RESEARCH ARTICLE

Open Access

Molecular evolutionary analysis of human primary microcephaly genes

Nashaiman Pervaiz[1], Hongen Kang[2], Yiming Bao[2*] and Amir Ali Abbasi[1*]

Abstract

Background: There has been a rapid increase in the brain size relative to body size during mammalian evolutionary history. In particular, the enlarged and globular brain is the most distinctive anatomical feature of modern humans that set us apart from other extinct and extant primate species. Genetic basis of large brain size in modern humans has largely remained enigmatic. Genes associated with the pathological reduction of brain size (primary microcephaly-MCPH) have the characteristics and functions to be considered ideal candidates to unravel the genetic basis of evolutionary enlargement of human brain size. For instance, the brain size of microcephaly patients is similar to the brain size of *Pan troglodyte* and the very early hominids like the *Sahelanthropus tchadensis* and *Australopithecus afarensis*.

Results: The present study investigates the molecular evolutionary history of subset of autosomal recessive primary microcephaly (MCPH) genes; CEP135, ZNF335, PHC1, SASS6, CDK6, MFSD2A, CIT, and KIF14 across 48 mammalian species. Codon based substitutions site analysis indicated that ZNF335, SASS6, CIT, and KIF14 have experienced positive selection in eutherian evolutionary history. Estimation of divergent selection pressure revealed that almost all of the MCPH genes analyzed in the present study have maintained their functions throughout the history of placental mammals. Contrary to our expectations, human-specific adoptive evolution was not detected for any of the MCPH genes analyzed in the present study.

Conclusion: Based on these data it can be inferred that protein-coding sequence of MCPH genes might not be the sole determinant of increase in relative brain size during primate evolutionary history.

Keywords: Molecular evolution, Brain, Primary microcephaly, MCPH, Primates, Positive selection

Background

The enlarged and globular brain distinguish modern humans not only from extant non-human primates but also from their extinct Homo relatives [1]. The modern human brain is approximately three-fold larger in size than that of our closest extant relative, the chimpanzee, and extinct early hominids. From developmental perspective, the larger brain size in humans is attributed to human-specific fast and prolonged neonatal and postnatal brain growth and patterning [2–4]. The evolutionary expansion of human brain size is heterogeneous across brain regions. For instance, the most notable expansion occurred in the neocortex that has been directly related to the emergence of higher cognitive capabilities, such as language, intelligence, and social learning [5]. Most of the expansion in brain size occurred in the last 2–3 million years of human evolution [6, 7]. The genetic basis of divergence between the highly cognitive human brain and supposedly lesser cognitive non-human primate brain has largely remained enigmatic. It has been speculated that complexity of modern human brain arose through

*Correspondence: baoym@big.ac.cn; abbasiam@qau.edu.pk
[1] National Center for Bioinformatics, Program of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2] China National Center for Bioinformation and National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 2 of 9

changes in protein-coding genes and non-coding regulatory elements [8]. In particular, in relation to the evolutionary expansion of cerebral cortex, genes associated with primary microcephaly (MCPH) have been the focus of immense attention in the past couple of decades [9]. Primary microcephaly is an autosomal congenital disorder, which is characterized by significant reduction in brain mass, particularly the cerebral cortex with no other neuroanatomical abnormalities [10]. The brain size of microcephaly patients is similar to that of very primitive extinct hominids *Sahelanthropus tchadensis* (brain size; 370 cm^3) and *Australopithecus afarensis* (brain size; 450 cm^3). Therefore, the small brain size in microcephaly patients is considered as an example of evolutionary retrogression [7, 11]. Furthermore, microcephaly patients have mild to severe cognitive impairment [10]. MCPH is a heterogeneous disorder and associated with mutations in at least sixteen gene loci (MCPH1-16) [12]. The underlying MCPH genes are named as *MCPH1, WDR62, CDK5RAP2, CASC5, ASPM, CENPJ, STIL, CEP135, CEP152, ZNF335, PHC1, CDK6, SASS6, MFSD2A, CIT, and KIF14* [12–27]. Almost all MCPH-associated genes express predominantly in the fetal brain and regulate neurogenesis and brain size through participation in several important cellular mechanisms including DNA replication, repair, cytokinesis, intracellular transport and autophagy [27]. Therefore, investigation of MCPH genes can reveal molecular mechanisms that are critical in the regulation of brain size and complexity [28, 29]. Indeed, many previously reported molecular evolutionary investigations have implicated MCPH genes (like *ASPM* and *CDK5RAP2*) in the expansion and reduction of brain size during primate history [30, 31].

In this study, the roles of MCPH genes in the evolutionary enlargement of human brain size was explored through molecular evolutionary analysis of eight newly identified microcephaly genes; *CEP135, ZNF335, PHC1, CDK6, SASS6, MFSD2A, CIT,* and *KIF14*. Sequence data from 48 eutherian species was employed to investigate the signatures of episodic positive selection and diversifying selection during mammalian evolutionary history. Results obtained in the present study provide a broader perspective on the evolutionary link between primary microcephaly genes and human brain size.

Results

Molecular evolutionary analysis of MCPH genes
In order to investigate the genetic basis of the evolutionary expansion of human brain, eight newly identified *MCPH* genes (*CEP135, ZNF335, PHC1, CDK6, SASS6, MFSD2A, CIT,* and *KIF14*) were considered as candidates for evolutionary analysis. We assembled the coding sequence for each of this selected subset of MCPH genes

from a broad range of eutherian genomes which include 21 primatomorphans, 9 glirans, 6 carnivores, 2 perissodactylans, 5 cetartiodactylans, 3 chiropterans, and one animal each from Eulipotyphla and Proboscidea (Additional file 1: Figure S1). These 48 eutherian genomes provide enough genomic coverage and evolutionary breadth to perform a thorough molecular evolutionary genetic analysis. For each of these selected subsets of MCPH genes, orthologous sequence information was classified into three data sets. i.e., primates (20 genomes), nonprimate placental mammals (28 genomes), and placental mammals (48 genomes). Each of these selected subsets of datasets is subjected to maximum likelihood based codon substitution models.

Signature of pervasive positive selection in MCPH genes
In order to examine whether the pervasive positive selection has operated on selected subset of MCPH genes (*CEP135, ZNF335, PHC1, CDK6, SASS6, MFSD2A, CIT*, and *KIF14*), three pairs of site models (M1 & M2, M7 & M8, and M8a & M8) based on codon substitutions were used. These codon substitutions site models assume that selective pressure ω vary among the amino acid sites but not across the lineage [32]. The signature of positive selection is considered optimal only if two out of the three null models (M1, M7, and M8a) are rejected in the favor of more complex alternative models (M2 and M8). The results of codon substitutions site pair models (M1/M2, M7/M8, and M8a/M8) revealed the consistent signature of pervasive positive selection only for *KIF14* gene in primates (Table 1 and Additional file 2: Table S1). Furthermore, significant signatures of positive selection were also detected in non-primate placental mammals and placental mammals for three (*ZNF335, SASS6,* and *CIT*) and two MCPH genes (*ZNF335*, and *KIF14*) respectively (Table 1 and Additional file 2: Table S1). Positively selected codon sites were identified by using the Bayes Empirical Bayes (BEB) method implemented in M8 codon substitution site model [33] (Table 2).

Signatures of episodic positive selection
We next examined the imprints of transient or episodic positive selection in different lineages from ancestral primate branch to human terminal branch by employing branch site model (Additional file 2: Table S2). Branch site model allows the ratio of nonsynonymous (dN) to synonymous substitution (dS) rates ω to vary not only across the branches in the phylogeny but also across sites [34]. The significance of the transient imprint of adaptive evolution was determined by likelihood ratio tests (LRTs) of null model (similar to branch-site model except ω2 is restricted to one for the predefined lineage of interest) against branch-site model through log likelihood score

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 3 of 9

Table 1 Signature of pervasive positive selection in primary microcephaly genes

Genes	Data sets	No. of sequences	(M2-M1)			(M8-M7)			(M8-M8a)		
			LRT	P value	q Value	LRT	P value	q Value	LRT	P value	q Value
ZNF335	Nonprimate mammals	27	0	1	1	24.13	0.000005	0.000013	9.34	0.002	0.008
	Mammals	45	0	1	1	27.41	0.000001	0.000005	18.62	0.00002	0.0002
SASS6	Nonprimate mammals	25	0	1	1	21.58	0.00002	0.00004	7.196	0.007	0.02
CIT	Nonprimate mammals	22	0	1	1	38.84	< 0.0000001	< 0.0000001	18.21	0.00002	0.0002
	Mammals	37	0.000002	1	1	24.4	0.000005	0.00001	0.002	0.96	0.80
KIF14	Primates	17	13.46	0.002	0.00	20.76	0.00003	0.0006	11.63	0.0006	0.004
	Mammals	29	0	1	1	26.67	0.0000001	0.000005	10.36	0.001	0.005

Positive selection is inferred if two out of three site model comparisons significantly reject null hypothesis using the cutoff of q value 0.05. Null models: M1, M7, and M8a, Alternative models: M2 and M8, LRT: likelihood ratio test. False discovery rate q value corrections over p values were calculated using q value package in R

Table 2 Positive selected sites detected by using M8 codon substitution site model

Gene	Data sets	No. of codon under positive selection	Sites under positive selection (BEB > 80%)
ZNF335	Non-primate mammals	7	225A[α], 365P, 414G, 761S, 830P, 983V[α], 1133S
	Mammals	7	93H, 225A[α], 738A, 761S, 830P, 938V[β], 1133S
SASS6	Non-primate mammals	7	99A, 241I, 494 T, 524S, 550 T, 581I[α], 585C
CIT	Non-primate mammals	8	11P, 12L[α], 17A, 150R, 279T[β], 334F, 1479 T, 1723I
KIF14	Primates	12	107L, 203P, 204S, 342 V, 875H[β], 1413F, 1474S, 1499R[α], 1532Q, 1602 K, 1605H, 1612G
	Mammals	10	29A, 107L, 166I, 203P, 204S, 232 T, 1303I, 1525H, 1608S[α], 1632S

Positively selected sites are identifies using Bayes Empirical Bayes (BEB) analysis implemented in M8 codon substitution site model with cutoff of posterior probability > 0.80%. Sites with P value < = 0.05 are labeled as α and with P value < = 0.01 are labeled as β

for each model. False positive results obtained by branch-site model were eliminated by estimating the false discovery rate q value. Branch site calculations revealed no significant patterns of episodic positive selection in any lineage from primate ancestor to human terminal branch for all MCPH genes analyzed except for KIF14 in homininae ancestral branch (Additional file 2: Table S2).

Divergent selection of MCPH genes across mammals

Functional divergence among orthologous proteins during evolution may not necessarily be reflected as a signature of positive selection at the sequence level. Instead, sequence evolutionary rate variations among different clades of phylogeny can also be taken as a metric of adaptive functional diversity. For each of the eight MCPH coding regions, we estimated the divergent selective pressure between different partitions of mammalian phylogeny (primate vs. nonprimate, simians vs. nonsimians, catarrhini vs. noncatarrhini, hominidae vs. nonhominidae and hominini vs. nonhominini placental mammals) by employing clade model c (CmC) (Table 3). Similar to site and branch-site analysis, false positive data in CmC analysis was eradicated by q value [35]. The CmC analysis

revealed that selective pressure on different parts of phylogeny is not significantly divergent for all analyzed microcephaly loci except for *SASS6* in the comparison of simians and nonsimians placental mammals (Table 3). In this particular case of *SASS6*, approximately 34% of sites evolved under divergent selective pressure between simians ($\omega s = 0.499$) and nonsimians placental mammals ($\omega ns = 0.214$) (Table 3).

Discussion

Compared to other primates, including great apes, humans have very large brains. Weighing approximately 1,400 g, our brains are roughly three times larger than those of other great apes such as chimpanzees (395 g) and gorillas (490 g) [36]. In particular, during 4–5 million years of human evolution, an enormous increase in brain size has occurred, from a brain mass of 450 g found in Australopithecines to about 1400 g in modern Homo sapiens [37]. Similar to overall increase in brain size/mass, the neocortex of brain (which is involved in higher-order brain functions such as cognition, reasoning and language) has significantly enlarged in the hominin lineage after the divergence of closely

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 4 of 9

Table 3 Divergent selection constraint parameters estimation and likelihood scores for eight MCPH genes on different partition of mammalian phylogeny

Gene	Model	Parameter estimation (ω)	lnL	P value	q value
CEP135	CmC-Primate	$p_0 = 0.560$, $\omega_0 = 0.0231$, $p_1 = 0.142$, $\omega_1 = 1.0$, $p_2 = 0.298$, $\omega_{np} = 0.251$, $\omega_p = 0.315$	$-26{,}080.0893$	0.057	0.34
	CmC-Simians	$p_0 = 0.296$, $\omega_0 = 0.27$, $p_1 = 0.141$, $\omega_1 = 1.0$, $p_2 = 0.56$, $\omega_{ns} = 0.023$, $\omega_s = 0.021$	$-26{,}081.8471$	0.75	0.66
	CmC-catarrhini	$p_0 = 0.293$, $\omega_0 = 0.271$, $p_1 = 0.141$, $\omega_1 = 1.0$, $p_2 = 0.57$, $\omega_{nc} = 0.023$, $\omega_c = 0.0397$	$-26{,}081.3823$	0.31	0.58
	CmC-greatapes	$p_0 = 0.295$, $\omega_0 = 0.269$, $p_1 = 0.141$, $\omega_1 = 1.0$, $p_2 = 0.564$, $\omega_{ng} = 0.0235$, $\omega_g = 0.0184$	$-26{,}081.8801$	0.85	0.66
	CmC-hominini	$p_0 = 0.56$, $\omega_0 = 0.023$ $p_1 = 0.141$, $\omega_1 = 1.0$, $p_2 = 0.296$, $\omega_{nh} = 0.27$, $\omega_h = 0.59$	$-26{,}081.1383$	0.22	0.57
	M2a_rel	$p_0 = 0.295$, $\omega_0 = 0.269$, $p_1 = 0.0141$, $\omega_1 = 1.0$, $p_2 = 0.564$, $\omega_2 = 0.0235$	$-26{,}081.8982$	NA	NA
ZNF335	CmC-Primate	$p_0 = 0.75$, $\omega_0 = 0.015$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.221$, $\omega_{np} = 0.255$, $\omega_p = 0.285$	$-32{,}294.0420$	0.30	0.58
	CmC-simians	$p_0 = 0.751$, $\omega_0 = 0.0155$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.22$, $\omega_{ns} = 0.256$, $\omega_s = 0.312$	$-32{,}293.5981$	0.16	0.57
	CmC-catarrhini	$p_0 = 0.751$, $\omega_0 = 0.0154$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.22$, $\omega_{nc} = 0.266$, $\omega_c = 0.27$	$-32{,}293.5586$	0.15	0.57
	CmC-greatapes	$p_0 = 0.75$, $\omega_0 = 0.0154$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.22$, $\omega_{ng} = 0.259$, $\omega_g = 0.297$	$-32{,}294.4826$	0.66	0.66
	CmC-hominini	$p_0 = 0.75$, $\omega_0 = 0.0154$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.22$, $\omega_{nh} = 0.26$, $\omega_h = 0.25$	$-32{,}294.5720$	0.92	0.69
	M2a_rel	$p_0 = 0.75$, $\omega_0 = 0.015$, $p_1 = 0.029$, $\omega_1 = 1.0$, $p_2 = 0.221$, $\omega_2 = 0.26$	$-32{,}294.5769$	NA	NA
PHC1	CmC-Primate	$p_0 = 0.784$, $\omega_0 = 0.020$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.195$, $\omega_{np} = 0.339$, $\omega_p = 0.395$	$-20{,}737.4013$	0.29	0.58
	CmC-simians	$p_0 = 0.194$, $\omega_0 = 0.348$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.78$, $\omega_{ns} = 0.020$, $\omega_s = 0.018$	$-20{,}737.9248$	0.82	0.66
	CmC-catarrhini	$p_0 = 0.78$, $\omega_0 = 0.020$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.19$, $\omega_{nc} = 0.347$, $\omega_c = 0.364$	$-20{,}737.9367$	0.86	0.66
	CmC-greatapes	$p_0 = 0.78$, $\omega_0 = 0.020$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.19$, $\omega_{ng} = 0.349$, $\omega_g = 0.179$	$-20{,}737.3879$	0.29	0.58
	CmC-hominini	$p_0 = 0.78$, $\omega_0 = 0.020$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.19$, $\omega_{nh} = 0.35$, $\omega_h = 0.13$	$-20{,}737.5787$	0.38	0.65
	M2a_rel	$p_0 = 0.784$, $\omega_0 = 0.020$, $p_1 = 0.021$, $\omega_1 = 1.0$, $p_2 = 0.194$, $\omega_2 = 0.348$	$-20{,}737.9512$	NA	NA
SASS6	CmC-Primate	$p_0 = 0.612$, $\omega_0 = 0.017$, $p_1 = 0.077$, $\omega_1 = 1.0$, $p_2 = 0.310$, $\omega_{np} = 0.253$, $\omega_p = 0.270$	$-14{,}706.9216$	0.67	0.66
	CmC-simians	$p_0 = 0.58$, $\omega_0 = 0.013$, $p_1 = 0.082$, $\omega_1 = 1.0$, $p_2 = 0.34$, $\omega_{ns} = 0.214$, $\omega_s = 0.499$	$-14{,}697.4440$	**0.00001**	**0.0003**
	CmC-catarrhini	$p_0 = 0.31$, $\omega_0 = 0.26$, $p_1 = 0.077$, $\omega_1 = 1.0$, $p_2 = 0.62$, $\omega_{nc} = 0.017$, $\omega_c = 0.024$	$-14{,}706.9516$	0.72	0.66
	CmC-greatapes	$p_0 = 0.61$, $\omega_0 = 0.17$, $p_1 = 0.078$, $\omega_1 = 1.0$, $p_2 = 0.31$, $\omega_{ng} = 0.25$, $\omega_g = 0.47$	$-14{,}706.3050$	0.23	0.57
	CmC-hominini	$p_0 = 0.62$, $\omega_0 = 0.017$, $p_1 = 0.077$, $\omega_1 = 1.0$, $p_2 = 0.31$, $\omega_{nh} = 0.26$, $\omega_h = 1.082$	$-14{,}705.0071$	**0.045**	0.33
	M2a_rel	$p_0 = 0.614$, $\omega_0 = 0.017$, $p_1 = 0.077$, $\omega_1 = 1.0$, $p_2 = 0.308$, $\omega_2 = 0.259$	$-14{,}707.0141$	NA	NA
CDK6	CmC-Primate	$p_0 = 0.895$, $\omega_0 = 0.011$, $p_1 = 0.012$, $\omega_1 = 1.0$, $p_2 = 0.0926$, $\omega_{np} = 0.298$, $\omega_p = 0.257$	-5156.14749	0.76	0.66
	CmC-Simians	$p_0 = 0.897$, $\omega_0 = 0.012$, $p_1 = 0.014$, $\omega_1 = 1.0$, $p_2 = 0.088$, $\omega_{ns} = 0.271$, $\omega_s = 0.428$	-5155.82423	0.39	0.65
	CmC-catarrhini	$p_0 = 0.902$, $\omega_0 = 0.012$, $p_1 = 0.011$, $\omega_1 = 1.0$, $p_2 = 0.087$, $\omega_{nc} = 0.32$, $\omega_c = 0.16$	-5156.01860	0.55	0.66
	CmC-greatapes	$p_0 = 0.089$, $\omega_0 = 0.298$, $p_1 = 0.012$, $\omega_1 = 1.0$, $p_2 = 0.899$, $\omega_{ng} = 0.012$, $\omega_g = 0.00$	-5156.08170	0.63	0.66
	CmC-hominini	$p_0 = 0.089$, $\omega_0 = 0.296$, $p_1 = 0.012$, $\omega_1 = 1.0$, $p_2 = 0.898$, $\omega_{nh} = 0.012$, $\omega_h = 0.00$	-5156.12730	0.71	0.66
	M2a_rel	$p_0 = 0.897$, $\omega_0 = 0.011$, $p_1 = 0.012$, $\omega_1 = 1.0$, $p_2 = 0.0906$, $\omega_2 = 0.294$	-5156.19494	NA	NA
MFSD2A	CmC-Primate	$p_0 = 0.731$, $\omega_0 = 0.0152$, $p_1 = 0.062$, $\omega_1 = 1.0$, $p_2 = 0.207$, $\omega_{np} = 0.257$, $\omega_p = 0.279$	$-11{,}299.7769$	0.66	0.66
	CmC-simians	$p_0 = 0.719$, $\omega_0 = 0.014$, $p_1 = 0.066$, $\omega_1 = 1.0$, $p_2 = 0.21$, $\omega_{ns} = 0.23$, $\omega_s = 0.38$	$-11{,}297.2803$	**0.02**	0.30
	CmC-catarrhini	$p_0 = 0.725$, $\omega_0 = 0.015$, $p_1 = 0.065$, $\omega_1 = 1.0$, $p_2 = 0.21$, $\omega_{nc} = 0.25$, $\omega_c = 0.33$	$-11{,}299.5559$	0.43	0.66
	CmC-greatapes	$p_0 = 0.73$, $\omega_0 = 0.015$, $p_1 = 0.062$, $\omega_1 = 1.0$, $p_2 = 0.21$, $\omega_{ng} = 0.262$, $\omega_g = 0.289$	$-11{,}299.8560$	0.85	0.66
	CmC-hominini	$p_0 = 0.73$, $\omega_0 = 0.015$, $p_1 = 0.061$, $\omega_1 = 1.0$, $p_2 = 0.21$, $\omega_{nh} = 0.266$, $\omega_h = 0.196$	$-11{,}299.8222$	0.75	0.66
	M2a_rel	$p_0 = 0.732$, $\omega_0 = 0.0154$, $p_1 = 0.061$, $\omega_1 = 1.0$, $p_2 = 0.206$, $\omega_2 = 0.265$	$-11{,}299.8738$	NA	NA
CIT	CmC-Primate	$p_0 = 0.89$, $\omega_0 = 0.0058$, $p_1 = 0.014$, $\omega_1 = 1.0$, $p_2 = 0.095$, $\omega_{np} = 0.24$, $\omega_p = 0.19$	$-35{,}246.2438$	0.18	0.57
	CmC-simians	$p_0 = 0.095$, $\omega_0 = 0.226$, $p_1 = 0.015$, $\omega_1 = 1.0$, $p_2 = 0.89$, $\omega_{ns} = 0.0059$, $\omega_s = 0.0043$	$-35{,}246.9740$	0.57	0.66
	CmC-catarrhini	$p_0 = 0.094$, $\omega_0 = 0.227$, $p_1 = 0.015$, $\omega_1 = 1.0$, $p_2 = 0.89$, $\omega_{nc} = 0.0057$, $\omega_c = 0.0088$	$-35{,}246.8939$	0.49	0.66
	CmC-greatapes	$p_0 = 0.89$, $\omega_0 = 0.0058$, $p_1 = 0.015$, $\omega_1 = 1.0$, $p_2 = 0.094$, $\omega_{ng} = 0.2279$, $\omega_g = 0.2074$	$-35{,}247.1167$	0.86	0.66
	CmC-hominini	$p_0 = 0.89$, $\omega_0 = 0.0057$, $p_1 = 0.015$, $\omega_1 = 1.0$, $p_2 = 0.095$, $\omega_{nh} = 0.224$, $\omega_h = 0.379$	$-35{,}246.9556$	0.55	0.66
	M2a_rel	$p_0 = 0.89$, $\omega_0 = 0.0058$, $p_1 = 0.015$, $\omega_1 = 1.0$, $p_2 = 0.095$, $\omega_2 = 0.227$	$-35{,}247.1321$	NA	NA
KIF14	CmC-Primate	$p_0 = 0.43$, $\omega_0 = 0.043$, $p_1 = 0.15$, $\omega_1 = 1.0$, $p_2 = 0.42$, $\omega_{np} = 0.36$, $\omega_p = 0.39$	$-31{,}056.0434$	0.58	0.66
	CmC-simians	$p_0 = 0.43$, $\omega_0 = 0.043$, $p_1 = 0.14$, $\omega_1 = 1.0$, $p_2 = 0.42$, $\omega_{ns} = 0.36$, $\omega_s = 0.42$	$-31{,}055.4805$	0.23	0.57
	CmC-catarrhini	$p_0 = 0.43$, $\omega_0 = 0.042$, $p_1 = 0.15$, $\omega_1 = 1.0$, $p_2 = 0.42$, $\omega_{nc} = 0.36$, $\omega_c = 0.41$	$-31{,}055.9687$	0.50	0.66
	CmC-greatapes	$p_0 = 0.42$, $\omega_0 = 0.37$, $p_1 = 0.15$, $\omega_1 = 1.0$, $p_2 = 0.43$, $\omega_{ng} = 0.042$, $\omega_g = 0.13$	$-31{,}054.6979$	0.083	0.42
	CmC-hominini	$p_0 = 0.42$, $\omega_0 = 0.37$, $p_1 = 0.15$, $\omega_1 = 1.0$, $p_2 = 0.43$, $\omega_{nh} = 0.042$, $\omega_h = 0.25$	$-31{,}053.9072$	**0.03**	0.30
	M2a_rel	$p_0 = 0.42$, $\omega_0 = 0.37$, $p_1 = 0.15$, $\omega_1 = 1.0$, $p_2 = 0.43$, $\omega_2 = 0.043$	$-31{,}056.1959$	NA	NA

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 5 of 9

related chimpanzee lineage aproximately ~ 6–7 million years ago [7]. This expansion in neocortex size in the hominin lineage might have occurred prior to the split of anatomically modern humans from archaic hominins (Neanderthals and Denisovans) approximately 550,000–750,000 years ago, as both modern humans and Neanderthals exhibit comparable overall brain size [38, 39]. However, cranial lobe size (that demarcate different regions of neocortex) does differ between anatomically modern humans and Neanderthals, most prominently parieto-temporal lobe of the neocortex has increased and orbitofrontal cortex is wider in modern humans as compared to Neanderthals [38, 40]. This indicates that certain neocortical regions have evolved after the split of Neanderthals and anatomically modern humans. The size of the neocortex is predominantly determined by the magnitude of neurogenesis and cytokinesis during fetal development. During neurogenesis, cortical neurons originate from a progenitor cell in the ventricular zone of the developing brain [41]. The progenitor cells undergo successive cycles of proliferative division before entering to neurogenic division and formation of the subventricular zone [42–45]. Massive expansion in the neocortex size during human evolution has been attributed to extraordinary expansion in neuron number through increased rate of neural progenitor cell division [5]. This expansion can be explained in the context of the "*radial unit hypothesis*" of cortical development [46]. This hypothesis proposes a general mechanism for a rapid increase in neocortical surface area during evolution through prolonged proliferative symmetric division period. This could yield an increased number of radial columnar units that ultimately generate neurons and consequently expand the neocortical surface area [46]. An alternate hypothesis, "*intermediate progenitor model*" suggests that during mammalian evolution the expansion in neocortical surface area and folding have occurred due to the escalation in basal progenitor pool size (BP originates from apical radial glia; the main neural progenitor cells in the ventricular zone) and their subsequent expansion in the subventricular zone [47]. Recently, another hypothesis linked the evolutionary expansion of neocortex to increased proliferative capacity of basal radial glia (bRG) [48]. Though the timing of brain development is conserved across mammals, however species specific differences in the duration of cortical neurogenesis (6 days in mice, 60 days in macaque, and 100 days in humans) might have contributed to the differences in neocortex size and complexity during primate history [5, 49]. Species-specific difference in BP pool size abundance and temporal aspects of neocorticogenesis might have some lineage specific genetic underpinnings [48].

The majority of MCPH gene products are known to play an important role in the regulation of duration and mode of cell division and hence identified as prime suspect in evolutionary expansion of mammalian/primate cerebral cortex [43, 50, 51]. Initial evaluation of MCPH genes has revealed that *MCPH1, CDK5RAP2, ASPM,* and *CENPJ* evolved adaptively in the human lineage [9, 52, 53]. Further investigations have extended the signatures of positive selection in these MCPH loci (*MCPH1, CDK5RAP2, ASPM,* and *CENPJ*) from human to all anthropoid primates [31]. Recently, the inclusion of the non-primate eutherian species in evolutionary genetic studies of MCPH genes revealed the signatures of pervasive positive selection on *ASPM, CDK5RAP2, MCPH1, CENPJ, CEP152* and *WDR62* throughout the eutherians history [54].

The present study investigates the molecular evolutionary history of eight newly identified MCPH genes by employing sequence data from 48 eutherian genomes and rigorous maximum likelihood based codon substitution models. The codon substitutions site analysis indicated that positive selection occurred during different stages of eutherian evolution in four MCPH genes, *ZNF335, SASS6, CIT,* and *KIF14* (Table 1 and Additional file 2: Table S1). Furthermore, the codon based site models revealed an inconsistent signature of pervasive positive selection in primates for all of the microcephaly genes analyzed in the present study except *KIF14* (Table 1 and Additional file 2: Table S1). Positive selection often acts for a brief period of evolutionary time or transiently on protein-coding intervals [34]. During the course of primate history brain enlargement seems to have happened episodically, therefore, transient or episodic positive selection could be of particular relevance to genes involved in brain size expansion [7]. Intriguingly, for the majority of MCPH genes analyzed in the present study, the branch-site model was unable to identify any signatures of episodic positive selection from primate ancestor to human terminal branch (Additional file 2: Table S2). These data obtained through codon substitutions site and branch-site models were corroborated further by employing clade model C (CmC) and conclusively showed that majority of the MCPH genes have

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 6 of 9

maintained their conserved functions throughout the history of placental mammals (Table 3).

Multiple sequence alignments of MCPH proteins analyzed in the present study have shown human specific substitutions for ZNF335, KIF14, PHC1, MFSD2A and CIT (Additional file 2: Table S3). Majority of these human specific substitutions appear to have fixed prior to the divergence of fully modern humans from archaic humans (Neandertals and Denisovans) some 450,000 years ago (Additional file 2: Table S3). Regardless the fact that human-specific adoptive evolution has not been detected in any of the MCPH genes analyzed here, we speculate that human specific replacements in subset of MCPH proteins could potentially be important in modifying their functions during the course of hominin evolution, The evolutionary relevance of these hominin-specific amino acid substitutions in evolution of brain size and complexity needs to be validated through further studies.

Conclusion

The present study demonstrates that the evolutionary enlargement of human brain cannot be attributed solely to the protein-coding sequences of MCPH genes [55]. Instead, complex conditional effects of human specific coding and non-coding regulatory changes in MCPH and other brain related loci might have been instrumental in the evolution of human brain size during the Pliocene–Pleistocene era.

Methods

Sequence acquisition and alignment

Full-length coding and amino acid sequences of eight MCPH genes (*CEP135, ZNF335, PHC1, CDK6, SASS6, MFSD2A, CIT, and KIF14*) for 48 eutherian species were retrieved from Ensemble and NCBI (National Center for Biotechnology Information) by using bidirectional BLAST hit strategy [56–58]. These sequence data include 20 primates and 28 nonprimate eutherian species (Additional file 1: Figure S1 and Additional file 3: Data S1).

The complete genomic sequences of archaic humans, the Neandertals and Denisovans were downloaded from Max Planck Institute for Evolutionary Anthropology website (http://www.eva.mpg.de/) in binary SAM (BAM) file format with $50 \times$ and $30 \times$ sequence coverage, respectively [59, 60]. MCPH gene sequences were retrived by calculating the consensus sequences of respective chromosomes from BAM (binary alignment Map) files of archaic genomes and was compared with human MCPH genes by using UGENE software [61].

The orthologous coding sequences were aligned through PRANK with default parameters for the empirical codon model by using phylogenetic information as a guide [62]. PRANK concedes insertion and deletion as a

distinct evolutionary event and introduces indels instead of aligning too divergent sequences and therefore reduces the number of false positive for evolutionary analysis [63, 64].

Analysis of substitution parameters and sites under positive selection

At the protein level, the ratio of nonsynonymous (dN) to synonymous (dS) substitution rates ω measures the selective pressure [32, 65]. The value of ω delineates the strength and direction of natural selection operating on protein-coding sequence; $\omega = 1$, $\omega < 1$, and $\omega > 1$, indicate neutral evolution, negative selection and positive selection respectively. For each of eight MCPH genes within every three datasets (primates, nonprimate eutherians, all eutherians), selective pressure ω was estimated using five different codon substitution site models (M1, M2, M7, M8 and M8a) implemented in CodeML program from PAML4.7 software [32, 66, 67]. In this study, a well-accepted phylogeny of eutherians was used for each gene [68]. In order to estimate the sites under positive selection, we compared the LRT of three pairs of site models. The first pair compare the null model M1 (nearly neutral model that assume the existence of two classes of sites with $\omega = 1$ and $\omega < 1$) and alternative model M2 (positive selection model that assume an additional third class of site with $\omega > 1$) [33, 67]. The other two pairs are null model M7 (beta) and alternative model M8 (beta, and $\omega_2 > 1$), and the last pair comparison between null model M8a (beta and $\omega_2 = 1$) and alternative model M8 [65, 69]. For this study, positive selection is inferred, if two out of three site pair models significantly reject the null model in the favor of alternative model. In addition, positively selected sites were identified by using Bayes empirical Bayes method implemented in M8 codon substitution site model [33].

Analysis of episodic positive selection

To evaluate whether or not the signature of positive selection was restricted to specific lineage, we used branch site approach implemented in CodeML [34]. This model allows for ω variation not only among prespecified branches but also among sites. The branch site model allows that phylogeny can be divided into prespecified foreground branch ($\omega_2 > = 1$, proportion of sites may be under positive selection) and background branch (where proportion of sites experienced either purifying selection or neutral evolution $0 < \omega_2 < = 1$). The inference of positive selection was conducted by calculating LRT between this branch site model and null model (it is the same as branch site model but with $\omega_2 = 1$ for foreground branch) [34]. We used eutherian phylogenetic tree as an input for the detection of episodic selection at different

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 7 of 9

evolutionary time point from primate ancestral branch to human terminal branch.

Analysis of divergent selection constraint

To determine the pattern of divergence in selective constraint across the phylogeny, we undertook the clade model C (CmC) approach implemented in CodeML [70]. Clade model C assumes that proportions of sites have evolved under divergent selective pressure but not necessarily under positive selection in two or more partition of phylogeny defined as priory. The LRT was conducted by comparing the CmC model with null model M2a-rel [71]. Both alternative CmC and null M2a-rel models have possessed three classes of sites with $\omega = 1$, $\omega < 1$. The third class of site in M2a-rel has single ω ratio ($\omega_2 > 0$) that is shared between all clades of phylogeny while CmC third class of site has ω ratios equal to the partitions of the phylogeny and varies among the partition of phylogeny [70, 71].

Statistical analysis

P values were calculated in Chi-square program of PAMLX 1.2 package [72]. The P values of the multiple testing hypothesis were corrected for false discovery rate by using the q value package in R3.5.0 [35, 73]. The q values were calculated for each analysis, ranked in ascending order by applying the bootstrap method for π_0 estimation and specified fdr.level $= 0.05$ in q value package in R3.5.0 [74].

Abbreviations

MCPH: Autosomal recessive primary microcephaly; BEB: Bayes Empirical Bayes; dN: Nonsynonymous substitutions rate; dS: Synonymous substitutions rate; LRTs: Likelihood ratio tests; BP: Basal progenitor; bRG: Basal radial glia; CmC: Clade model C; NCBI: National Center for Biotechnology Information; BAM: Binary alignment map.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-021-01801-0.

Additional file 1: Supplemental Figures.

Additional file 2: Supplemental Tables.

Additional file 3. Data S1.

Authors' contributions

N.P designed and performed the experiments, analyzed the data and wrote the paper. H.K analyzed the data. Y.B designed the experiment, analyzed the data and wrote the paper. A.A.A conceived and designed the experiment, analyzed the data and wrote the paper. All authors have read and approved the final manuscript.

Availability of data and materials

The datasets analyzed during the current study are available in the Ensembl database (http://www.ensembl.org), NCBI database (https://www.ncbi.nlm.nih.gov/) and Additional file 3: Data S1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing financial interests.

References

1. Neubauer S, Hublin J-J, Gunz P. The evolution of modern human brain shape. Sci Adv. 2018;4(1):5961.
2. Sakai T, Hirata S, Fuwa K, Sugama K, Kusunoki K, Makishima H, Eguchi T, Yamada S, Ogihara N, Takeshita H. Fetal brain development in chimpanzees versus humans. Curr Biol. 2012;22(18):R791–2.
3. Leigh SR. Brain growth, life history, and cognition in primate and human evolution. Am J Primatol. 2004;62(3):139–64.
4. Ardesch DJ, Scholtens LH, Li L, Preuss TM, Rilling JK, van den Heuvel MP. Evolutionary expansion of connectivity between multimodal association areas in the human brain compared with chimpanzees. Proc Natl Acad Sci. 2019. https://doi.org/10.1073/pnas.1818512116.
5. Geschwind DH, Rakic P. Cortical evolution: judge the brain by its cover. Neuron. 2013;80(3):633–47.
6. Shultz S, Maslin M. Early human speciation, brain expansion and dispersal influenced by African climate pulses. PLoS ONE. 2013;8(10):e76750.
7. McHenry HM. Tempo and mode in human evolution. Proc Natl Acad Sci. 1994;91(15):6780–6.
8. Vallender EJ, Mekel-Bobrov N, Lahn BT. Genetic basis of human brain evolution. Trends Neurosci. 2008;31(12):637–44.
9. Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Hum Mol Genet. 2004;13(5):489–94.
10. Jamieson CR, Fryns J-P, Jacobs J, Matthijs G, Abramowicz MJ. Primary autosomal recessive microcephaly: MCPH5 maps to 1q25-q32. Am J Hum Genet. 2000;67(6):1575–7.
11. Zollikofer CP, de León MSP, Lieberman DE, Guy F, Pilbeam D, Likius A, Mackaye HT, Vignaud P, Brunet M. Virtual cranial reconstruction of Sahelanthropus tchadensis. Nature. 2005;434(7034):755.
12. Li H, Bielas SL, Zaki MS, Ismail S, Farfara D, Um K, Rosti RO, Scott EC, Tu S, Chi NC. Biallelic mutations in citron kinase link mitotic cytokinesis to human primary microcephaly. Am J Hum Genet. 2016;99(2):501–10.
13. Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, Carr IM, Roberts E, Hampshire DJ, Crow YJ, Mighell AJ. Identification of microcephalin, a protein implicated in determining the size of the human brain. Am J Hum Genet. 2002;71(1):136–42.
14. Nicholas AK, Khurshid M, Désir J, Carvalho OP, Cox JJ, Thornton G, Kausar R, Ansar M, Ahmad W, Verloes A. WDR62 is associated with the spindle pole and is mutated in human microcephaly. Nat Genet. 2010;42(11):1010.

15. Bond J, Roberts E, Springell K, Lizarraga S, Scott S, Higgins J, Hampshire DJ, Morrison EE, Leal GF, Silva EO. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. Nat Genet. 2005;37(4):353.

16. Guernsey DL, Jiang H, Hussin J, Arnold M, Bouyakdan K, Perry S, Babineau-Sturk T, Beis J, Dumas N, Evans SC. Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. Am J Hum Genet. 2010;87(1):40–51.

17. Bond J, Roberts E, Mochida GH, Hampshire DJ, Scott S, Askham JM, Springell K, Mahadevan M, Crow YJ, Markham AF. ASPM is a major determinant of cerebral cortical size. Nat Genet. 2002;32(2):316.

18. Kumar A, Girimaji SC, Duvvari MR, Blanton SH. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. Am J Hum Genet. 2009;84(2):286–90.

19. Hussain MS, Baig SM, Neumann S, Nürnberg G, Farooq M, Ahmad I, Alef T, Hennies HC, Technau M, Altmüller J. A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. Am J Hum Genet. 2012;90(5):871–8.

20. Genin A, Desir J, Lambert N, Biervliet M, Van Der Aa N, Pierquin G, Killian A, Tosi M, Urbina M, Lefort A. Kinetochore KMN network gene CASC5 mutated in primary microcephaly. Hum Mol Genet. 2012;21(24):5306–17.

21. Awad S, Al-Dosari MS, Al-Yacoub N, Colak D, Salih MA, Alkuraya FS, Poizat C. Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. Hum Mol Genet. 2013;22(11):2200–13.

22. Khan MA, Rupp VM, Orpinell M, Hussain MS, Altmüller J, Steinmetz MO, Enzinger C, Thiele H, Höhne W, Nürnberg G. A missense mutation in the PISA domain of HsSAS-6 causes autosomal recessive primary microcephaly in a large consanguineous Pakistani family. Hum Mol Genet. 2014;23(22):5940–9.

23. Hussain MS, Baig SM, Neumann S, Peche VS, Szczepanski S, Nürnberg G, Tariq M, Jameel M, Khan TN, Fatima A. CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. Hum Mol Genet. 2013;22(25):5199–214.

24. Yang YJ, Baltus AE, Mathew RS, Murphy EA, Evrony GD, Gonzalez DM, Wang EP, Marshall-Walker CA, Barry BJ, Murn J. Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation and differentiation. Cell. 2012;151(5):1097–112.

25. Gul A, Hassan MJ, Hussain S, Raza SI, Chishti MS, Ahmad W. A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. J Hum Genet. 2006;51(9):760–4.

26. Moawia A, Shaheen R, Rasool S, Waseem SS, Ewida N, Budde B, Kawalia A, Motameny S, Khan K, Fatima A. Mutations of KIF14 cause primary microcephaly by impairing cytokinesis. Ann Neurol. 2017;82(4):562–77.

27. Basit S, Al-Harbi KM, Alhijji SA, Albalawi AM, Alharby E, Eldardear A, Samman MI. CIT, a gene involved in neurogenic cytokinesis, is mutated in human primary microcephaly. Hum Genet. 2016;135(10):1199–207.

28. Pervaiz N, Abbasi AA. Molecular evolution of WDR62, a gene that regulates neocorticogenesis. Meta gene. 2016;9:1–9.

29. Mochida GH. Genetics and biology of microcephaly and lissencephaly. In: Seminars in pediatric neurology: 2009. Amsterdam: Elsevier; 2009. p. 120–6.

30. Montgomery SH, Mundy NI. Evolution of ASPM is associated with both increases and decreases in brain size in primates. Evolution. 2012;66(3):927–32.

31. Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. Mol Biol Evol. 2010;28(1):625–38.

32. Yang Z, Nielsen R, Goldman N, Pedersen A-MK. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 2000;155(1):431–49.

33. Yang Z, Wong WS, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22(4):1107–18.

34. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–9.

35. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci. 2003;100(16):9440–5.

36. Semendeferi K, Damasio H. The brain and its main anatomical subdivisions in living hominoids using magnetic resonance imaging. J Hum Evol. 2000;38(2):317–32.

37. Carroll SB. Genetics and the making of Homo sapiens. Nature. 2003;422(6934):849–57.

38. Florio M, Borrell V, Huttner WB. Human-specific genomic signatures of neocortical expansion. Curr Opin Neurobiol. 2017;42:33–44.

39. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43.

40. Bastir M, Rosas A, Gunz P, Peña-Melian A, Manzi G, Harvati K, Kruszynski R, Stringer C, Hublin J-J. Evolution of the base of the brain in highly encephalized human species. Nat Commun. 2011;2:588.

41. Mo Z, Moore AR, Filipovic R, Ogawa Y, Kazuhiro I, Antic SD, Zecevic N. Human cortical neurons originate from radial glia and neuron-restricted progenitors. J Neurosci. 2007;27(15):4132–45.

42. Rakic P. Specification of cerebral cortical areas. Science. 1988;241(4862):170–6.

43. Rakic P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. Trends Neurosci. 1995;18(9):383–8.

44. Bystron I, Blakemore C, Rakic P. Development of the human cerebral cortex: Boulder Committee revisited. Nat Rev Neurosci. 2008;9(2):110.

45. Stancik EK, Navarro-Quiroga I, Sellke R, Haydar TF. Heterogeneity in ventricular zone neural precursors contributes to neuronal fate diversity in the postnatal neocortex. J Neurosci. 2010;30(20):7028–36.

46. Rakic P. Radial unit hypothesis of neocortical expansion. In: Novartis foundation symposium: 2000. Chichester: Wiley; 1999. p. 30–52.

47. Kriegstein A, Noctor S, Martínez-Cerdeño V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. Nat Rev Neurosci. 2006;7(11):883.

48. Nonaka-Kinoshita M, Reillo I, Artegiani B, Martínez-Martínez MÁ, Nelson M, Borrell V, Calegari F. Regulation of cerebral cortex size and folding by expansion of basal progenitors. EMBO J. 2013;32(13):1817–28.

49. Finlay BL, Darlington RB. Linked regularities in the development and evolution of mammalian brains. Science. 1995;268(5217):1578–84.

50. Caviness V Jr, Goto T, Tarui T, Takahashi T, Bhide P, Nowakowski R. Cell output, cell cycle duration and neuronal specification: a model of integrated mechanisms of the neocortical proliferative process. Cereb Cortex. 2003;13(6):592–8.

51. Kornack DR, Rakic P. Changes in cell-cycle kinetics during the development and evolution of primate neocortex. Proc Natl Acad Sci. 1998;95(3):1242–6.

52. Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science. 2005;309(5741):1717–20.

53. Evans PD, Vallender EJ, Lahn BT. Molecular evolution of the brain size regulator genes CDK5RAP2 and CENPJ. Gene. 2006;375:75–9.

54. Montgomery SH, Mundy NI. Microcephaly genes evolved adaptively throughout the evolution of eutherian mammals. BMC Evol Biol. 2014;14(1):120.

55. Boyd JL, Skove SL, Rouanet JP, Pilaz L-J, Bepler T, Gordân R, Wray GA, Silver DL. Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. Curr Biol. 2015;25(6):772–9.

56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

57. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T. The Ensembl genome database project. Nucleic Acids Res. 2002;30(1):38–41.

58. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005;33(1):D501–4.

59. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43.

60. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, et al. A high-coverage genome sequence from an Archaic Denisovan individual. Science. 2012;338(6104):222–6.

61. Okonechnikov K, Golosova O, Fursov M, Team U. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166–7.

62. Löytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science. 2008;320(5883):1632–5.

Pervaiz *et al. BMC Ecol Evo*   (2021) 21:76

Page 9 of 9

63. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Multiple sequence alignment methods. Berlin: Springer; 2014. p. 155–70.
64. Fletcher W, Yang Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. Mol Biol Evol. 2010;27(10):2257–67.
65. Yang Z, Swanson WJ. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol. 2002;19(1):49–57.
66. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
67. Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 2004;168(2):1041–51.
68. Nikolaev S, Montoya-Burgos JI, Margulies EH, Rougemont J, Nyffeler B, Antonarakis SE, Program NCS. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. PLoS Genet. 2007;3(1):e2.
69. Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 2003;20(1):18–20.
70. Bielawski JP, Yang Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol. 2004;59(1):121–32.
71. Weadick CJ, Chang BS. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. Mol Biol Evol. 2011;29(5):1297–300.
72. Xu B, Yang Z. PAMLX: a graphical user interface for PAML. Mol Biol Evol. 2013;30(12):2723–4.
73. Team RC. R: a language and environment for statistical computing. 2018.
74. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc Ser B (Stat Methodol). 2004;66(1):187–205.

Publisher's Note