
65. Schaner P, Richards N, Wadhwa A, Aksentijevich I, Kastner D, et al. (2001) Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. Nat Genet 27: 318–321.
66. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31: 3812–3814.
67. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30: 3894–3900.
68. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res 8: 967–974.
69. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
70. Yang Z, Goldman N, Friday A (1994) Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Mol Biol Evol 11: 316–324.
71. Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 15: 496–503.
72. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556.
73. Choi SS, Lahn BT (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. Genome Res 13: 2252–2259.