AUTHOR COPY

Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids

Ying Li[*1–3], Chun Ye[*1], Peng Shi[1], Xiao-Ju Zou[1], Rui Xiao[1], Yuan-Ying Gong[1] and Ya-Ping Zhang[1,2]

[1]Laboratory of Molecular Evolution and Genome Diversity, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

[2]Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming 650091, China

[3]Graduate School, Chinese Academy of Sciences, Beijing 100039, China

(Requests for offprints should be addressed to Y-P Zhang; Email: zhangyp@public.km.yn.cn)

[*](Ying Li and Chun Ye contributed equally to this paper)

Abstract

The growth hormone (GH) gene family represents an erratic and complex evolutionary pattern, involving many evolutionary events, such as multiple gene duplications, positive selection, the birth-and-death process and gene conversions. In the present study, we cloned and sequenced GH-like genes from three species of New World monkeys (NWM). Phylogenetic analysis strongly suggest monophyly for NWM GH-like genes with respect to those of Old World monkeys (OWM) and hominoids, indicating that independent gene duplications have occurred in NWM GH-like genes. There are three main clusters of genes in putatively functional NWM GH-like genes, according to our gene tree. Comparison of the ratios of nonsynonymous and synonymous substitutions revealed that these three clusters of genes evolved under different kinds of selective pressures. Detailed analysis of the evolution of pseudogenes showed that the evolutionary pattern of this gene family in platyrrhines is in agreement with the so-called birth-and-death process.

Journal of Molecular Endocrinology (2005) **35,** 399–409

Introduction

Growth hormone (GH) is a key protein in stimulating the growth and metabolism of muscle, bone and cartilage cells. In mammals generally, the pituitary GH gene evolved very slowly, with the two exceptions of the primates and the artiodactyls. Especially in the higher primates, a burst of rapid evolution occurred, in which about 62–64 residues changed in the mature hormone sequences compared with the proposed ancestor sequence of eutherian GH (Wallis 1994). The period of the rapid evolution has been dated to before the separation of lineages leading to New World monkeys (NWM) and Old World monkeys (OWM)/hominoids, but after the divergence of the tarsier from the higher primates (Liu *et al.* 2001, Wallis *et al.* 2001).

Most mammals and lower primates, such as the bush baby and the slow loris, have only a single copy of the GH gene in their genomes (Adkins *et al.* 2001, Wallis *et al.* 2001), while, as the result of a series of gene duplications, a cluster of GH-like genes has been observed in the marmoset, spider monkey, rhesus monkey and man (Chen *et al.* 1989, Golos *et al.* 1993, Wallis *et al.* 2001, Wallis & Wallis 2002, Mendoza *et al.* 2004). Humans have five GH-related genes, all located in chromosome 17q22–24 (George *et al.* 1981). These genes include the pituitary-expressed GH gene (hGHN) and four placenta-expressed genes: GH variant (hGHV), chorionic somatomammotropins A and B (hCSHA and hCSHB) and a pseudogene (hCSL). Functional members of the hGH/hCSH family have 191 amino-acid residues encoded by five exons, and the sequence similarity among members reaches more than 90% at the DNA level (Chen *et al.* 1989). In the rhesus monkey, one pituitary-expressed GH gene and four placenta-expressed genes have been reported (Golos *et al.* 1993). However, in the marmoset (a NWM), all the GH-like genes cluster together in the phylogenetic tree, and none cluster with any GH-like gene of human or rhesus monkey, indicating that the gene duplications that gave rise to the marmoset GH-like gene cluster were independent of those of man and rhesus monkey (Wallis & Wallis 2002). However, it is unclear whether this independent gene duplication is species-specific, lineage specific, or merely a placental lactogen (also known as chorionic somatomammotropins (CSH) in primates) specific event, as suggested by Mendoza *et al.* (2004). Thus, a comparative study of GHs, focusing on several closely related species in NWM, is necessary to clarify the evolution of the GH gene family in primates.

In this study, we cloned and sequenced GH-like genes from three representative species of NWM. In total,

DOI: 10.1677/jme.1.01778
0952–5041/05/035–399 © 2005 Society for Endocrinology *Printed in Great Britain*
Online version via http://www.endocrinology-journals.org

AUTHOR COPY

14 GH-like sequences were identified by at least two independent PCRs. Phylogenetic analysis indicates that the gene duplications that gave rise to the NWM and OWM/hominoid GH gene family occurred independently. Moreover, these independent gene duplications are calculated to have occurred after the division of NWM and OWM, further confirming that the GH genes in catarrhines and platyrrhines have expanded independently. Detailed analyses show that the evolutionary pattern of the GH gene family in NWM is coincident with the birth-and-death process. We also note that the three functional members of GH-like genes in these lineages evolved under different selective constraints.

Materials and methods

DNA extraction, PCR, cloning and sequencing

Genomic DNA was prepared from three species of NWM, the red howler (*Alouatta seniculus*), the dusky titi (*Callicebus moloch*) and the white-faced saki (*Pithecia pithecia*). The PCR amplification for GHs was successfully performed with the primer pair that Wallis *et al.* (2001) have reported for amplification of marmoset (*Callithrix jacchus*) GHs; the sense primer (5′ upstream) is TGGCTATCCTGACATCCTTTCCCGC and the antisense primer (3′ downstream) is CCACCCCATAA TATTAGAGAAGGACAC. The PCR protocol was similar to that reported by Wallis *et al.* (2001); that is, 50–100 ng genomic DNA, 2·5 units Pfu Ultra High-Fidelity DNA polymerase (Stratagene, La Jolla, CA, USA) or 2·5 units Taq DNA polymerase (Huamei, China); 0·2 mM each of dNTP; and 5 μl 10 × reaction buffer, adjusted to a final 50 μl volume with ddH_2O. A long denaturing step of 94 C° for 5 min was followed by 57–60 °C for 3 min and 72 °C for 5 min, and then 30 cycles of 94 °C for 1 min, 57–60 °C for 1 min and 72 °C for 5 min, continued by a final extension step of 72 °C for 10 min. The PCR products prepared with Pfu polymerase were cloned into the PCR-Script Amp cloning kit, according to the manufacturer's protocol (Stratagene), and transformed into an ultracompetent *E. coli* cell (Takara, China). Those PCR products obtained by using Taq polymerase were cloned into PMD 18-T Vector (Takara, China) and transferred into an ultracompetent cell (Takara, China). Plasmids carrying a PCR fragment were extracted and sequenced directly with an ABI 3700 automatic sequencer. To reduce single nucleotide variants and artificial recombination occurring by PCR errors, we repeated independent PCR three times (twice with Taq polymerase and once with Pfu polymerase), cloning and sequencing; 10–40 clones from each PCR were chosen at random for sequencing. Those sequences obtained by at least two independent PCRs were taken into account. Those sequences identified by only one PCR were confirmed by another independent PCR using specific primer PCR (the specific primers are not shown) and sequencing.

Evolutionary analysis

Sequences were aligned by the ClustalX program (Jeanmougin *et al.* 1998) with default settings and checked for obvious errors by manual inspection. The phylogenetic relationships of GHs among different primate lineages were inferred from full-length sequences. Kimura's (K80) (Kimura 1980) model was used to infer a neighbor-joining (NJ) (Saitou & Nei 1987) tree implemented in the program MEGA 2·1 (Kumar *et al.* 2001). A maximum-parsimony (MP) tree for potential functional GH genes was constructed by PAUP* 4·0 (Swofford 1998), and was generated by a heuristic search option with 1000 random sequence addition followed by tree bisection–reconnection (TBR) branch swapping. A consensus tree was inferred from the four most parsimonious trees by using the 50% major rule, and then 1000 replicate bootstrap analysis was conducted to test the reliability of this consensus tree. Furthermore, all available intron region and coding sequences for potential functional GHs were used to construct the phylogenetic tree by NJ and MP methods.

To test for a possibly inconsistent rate of sequence substitution, the method of Li and Bousquet (1992), implemented in RRTree (Robinson-Rechavi & Huchon 2000), was used for the relative rate test. Statistical tests for gene conversion by the method of Sawyer (1989) were performed with the GENECONV 1·81 program (Sawyer 2000), with full-length (including five exons and four introns) sequences as the input data. This procedure can rank the possible gene conversion events in an alignment and provide the probability (*P* value) of a segment of silent polymorphic sites as long as or longer than observed under the null hypothesis of no gene conversion. *P* values are computed from 10 000 permutations. *P* values lower than 0·05 are considered to be evidence of gene conversion. Since our alignment was originated from multiple species, some polymorphic sites might vary between species but not within species, and since using these sites artificially increases the probability of finding gene conversion events, we defined sequences from the same species to be a group. Mismatch was either not allowed (default setting) or allowed and given a penalty (gscale) of 1, 2 or 3.

Pairwise nonsynonymous substitutions (d_N) and synonymous substitutions per site (d_S) were calculated by the modified N-G method (Zhang *et al.* 1998) implemented in MEGA 2·1 (Kumar *et al.* 2001). Maximum likelihood analysis of variation of selective pressure during evolution of the GH gene family in NWM was performed with the codeml program in the PAML 3·13 package (Yang 1997), in which transition/transversion

AUTHOR COPY

Table 1 Sequences used in this study

Gene	Accession no.	Gene	Accession no.
Human GH/CSH cluster	J03071	H.leu7 (gibbon ghlp7)	AY621641
Chimpanzee GHN	AF374232	R.rox1 (golden monkey ghlp1)	AY621642
Chimpanzee GHV	AF374233	R.rox2 (golden monkey ghlp2)	AY621643
Macaque GHV	L16555	R.rox3 (golden monkey ghlp3)	AY621644
Spider monkey GHN	AF374234	R.rox4 (golden monkey ghlp4)	AY621645
Spider monkey GHB	AF374235	P. nem1 (langur ghlp1)	AY621646
Spider monkey GHC	AY435434	P.nem2 (langur ghlp2)	AY621647
Marmoset GHN	AJ297563	P.nem3 (langur ghlp3)	AY621648
Marmoset GHlp2	AJ489807	P.nem4 (langur ghlp4)	AY621649
Marmoset GHlp3	AJ489808	P.nem5 (langur ghlp5)	AY621650
Marmoset GHlp5	AJ4898010	M.ass1 (Assamese macaque ghlp1)	AY621651
Marmoset GHlp6	AJ4898011	M.ass2 (Assamese macaque ghlp2)	AY621652
Marmoset GHlp7	AJ4898012	M.ass3 (Assamese macaque ghlp3)	AY621653
Marmoset GHlp8	AJ4898013	M.ass4 (Assamese macaque ghlp4)	AY621654
Squirrel monkey GH	AF339060	M.ass5 (Assamese macaque ghlp5)	AY621655
Western tarsier GH	AF339081	A.sen1 (red howler ghlp1)	AY744451
Phillippine tarsier GH	AF339080	A.sen2 (red howler ghlp2)	AY744452
Slow loris GH	AJ297562	A.sen3 (red howler ghlp3)	AY744453
Bush baby GH	AF292938	A.sen4 (red howler ghlp4)	AY744454
Chimpanzee PL-A	AY146625	A.sen5 (red howler ghlp5)	AY744455
Chimpanzee PL-B	AY146626	C.mol1 (titi ghlp1)	AY744456
Chimpanzee PL-C	AY146627	C.mol2 (titi ghlp2)	AY744457
Chimpanzee PL-D	AY146628	C.mol3 (titi ghlp3)	AY744458
H.leu1 (gibbon ghlp1)	AY621635	C.mol4 (titi ghlp4)	AY744459
H.leu2 (gibbon ghlp2)	AY621636	C.mol5 (titi ghlp5)	AY744460
H.leu3 (gibbon ghlp3)	AY621637	P.pit1 (saki ghlp1)	AY744461
H.leu4 (gibbon ghlp4)	AY621638	P.pit2 (saki ghlp2)	AY744462
H.leu5 (gibbon ghlp5)	AY621639	P.pit3 (saki ghlp3)	AY744463
H.leu6 (gibbon ghlp6)	AY621640	P.pit4 (saki ghlp4)	AY744464

ghlp: growth hormone-like protein; PL: placental lactogen.

and codon usage bias were both taken into consideration (Yang 1997). F3 × 4 was used as the codon-substitution model. The 'branch-site' model accounts for variation in selective pressure both among sites and among lineages. It is used to test selective constraints along the predefined branches (foreground), affecting only a few amino-acid sites (Yang & Nielsen 2002) with other branches as background. The model assumes four classes of sites. Along all lineages in the phylogeny, the first two site classes have ω_0 and ω_1. The third and fourth site classes have ω_0 and ω_1 along all branches except a few branches of interest, which have ω_2. Yang and Nielsen (2002) proposed two models, A and B. In model A, ω_0 and ω_1 are fixed to be 0 and 1 individually, but this assumption seems unrealistic, while in model B, the ω ratios are estimated from the data as free parameters. Therefore, we used only model B here. When the estimate of ω_2 is greater than 1, some sites evolved under positive selection along the branches of interest were indicated, and sites belonging to this class of ω_2 were estimated by Bayesian prediction. This model can be compared with a 'site-specific' model, M3 (Yang *et al.* 2000) (K=2), to construct a likelihood ratio test (LRT) (Yang & Nielsen 2002), in which the test statistic is calculated as $2\delta l$ and compared with a chi-square distribution with the freedom value equal to the difference in the number of parameters between the two models.

Results

Sequence obtained

Fourteen different sequences of GH-like genes were obtained from the three NWM species. These sequences covered entire encoding regions and some 5′ upstream of GH-related genes about 1·9 kb long. Their GenBank accession numbers are shown in Table 1. Aligning those sequences together with other primate GH genes available from GenBank (Table 1), we revealed a GT to AT mutation of the 5′ splice site at the exon 3 and intron 3 boundary and two cases of single base deletions in C.mol3 and P.pit3. These features were all shared with a previously defined pseudogene, marmoset ghlp7 (Wallis & Wallis 2002). In addition, a one-base deletion in sequence C.mol4 was also determined to be shared with marmoset ghlp7. These mutations may lead to a shift of open reading frame and altered native splicing,

AUTHOR COPY

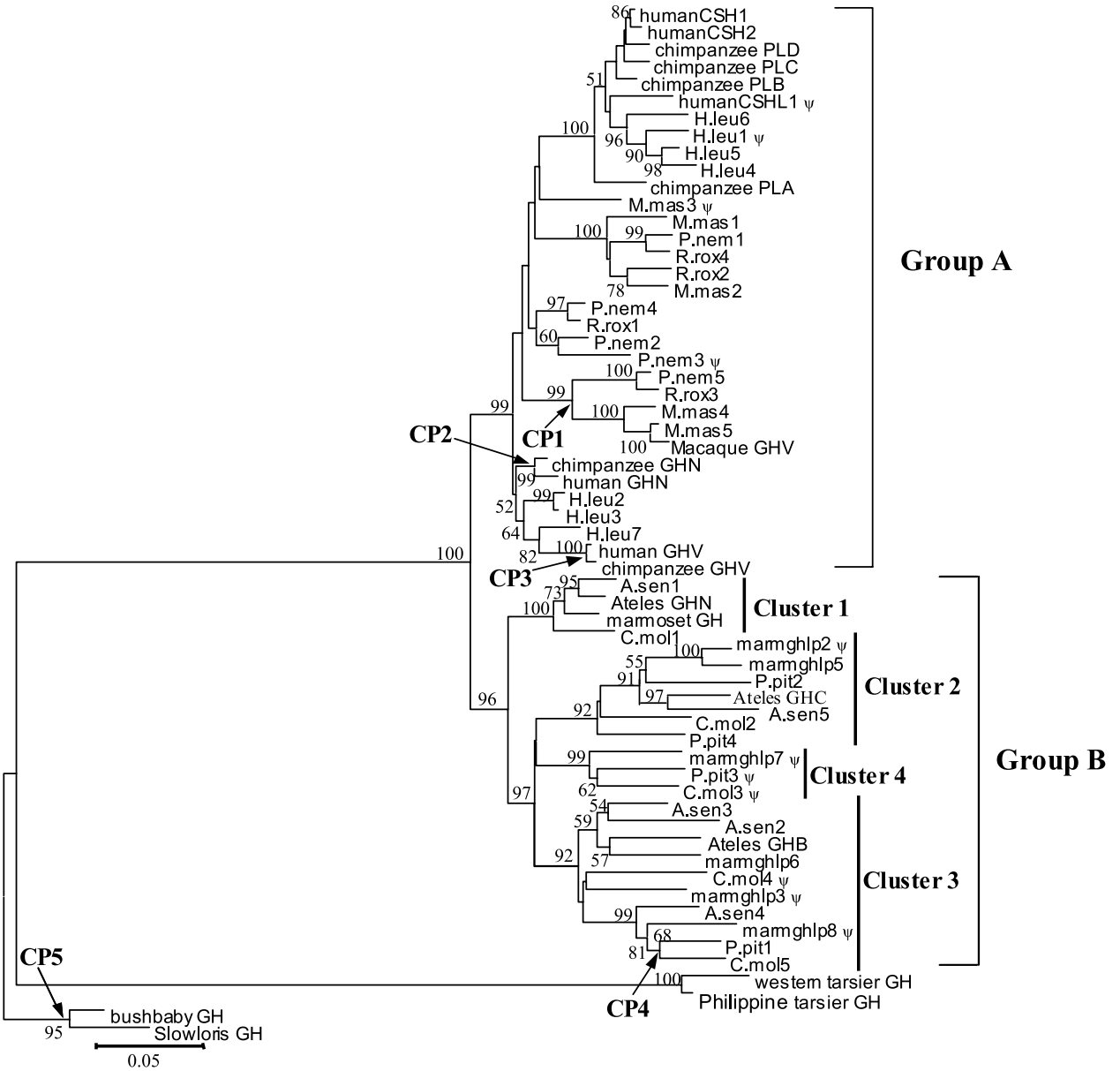

Figure 1 Neighbor-joining tree based on the full-length sequence (contain five exons and four introns) of GH/CSH genes of primates. We used four prosimian GH sequences (bush baby, slow loris, western tarsier and Philippine tarsier) as outgroups to locate the root of this tree. Kimura 2-parameter distances were used. Bootstrap percentages (>50%) by 1000 replications are shown. The sequences identified by our research are designated by the first letter of the genus name and the first three letters of the species name added to the clone number. CP represents calibration points. 'Ψ'indicates pseudogene.

indicating that the three sequences are probably pseudogenes.

Phylogenetic analysis

We used different methods (NJ and MP) and different regions (introns, exons, and full-length sequences) of GH genes for phylogenetic analyses and got similar topologies of gene tree overall. Figure 1 shows the phylogenetic tree inferred by the NJ method with the GH full-length (including all five exons and four introns) sequences. The previous inferred pseudogenes marmoset ghlp4 (Wallis & Wallis 2002) and spider monkey GHD (Mendoza *et al.* 2004) are not used here because they contain large deletions. This tree indicates three noticeable features during GH gene family evolution in

AUTHOR COPY

primates. Firstly and most interestingly, all members of the family (suborder Anthropoidea) are subdivided into two groups, A and B, which are supported by high bootstrap values (99% and 96% respectively). All sequences from NWM clustered together and formed group B, and all sequences from OWM and hominoids clustered to group A. Wallis and Wallis (2002) considered that the duplications that gave rise to the marmoset GH-gene cluster occurred independently of those that gave rise to the corresponding cluster in man, because no marmoset gene is clearly equivalent to any human GH/CSH gene. After getting more information from our study, we observed that, according to the phylogenetic relationship, the gene duplications leading to a cluster of GH-like genes in NWM are independent of those of man and the macaque. Secondly, within each group, GHN-like genes, GHV-like genes, CSH-like genes and other GH-like genes, of unknown function, formed several monophyletic clusters respectively. Moreover, the relationships of these sequences agree with the expected taxonomic placement of these taxa within each cluster. Thirdly, our phylogenetic tree suggested the following four main clusters (clusters 1, 2, 3 and 4) in NWM:

1. Marmoset GH, spider monkey GHN and other GH-like genes form a cluster (cluster 1) with bootstrap support of 100%.
2. Marmoset GHlp5 and other GH-like genes form a single monophyletic cluster (cluster 2) with bootstrap support of 92%.
3. Marmoset GHlp6, spider monkey GHB and other GH-like genes of NWM form a single cluster (cluster 3) with a bootstrap support of 92%.
4. Three putative pseudogenes, marmoset ghlp7, C.mol 3 and P.pit 3, form a single cluster (cluster 4) with 99% bootstrap support.

Gene conversions in GH/CSH genes

We expect to observe a whole tree divided into two (one half with GH-like sequences and the other with CSH-like sequences) under the condition that all GH-like genes observed in the suborder Anthropoidea originated from the common ancestor gene duplication, and paralogous genes evolved independently. In fact, we observed a tree of intermixed GH-like and CSH-like sequences, but, very interestingly, sequences from NWM and OWM/hominoids do not intermingle with each other. Frequently, gene conversion will make two parologous genes very similar in their sequences (Li 1997, Zhang 2003), and this may alter the normal tree topology of the gene family.

There are many methods to test for events of gene conversion. The statistical method proposed by Sawyer (1989) is considered to be a useful tool for inferring gene conversion (Drouin *et al.* 1999). Using this method, we tested all available NWM GH-like genes (in total 24 sequences) from five species, sequences from each species being defined as a group. Marmoset ghlp4 (Wallis & Wallis 2002) and spider monkey GHD (Mendoza *et al.* 2004) were not used because of their large deletions. Different mismatch penalties give slightly different lists of gene conversion events. Here we show the fragments indicated by any mismatch penalty (gscale values) and the longest estimate of the converted region. Eighteen statistically possible cases of gene conversion were identified by this criterion in the NWM GH gene family (Table 2).

Relative-rate test and gene duplication time estimation

To examine the rate variation among members of the GH/CSH gene family in their intron regions, we divided our data into the following 10 groups: hominoid GHN genes, hominoid GHV genes, hominoid CSH genes, OWM GHN genes, OWM GHV genes, OWM CSH genes, and NWM clusters 1, 2, 3 and 4 genes, based on the tree topology shown in Fig. 1. We compared these 10 groups of gene pairs with the bush baby and the slow loris, which have only a single copy of the GH gene, as outgroups. No significant rate variation was detected in any comparison, suggesting relatively constant evolutionary rates in the intron regions of the GH-like genes. Thus, we could estimate the duplication time leading to NWM and OWM/hominoid GH-like gene cluster by GH intron sequences. Five calibration points (CPs) were employed (shown on Fig. 1) by the estimation of Goodman *et al.* (1998), which assumes that the chimpanzee and man diverged about 6 million years ago (Mya), the bush baby and slow loris 23 Mya, the dusky titi and the white-faced saki 17 Mya, and the Colobini and the Cercopithecini 14 Mya. Using these five CPs, we calculated that the GH intron sequence evolutionary rates are $2{\cdot}38 \times 10^{-9}$, $1{\cdot}77 \times 10^{-9}$, $1{\cdot}23 \times 10^{-9}$, $1{\cdot}79 \times 10^{-9}$ and $2{\cdot}44 \times 10^{-9}$ substitutions per site per year respectively. The average rate is $1{\cdot}92 \times 10^{-9}$ substitutions per site per year. Applying this average rate and the intron sequences of the primate GH/CSH gene family, we estimate the gene duplication that gave rise to the GH gene family in NWM lineage to be 32·55 Mya (29·04–37·37 Mya with 95% bootstrap confidence interval), and that to OWM/hominoids to be 28·02 Mya (24·27–31·77 Mya with 95% bootstrap confidence interval). Thus, the two gene duplication events occurred after the divergence of OWM/hominoids from NWM, which is estimated to be 40 Mya (Goodman *et al.* 1998). This further confirmed our previous conclusion that the GHs in the Platyrrhini and the Catarrhini originated independently.

AUTHOR COPY

Table 2 GH/CSH-like genes identified as likely gene conversion

		Aligned					
	***P* value**	Begin	end	**Len**	**Num poly**	**Num Dif**	**Total Difs**
Sequence name							
A.sen2;A.sen5	0·0000	1054	1682	629	279	20	161
A.sen2;A.sen3	0·0003	17	1171	1155	450	18	81
A.sen2;A.sen4	0·0123	429	1171	743	244	30	162
A.sen1;A.sen3	0·0202	1068	1244	177	88	5	166
C.mol2;C.mol3	0·0000	736	1640	905	287	35	180
C.mol2;C.mol4	0·0419	1085	1464	380	160	22	190
C.mol3;C.mol4	0·0047	1085	1644	560	239	33	177
P.pit3;P.pit4	0·0022	1079	1450	372	160	17	179
marmghlp6;marmghlp7	0·0093	1504	1654	151	74	3	187
marmghlp3;marmghlp5	0·0099	1526	1656	131	64	4	210
marmghlp3;marmghlp7	0·0123	1526	1658	133	66	2	189
marmghlp2;marmghlp3	0·0035	1526	1664	139	69	4	207
marmghlp2;marmghlp7	0·0008	1270	1658	389	150	15	184
marmghlp2;marmghlp6	0·0010	1504	1654	151	74	5	216
marmosetGH;marmghlp6	0·0435	1461	1650	190	82	6	187
marmosetGH;marmghlp2	0·0433	1392	1650	259	110	10	178
marmghlp5;marmghlp6	0·0000	1504	1670	167	84	4	210
marmghlp5;marmghlp7	0·0003	1411	1656	246	105	7	191

P value represents global permutation *P* value. Aligned 'begin' and 'end' are the estimated boundaries of the sequence fragment affected by gene conversion. Num poly is the number of polymorphic sites shared by the two sequences in the inferred conversion region. Num Dif is the number of differences in the converted region. Tot Difs is the number of total differences between the two compared sequences.

Different members of the GH gene family in NWM evolved under different selective constraints

The nonsynonymous (amino-acid altering) to synonymous (silent) substitution rate ratio ($\omega = d_N/d_S$) provides a measure of natural selection at the protein level, with $\omega = 1$, <1 and >1 indicating neutral evolution, purifying selection and positive selection respectively. To explore the evolutionary forces driving the evolution of GHs genes in NWM, we first calculated pairwise nonsynonymous substitutions per nonsynonymous sites (d_N) and synonymous substitutions per synonymous sites (d_S) separately within the three putatively functional clusters of NWM GHs (clusters 1, 2 and 3 in Fig. 2). Cluster 4 genes were not used here because they were all pseudogenes. The average d_N, d_S, and potentially synonymous (S) and nonsynonymous sites (N) are listed in Table 3. We noted that the d_N/d_S values varied greatly among members of GH/CSH gene family. All comparisons between cluster 1 genes showed smaller d_N than d_S, and the average d_N (Table 3) was significantly smaller than the average d_S ($Z = 4{\cdot}40$, $P < 0{\cdot}001$ by the one-tailed Z test). This indicates that purifying selection is dominant during the evolution of cluster 1 genes. While almost all d_N values were greater than d_S in cluster 2 (Fig. 2), the average d_N was significantly higher than the average d_S ($Z = 3{\cdot}22$, $P < 0{\cdot}001$, by the one-tailed Z test). This suggests that positive selection is dominant for cluster 2 genes. As for cluster 3 genes, all comparisons show d_N to be very close to d_S values, The average d_N was $0{\cdot}129 \pm 0{\cdot}012$, which is very close to $d_S = 0{\cdot}122 \pm 0{\cdot}017$, but the difference between the two values was not statistically significant ($\mathcal{Z} = 0{\cdot}34$ $P = 0{\cdot}37$, by the two-tailed Z test). This indicates that NWM cluster 3 genes may be subject to neutral or near neutral selection. Thus, different members of the GH/CSH gene family in NWM evolved under different kind of selection constraints.

Test of variable ω among lineages and sites by the maximum-likelihood method

Our d_N and d_S values were computed by averaging overall sites in the protein. Actually, positive selection may affect only a few sites at a few time points (Gillespie 1991, Yang & Nielsen 2002). Therefore, we reanalyzed our data by the branch-site-specific model, which can detect positive selection at a proportion of sites along a predefined lineage (Yang & Nielsen 2002). Given that the GH/CSH gene family in NWM and OWM/hominoids expanded independently, we used only the NJ tree (shown in Fig. 3), which was constructed by all putatively functional NWM GH genes based on K2 distance, with the bush-baby GH gene as outgroup, as imputed tree file. We defined five subsets (1, 2, 3, 4 and 5) with different branches as foregrounds (Fig. 3) and all other branches as backgrounds. Parameters estimated by model B are shown in Table 4. We observed that the ω value increased slightly for a proportion of sites (10%)

AUTHOR COPY

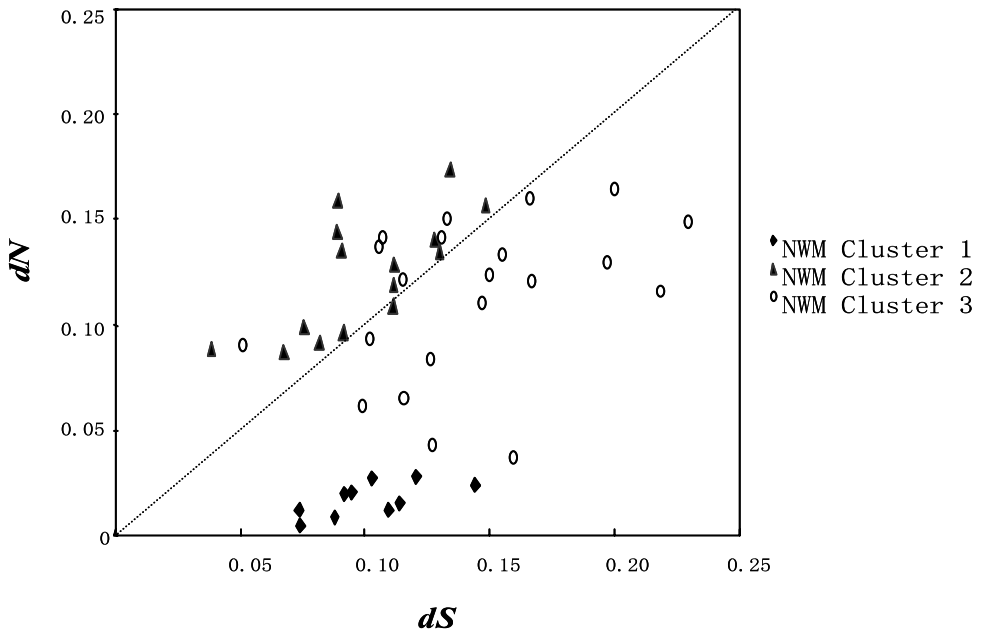

Figure 2 Pairwise synonymous (d_S) and nonsynonymous (d_N) nucleotide distances of the GH gene family in NWM. Dotted line: d_N/d_S =1. The values were obtained by the modified Nei–Gojobori method (Zhang *et al.* 1998) employing the MEGA 2·1 program.

after the first gene duplication event (subset 1), but model B was not significantly better than the null model M3 (K=2). However, 7% of sites in subset 2 showed ω_2 significantly greater than 1, indicating that positive selection occurred immediately after the second gene duplication event. Interestingly, parameter estimates suggested that most sites (78%) under purifying selection for the cluster 1 genes (ω_2=0·0001) and a proportion of sites under positive selection for cluster 2 genes (ω_2=2·28) had significant LRT results (Table 4). As for cluster 3 genes, 12% of sites were inferred to have ω_2=1·98; however, MB was no significant improvement over M3. In this situation, positive selection was not supported by solid evidence; and neutral selection evolution still could not be ruled out in the evolution of cluster 3 genes. These results are consistent with our above conclusion that these three clusters of genes represent three different kinds of selection constraints.

It is believed that the maximum-likelihood method used here is somewhat liberal (Suzuki & Nei 2002, 2004, Zhang 2004). To verify further our results, we reanalyzed our data by the relatively conserved phylogeny-based, branch-specific test (Zhang *et al.* 1998) and also the updated Yang method implemented in PAML 3·14. The results inferred by these methods were all consistent with those shown above.

Discussion

Wallis and Wallis (2002) argue that the duplications that gave rise to a cluster of marmoset GH-like genes were independent of those that gave rise to these genes in man. After getting more species' GH-like genes in NWM, we were able to conclude that the gene duplication which gave rise to GH-like sequences in NWM is independent of those in OWM/hominoids. The conclusion is strongly supported, and this support is independent of the methods and sequences regions applied (other NJ and MP trees are available on

Table 3 Evolutionary rates for members of GH/CSH gene family in NWM

	R=Ts/Tv	N	S	d_N	d_S
Gene					
NWM cluster 1	3·47	393	180	0·014±0·004	0·078±0·014
NWM cluster 2	1·51	339	131	0·143±0·013	0·079±0·015
NWM cluster 3	2·09	403	167	0·129±0·012	0·122±0·017

Ts: transition; Tv: transversion; N: number of potential nonsynonymous sites; S: number of potential synonymous sites.

AUTHOR COPY

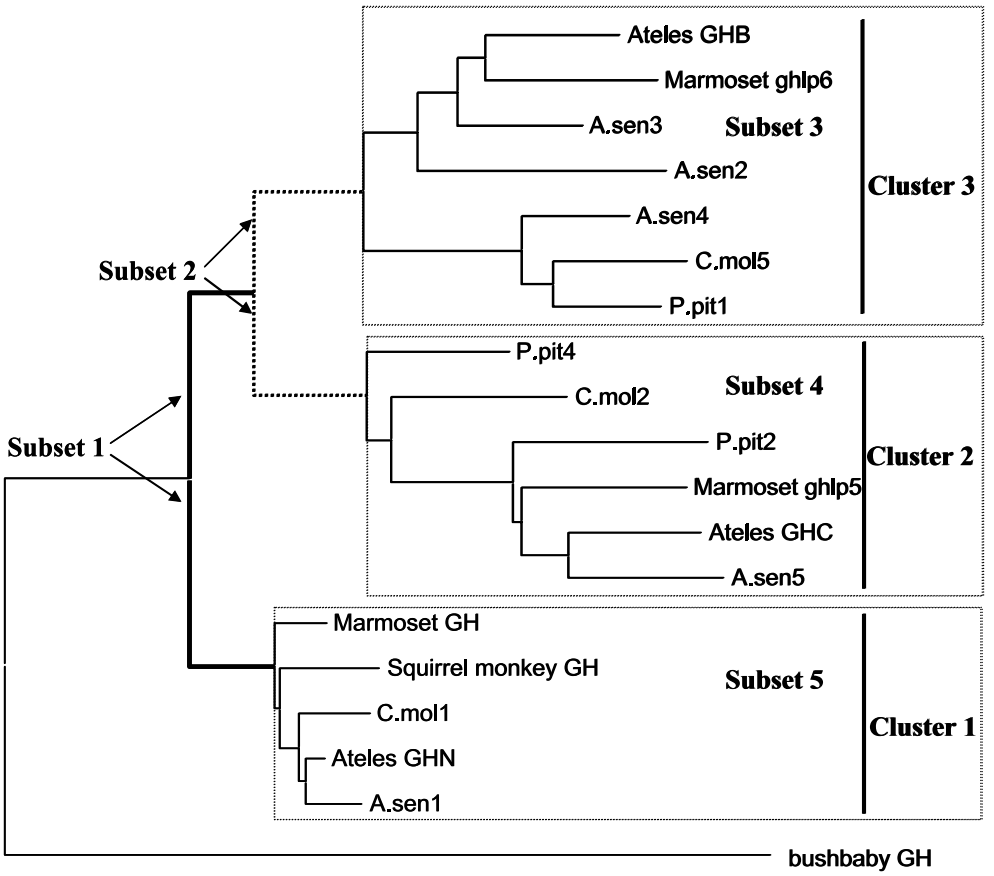

Figure 3 Phylogenies for GH-like genes in NWM implemented as user-defined tree in maximum-likelihood analyses for positive selection.

request). Mendoza *et al.* (2004) got a different tree topology based on smaller species' sequences and the deduced amino-acid distance for the following reasons: the pituitary GH gene evolved under different rates in different lineages (Wallis 1994, Liu *et al.* 2001); as our analysis shows, different family members evolved under different selection constraints; and amino-acid sequences generally offer less information than full-length nucleotide sequences. Thus, our phylogenetic tree should be more reliable. Furthermore, our estimation of gene duplication times leading to NWM GH-like genes and that to OWM/hominoids are 29·04–37·37 and 24·27–31·77 Mya respectively, times which are both after the divergence of NWM and OWM/hominoids. This further indicates independent expansion of GH-like genes in NWM and OWM/hominoids.

Moreover, in studying the relationships of members of a gene family, gene conversion is one commonly invoked explanation when orthologous genes do not form a monophylogeny. Previous studies have identified several gene conversion events in the evolution of the GH gene family (Hirt *et al.* 1987, Chen *et al.* 1989, Giordano *et al.* 1997, Krawczak *et al.* 1999, Horan *et al.* 2003, Mendoza *et al.* 2004, Ye *et al.* 2005). In the present study, we also detected some cases of gene conversion events in NWM by a statistical method. Obviously, gene conversions have occurred in the evolution of this gene family, but the relative importance of this force in forming the present evolutionary pattern of GH family is unclear. We noted that no case shows more similar relationships among all paralogous genes than orthologous genes in any species, a fact which indicates that no recent gene conversion has occurred. Furthermore, the sequences of NWM or OWM/hominoids GH-like genes are not homogenized. Therefore, gene conversion alone cannot explain the great similarity in all sequences of OWM/hominoids and NWM respectively. The combined evidence indicates that gene duplication and expansion independently occurred in the playrrhines and catarrhines. Interestingly, previous research found two different GH alleles in sheep, one allele containing a single GH1 gene and the other containing two duplicated genes, GH-N and GH-Z, and this gene duplication is independent of that in man (Ofir & Gootwine 1997).

AUTHOR COPY

Table 4 Log likelihood values and parameters estimate under branch-site models

	np	*l*	Estimates of parameters	Positively selected sites	2 δ L
Subset					
1	39	−4127·01	p0=0·40 ω 0=0·15; p1=0·50 ω 1=1·14; p2=0·10 ω 2=1·52	foreground: none	0·24
2	39	−4119·99	p0=0·44 ω 0=0·17; p1=0·50 ω 1=1·12; p2=0·07 ω 2=35	foreground: **−20M** −3T 3R 19Q 34N 35C 45L 46R 49K 62F 67V **69A 85S** 153D	14·28**
3	39	−4125·36	p0=0·42 ω 0=0·15; p1=0·46 ω 1=1·14; p2=0·12 ω 2=1·98	foreground: 26D 125T 131G 161W 178H	3·54
4	39	−4123·59	p0=0·40 ω 0=0·14; p1=0·45 ω 1=1·11; p2=0·15 ω 2=2·28	foreground: 37P 41Y 45L 56E 91Q 92P 99N 188S	7·08*
5	39	−4107·79	p0=0·09 ω 0=0·17; p1=0·13 ω 1=1·34; p2=0·78 ω 2=0·00001	foreground: none	36·68**
M3(k=2)	37	−4127·13	p0=0·45 ω 0=0·16; p1=0·55 ω 1=1·15	/	/

np, number of parameters; sites potentially under positive selection are identified using the human sequences as the reference. Those sites with posterior possibility >90% are shown in bold. Positive selected sites for all lineages are not shown. *indicate statistically significant at 95% level, **indicate statistically significant at 99% level.

The biologic and evolutionary meaning of the independent gene duplications is unknown due to lack of knowledge of the function and expression patterns of NWM GH-like genes. In man and the macaque, the three main functions of the GH gene family are as follows: the GHN gene promotes growth and expresses in the pituitary, the GHV gene has the same function but expresses in the placenta, and the CSH gene has prolactin-like function and expresses in the placenta. Coincidentally, the putatively functional NWM GH-like genes can be distinctly divided into three clusters. Whether these three clusters of genes have the corresponding function of human and macaque GH genes is an interesting question and deserves further study.

Figure 1 shows three potential pseudogenes, the P.pit3, C.mol3 and marmoset-ghlp7 forming a single clade (cluster 4) with high bootstrap support. This suggests that the gene duplication giving rise to this cluster of pseudogenes occurred before the divergence of the Cebidae and Callicebini families, inferred to be 25 Mya (Goodman 1998). Thus, this cluster of pseudogenes has existed in the genome for a relatively long time. However, other pseudogenes, such as marmoset ghlp2, are phylogenetically close to the functional marmoset ghlp5 (Wallis & Wallis 2002), indicating that this gene may have been pseudogenized recently. The above analysis indicates that gene duplication and pseudogenization both happened in the evolution of GH-like genes in NWM. This evolutionary pattern is reminiscent of the birth-and-death model, which was proposed to explain the evolutionary processes for several multiple gene families, such as the MHC, buiquitin, immunoglobulin and histone H4 genes (Ota & Nei 1994, Nei *et al.* 1997, 2000, Piontkivska *et al.* 2002). This model of evolution is characterized by frequently occurring gene duplication, and some of the duplicated genes are preserved as functional genes, while others become pseudogenes or are lost from the genome (Nei *et al.* 1997). Roelofs and Rooney (2003) recently proposed that three general features – sequence grouped by gene duplication order rather than by species, relatively low sequence homogeneity, and evidence of gene loss/deletion or pseudogene formation – should be observed in the phylogeny of those gene families evolved by the birth-and-death process. The evolutionary pattern of the GH gene family in NWM is coincident with these criteria in the following three features:

1. As Fig. 1 shows, no species has all paralogous genes clustering together.

2. The sequences in NWM are not homogeneous, since the differences between any two genes in NWM are no less than 17 nucleotides in the coding region.

3. Several pseudogenes are observed.

Therefore, the birth-and-death process is a main evolutionary mechanism in NWM GH-like genes. Most intriguingly, a similar evolutionary pattern was observed in OWM/hominoids GH genes in our previous study (Ye *et al.* 2005).

The episodic evolutionary pattern is an interesting and unusual feature of GH genes among both mammals and primates (Wallis 1994, 1996). However, the driving force of this rapid evolution, whether adaptive evolution or relaxation of purifying selection constraint, is controversial. Ohta (1993) suggested positive selection as a possible cause, but its relative importance is uncertain.

AUTHOR COPY

Wallis (1997) proposed a mechanism of 'function switch' to explain this burst of evolution without involving gene duplication; that is, in one certain period of time, the ancestor of GH displayed a new lactogenic function in addition to the original growth-promoting role. Acquisition of this new function may have required some adaptive selective pressure on the molecule. Therefore, positive selection, rather than relaxation of purify selection, was the driving force of this fast evolution. In fact, Liu *et al.* (2001) found that the amino-acid substitution at functionally important sites occurred more frequently than at other sites, suggesting a role of positive selection during rapid evolution. Moreover, they argued that, except for adaptive evolution, relaxation of selective constraint might also be a cause.

In this study, we focus on the selective pressures among members of GH-like genes in NWM. Interestingly, we observed that different gene clusters are subjected to different kinds of selective constraints. Purifying selection is dominant in cluster 1 genes. In contrast, positive selection is dominant in cluster 2 genes. As for cluster 3 genes, d_N and d_S values show no significant difference, in agreement with the criterion of neutral selection. Therefore, relaxation of purifying selection after duplication played an important role in the evolution of cluster 3 genes. Interestingly, different members of the GH gene family also evolved under different selective constraints in the OWM/hominoid lineage (Ye *et al.* 2005). Without more analyses of the structure, function and expression of the three clusters of GH-like genes in NWM, we are unable to determine the biologic significance of this unusually great variation of selection constraint and whether these duplicated genes are differentially used in time and space. Therefore, further study of the functions of different GH gene family members in NWM is necessary to elucidate the evolution of this gene family in primates.

Acknowledgements

This work was supported by a National Natural Science Foundation of China grants to Y-P Z (30021004 and 30430110) and by the Chinese Academy of Sciences Foundation.

References

Adkins RM, Nekrutenko A & Li WH 2001 Bush baby growth hormone is more similar to nonprimate growth hormone than to rhesus monkey and human growth hormone. *Molecular Biology and Evolution* **18** 55–60.

Chen EY, Liao YC, Smith DH, Saldana BHA, Gelinas RE & Seeburg PH 1989 The human growth hormone locus: nucleotide sequence, biology, and evolution. *Genomics* **4** 479–497.

Cunningham BC & Wells JA 1989 High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244** 1081–1085.

Cunningham BC & Wells JA 1993 Comparison of a structural and a functional epitope. *Journal of Molecular Biology* **234** 554–563.

Drouin G, Prat F, Ell M & Clarke GD 1999 Detecting and characterizing gene conversions between multigene family members. *Molecular Biology and Evolution* **16** 1369–1390.

George DL, Phillips JA, Francke RDU & Seeburg PH 1981 The genes for growth hormone and chorionic somatomammotropin are on the long arm of human chromosome 17 in region q21 to qter. *Human Genetics* **57** 138–141.

Gillespie JH 1991 The causes of molecular evolution. Oxford University Press, Oxford.

Giordano M, Marchetti C, Chiorboli GB & Richiardi PM 1997 Evidence for gene conversion in the generation of extensive polymorphism in the promoter of the growth hormone gene. *Human Genetics* **100** 249–255.

Golos TG, Durning M, Fisher JM & Fowler PD 1993 Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. *Endocrinology* **133** 1744–1752.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G & Groves CP 1998 Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics and Evolution* **9** 585–598.

Hirt H, Kimelman J, Birnbaum MJ, Chen EY, Seeburg PH, Eberhardt NL & Barta A 1987 The human growth hormone gene locus: structure, evolution, and allelic variations. *DNA* **6** 59–70.

Horan M, Millar DS, Hedderich J, Lewis G, Newsway V, Mo N, Fryklund L, Procter AM, Krawcak M & Cooper DN 2003 Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. *Human Mutation* **21** 408–423.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG & Gibson TJ 1998 Multiple sequence alignment with Clustal X. *Trends in Biochemistry Science* **23** 403–405.

Kimura M 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16** 111–120.

Kumar S, Tamura K, Jakobsen IB & Nei M 2001 MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17** 1244–1245.

Krawczak M, Chuzhanova NA & Cooper DN 1999 Evolution of the proximal promoter region of the mammalian growth hormone gene. *Gene* **237** 143–151.

Li P & Bousquet J 1992 Relative-rate test for nucleotide substitutions between two lineages. *Molecular Biology and Evolution* **9** 1185–1189.

Li WH 1997 *Molecular Evolution*. Sunderland, MA: Simauer Associates.

Liu JC, Makova KD, Adkins RM, Gibson S & Li WH 2001 Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. *Molecular Biology and Evolution* **18** 945–953.

Mendoza ARD, Escobedo DE, Dávila IM & Saldaña HB 2004 Expansion and divergence of the GH locus between spider monkey and chimpanzee. *Gene* **336** 185–193.

Nei M, Gu X & Sitnikova T 1997 Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS* **94** 7799–7806.

Nei M, Rogozin LB & Piontkivska H 2000 Purifying selection and birth-and-death evolution in the ubiquitin gene family. *PNAS* **97** 10866–10871.

AUTHOR COPY

Ofir R & Gootwine E 1997 Ovine growth hormone gene duplication – structural and evolutionary implications. *Mammalian Genome* **8** 770–772.

Ohta M 1993 Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. *Genetics* **134** 1271–1276.

Ota T & Nei M 1994 Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. *Molecular Biology and Evolution* **11** 469–482.

Piontkivska H, Rooney AP & Nei M 2002 Purifying selection and birth-and-death evolution in the histone H_4 gene family. *Molecular Biology and Evolution* **19** 689–697.

Rechavi RM & Huchon D 2000 RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16** 296–297.

Robinson-Rechavi M & Huchon D 2000 RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16** 296–297.

Roelofs WL & Rooney AP 2003 Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *PNAS* **100** 9179–9184.

Saitou N & Nei M 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4** 406–425.

Sawyer SA 1989 Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* **6** 526–538.

Sawyer SA 2000 GENECONV: statistical tests for detecting gene conversion – Version 1·81. Department of Mathematics, Washington University, St Louis, MO.

Suzuki Y & Nei M 2002 Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid site. *Molecular Biology and Evolution* **19** 1865–1869.

Suzuki Y & Nei M 2004 False-positive selection identified by ML-based methods: examples from the *sgi1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Molecular Biology and Evolution* **21** 914–921.

Swofford DL 1998 PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4·0. Sunderland, MA: Sinauer Associates.

Wallis M 1994 Variable evolutionary rates in the molecular evolution of mammalian growth hormones. *Journal of Molecular Evolution* **38** 619–627.

Wallis M 1996 The molecular evolution of vertebrate growth hormone: a pattern of near-stasis interrupted by sustained bursts of rapid change. *Journal of Molecular Evolution* **43** 93–100.

Wallis M 1997 Function switching as a basis for bursts of rapid change during the evolution of pituitary growth hormone. *Journal of Molecular Evolution* **44** 348–350.

Wallis M & Wallis OC 2002 Characterizing of the GH gene cluster in a New World monkey, the marmoset (*Callithrix jacchus*). *Journal of Molecular Endocrinology* **29** 89–97.

Wallis OC, Zhang YP & M Wallis 2001 Molecular evolution of growth hormone (GH) in primates. Characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. *Journal of Molecular Endocrinology* **26** 249–258.

Yang Z 1997 PAML: a program package for phylogenetic analysis by maximum likelihood Computer Applications in the Biosciences **13**:555–556 http://abacus.gene.ucl.ac.uk/software/paml.html.

Yang Z & Nielsen R 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* **19** 908–917.

Ye C, Li Y, Shi P & Zhang YP 2005 Molecular evolution of growth hormone gene family in old world monkey and hominoids. *Gene* **350** 183–192.

Zhang J 2003 Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18** 292–298.

Zhang J 2004 Frequent false detection of positive selection by the likelihood method with branch-site models. *Molecular Biology and Evolution* **21** 1332–1339.

Zhang J, Rosenberg HF & Nei M 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* **95** 3708–3713.

Zhang J, Zhang YP & Rosenberg HF 2002 Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics* **30** 411–415.

Received 27 June 2005
Accepted 11 July 2005