図２　ヒト RefSeq mRNAs と相同性を示すチンパンジー5'-ESTs (10,540)
(BLASTN search with E=1e-120)

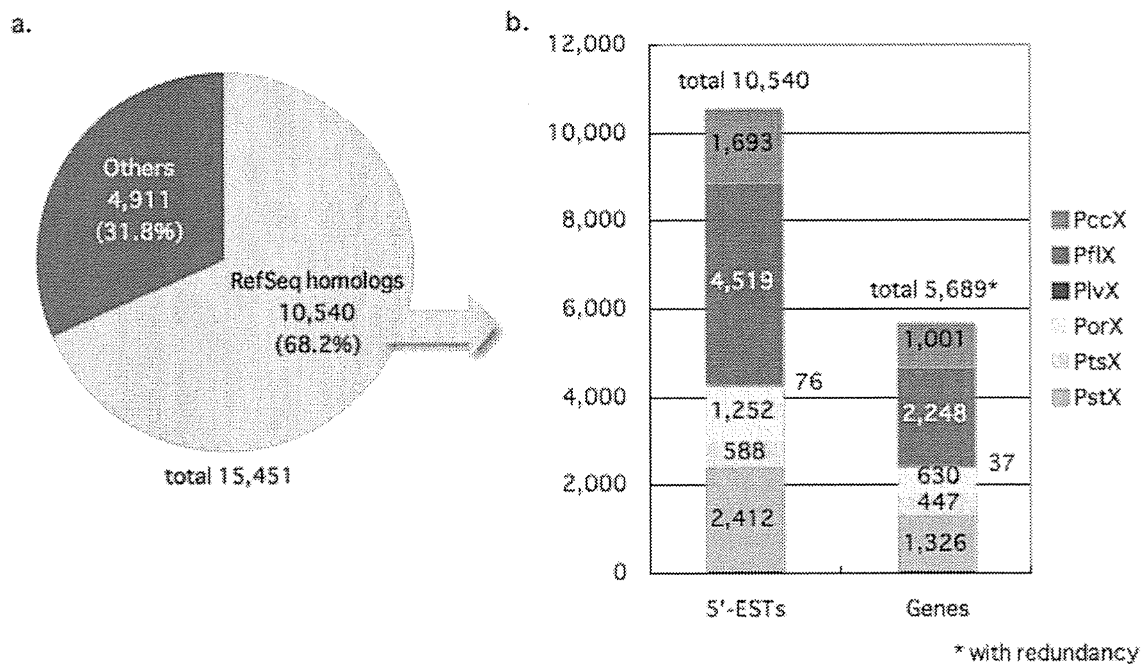

表２　　チンパンジークローン中の疾患関連遺伝子

疾患関連遺伝子	7,492 クローン	2,411 OMIM genes
心臓血管系疾患関連遺伝子	122 クローン	39 OMIM genes
肥満関連遺伝子	106 クローン	31 OMIM genes

表３．　心臓血管系疾患関連遺伝子の比較解析

遺伝子	ヒトhomologuとのアミノ酸比較結果
FBN1	１００%一致
FUCOSIDOSIS	チンプゲノム不完全（アカゲはある）
GBA	シグナルペプチド内1アミノ酸異なる　他種保存性なし
GJA1	アミノ配列１００%一致
GCCR	種間でアミノ酸置換３カ所（そのうち２カ所はモチーフ内でヒトはdog, mouse, ratと一致。チンプのみ異なる）そこにヒトSNPなし
MPS Type VII	アミノ酸異なる　　　　　　　　　　　　　　　他種保存性なし
HLP Type 1	チンプUCSC3'UTR含むギャップ、ヒトのみ異なるアミノ酸置換あるがモチーフの外
Methylmalonicaciduria	シグナルペプチド内1アミノ酸異なる、他種保存性なし。TSS隣の遺伝子(アンチセンス)と共有か
NF1	１００%一致　intron antisenseに神経繊維腫症の遺伝子などある
NP	homologGeneの誤り？（チンプ5' アミノ酸長い, チンプクローンはヒトと同じ）
TTR	SNP多数あるのにヒトのみ異なりSNPなし1カ所（チンプ、チキン、ワラビー、オポッサム、ラット、ドッグ、マウスで保存）、チンプのみ異なりSNPなし1カ所（ただしここはチキンでは保存されずに別のアミノ酸）
IVA	チンプゲノム不完全
LDLR	２アミノ酸ヒトのみ異なる、1アミノ酸チンプのみ異なる、いずれもSNP報告無し
GPX1	チンプhomoloGene登録なし
IL6	ヒトゲノム3'UTRにsimple repeat。チンプのrepeat数小さい。アカゲない。カニクイ対応クローン有り。アミノ酸1カ所異なるがSNP報告無し。他の種で保存されていない。
IGF1	１００%一致　　　homoloGeneはalternativeが登録されている
AGTR1	homoloGeneチンプ5' アミノ酸長いが、チンプクローンはヒトと同じ。1アミノ酸ヒトのみ異なる、SNP報告無し
ALDH2	チンプゲノム不完全、アカゲはある
NPR1	チンプhomoloGene登録なし。5'UTR内CT繰り返し、UCSCヒトCTx11回、UCSCチンプ6回、チンプクローン5回。クローンORF OK
PTGS2	homologGeneチンプ5' アミノ酸長いが、チンプクローンはヒトと同じ。1アミノ酸ヒトのみ異なる、1アミノ酸チンプのみ異なる(チキンも保存)、いずれもヒトSNP報告無し
ADAM10	１００%一致
ADM	１００%一致
CAV1	１００%一致
SELP	アミノ酸置換多い、血小板タンパク。免疫、アトピー、喘息に関連するSNPあり、ヒトRefSeqがマイナー型、ヒトメジャー型と他種の型の方が疾患に相関あり。７１１アミノ酸で１６アミノ酸置換（ヒト一チンプ）
DM2	チンプhomoloGene登録誤り？　チンプクローンはexon/intronはヒトと同じ、アミノ酸配列は不明、ORFは開く。
DNCL1	１００%一致
EPHB6	クローン；チンプ特異的alternative splicing in 5'UTR（ESTもない）。HomoloGene;ヒトのみ異なる3アミノ酸（そのうち1カ所でSNP報告有り、ヒトでマイナーが他種と同じアミノ酸。３カ所のうち２カ所はドメイン内。1アミノ酸チンプのみ異なる（SNP報告無し）
FOLH1	domain内に多数変異。Homologは哺乳類中ヒトとチンプしかない。マウス、ブタ、ラットも同じ名前の遺伝子あるが違う領域（ヒトでは同じ１１番染色体11q14.3）に位置して5'が異なる。そこにはヒトはこの遺伝子の別名の遺伝子が存在している。
AEF2AK3	3アミノ酸ヒトのみ異なる（そのうち1カ所でSNP報告有り、ヒトでメジャー型が他種と同じアミノ酸、RefSeqヒトがマイナー型。）、チンプのみ異なる4アミノ酸（SNPなし）。モチーフ内にも存在
HP	複数コピー、ハプトグロビン
PANT3	チンプhomolog登録なし
CUGBP2	homologeneチンプ5' アミノ酸長いが、チンプクローンnot fullで未確認。chimpのみputativehomologeneのせい？
WASF2	homologeneなし。ヒト、チンプchrXにホモログ有り。マウス、アカゲはない（アカゲは未シークエンシング？）クローンはnot full(3'UTR)
ATE1	チンプ5' アミノ酸長い, チンプクローンはヒトと同じ。homoloGeneの誤り？ 1アミノ酸ヒトのみ異なる、1アミノ酸チンプのみ異なる、いずれもヒトSNP報告無し。
BCAR1	チンプ5' アミノ酸長い, チンプクローンはヒトと同じ。homoloGeneの誤り？Alternative　2アミノ酸ヒトのみ異なる、SNP報告無し。
NOTCH2	チンプのみ5' アミノ酸配列ない, repeatがあるためアセンブリの誤り？チンプクローンはnot full。ヒトのみ異なる2アミノ酸、チンプのみ異なる1アミノ酸、うち2カ所ドメイン内、全てSNP報告無し。遺伝子自体は隣接遺伝子症候群のAlagille症候群に関与など。
GTF21	Homolog Alternatively spliced type
MAPK14	１００%一致
MAP3K3	チンプhomologeneない, UCSCゲノム不完全、アカゲはある

表４　　肥満関連遺伝子の比較解析

遺伝子	ヒトhomologとのアミノ酸比較結果
AGT	cSNPあり（ただしmature peptideの外）、ヒトでマイナーなチンプ型はhypertension, vascular demenia リスク低い
GCCR	種間でアミノ酸置換３カ所（そのうち２カ所はモチーフ内でヒトはdog,mouse,ratと一致。チンプのみ異なる）そこにヒトSNPなし
INSR	遺伝子の5'側上流に染色体１９と７間共通配列（約４０kb）転座の報告あり、チンパンジーは遺伝子の5'からギャップ
Hyperlipo Type1	チンプUCSC3'UTR含むギャップ、ヒトのみ異なるアミノ酸置換あるがモチーフの外
PAPD	１００%一致
PCSK1	イントロン内にヒトのみSINE挿入、アミノ酸１カ所異なるがチンプタイプのSNP報告有り（病気との関連の報告は無い）
GNAS	チンプゲノム配列が誤りか
IL8	１００%一致
IL6	ヒトゲノム3'UTRにsimple repeat。チンプのrepeat数小さい。アカゲない。カニクイ対応クローン有り。アミノ酸１カ所異なるがSNP報告無し。他の種で保存されていない。
HTR2A	アミノ酸１カ所異なるがヒトSNP報告無し。ヒト、マウス、ラットが一致。チンプはイヌと一致
CCK	アミノ酸１カ所ヒトのみ異なる。ヒトSNP報告無し。Mature peptide内。
FABP4	１００%一致
CPE	チンプ5'アミノ酸長い（モチーフない）、ただしチンプクローンはヒトと同じ UCSCチンプゲノムの誤りの可能性高い
Factor D	１アミノ酸ヒトのみ異なる、１アミノ酸チンプのみ異なる　どちらもヒトSNP無し
PTPN1	１００%一致
SCYA2	１００%一致
SDC1	チンプゲノム不完全
UCP2	１００%一致
ACDC	１アミノ酸ヒトのみ異なる、１アミノ酸チンプとイヌのみ異なる、どちらもヒトSNP報告無し。チンプ5'他種より長い、チンプゲノム、クローンともにヒトと同じ位置に翻訳開始のATGある。
MECP2	チンプ5'他種より長い、ヒトと他の亍種は一致
SCD	１アミノ酸チンプのみ異なる、ヒトSNP報告無し。１アミノ酸ヒトのみ異なる、SNP報告あり（疾患との関連の報告はなし）
NTRK2	１００%一致（イヌのみ大きく異なる）
SORBS1	３アミノ酸ヒトのみ異なる、２アミノ酸チンプのみ異なる、いずれもSNP報告無し
SEDL	１００%一致
BBS4	１００%一致
GNAS	チンプの登録homologはalternative typeらしい
NCOA1	２アミノ酸ヒトのみ異なる、いずれもSNP報告無し
LMNA	チンプゲノム不完全なため、homolog登録おかしい
MKKS	ヒトのみ異なる２アミノ酸、チンプのみ異なる２アミノ酸、いずれもSNP報告無し
GBFA2T1	チンプ5'他種と異なる、他はヒトと一致。Not fullなので5'確認できず
NSIG1	ヒトのみ異なる１アミノ酸、SNP報告無し
URB	チンプゲノム不完全なため、homolog登録おかしい

平成１８年度　厚生労働科学研究費補助金（ヒトゲノム・再生医療等研究事業）
分担研究報告書

カニクイザルES細胞分化のDNAチップを用いた解析

分担研究者　鳥居　隆三　滋賀医科大学・動物生命科学研究センター教授

研究要旨　再生医療においてES細胞は機能細胞の供給源として有用であるが、その実用化には移植細胞の厳密な品質管理、分化状態の情報化が必要である。そこでDNAチップを用いた網羅的な遺伝子発現解析の応用を試みる。まずヒトES細胞と類似し、同種移植実験が可能なカニクイザルES細胞を用いて機能細胞への分化系を開発する。分化細胞と生体組織をDNAチップ解析によりその遺伝子発現を比較し、分化、機能性を評価すると同時に品質管理としての利用方法を開発する。

A．研究目的

再生医療、細胞移植医療においてES細胞は機能細胞の供給源として非常に有用である。サルES細胞はヒトES細胞と非常に類似した特性を示し、サルへの同種移植実験が可能であるため、細胞移植医療の有効性、安全性を検証するための優れたモデルと考えられる。ES細胞を用いた再生医療実用化の重要な鍵は、ES細胞の機能細胞への効率良い分化系の開発と、移植用分化細胞の評価、品質管理である。そこで本研究は霊長類ES細胞の機能細胞への分化系の開発と共に、分化状態の情報化、分化細胞の品質管理にDNAチップを用いた網羅的遺伝子発現解析を応用することを目的とする。

B．研究方法

サルES細胞の心筋、脂肪細胞への分化を試みた。ハンギングドロップ培養を行い、自発分化およびインスリン、IBMX、デキサメサゾン等のホルモン添加によりそれぞれ、心筋、脂肪細胞へ分化させた。またsiRNAを用いた新たな分化方法の開発として、ES細胞にPPARγ-siRNAを導入し、骨芽細胞への分化を行った。さらに高品質DNAチップ作成のため、未分化サルES細胞cDNAライブラリー作製を試みた。すなわちフィーダー上で培養した状態の良い未分化ESコロニーを選択的にピックアップし、RNAを抽出することにより行った。

C．研究結果

サルES細胞の心筋細胞、脂肪細胞への分化に成功した。分化した心筋細胞はGATA4,心筋特異的トロポニンＩに加え、心臓ホルモンと呼ばれるナトリウム利尿ホルモン(ANP, BNP)を発現していることを明らかにした。さらに分化心筋の電気生理学的解析からペースメーカー細胞に分化可能なこと、ANP, BNPはトランスゴルジネットワークに存在していることがわかった。脂肪細胞はインスリンレセプター、GLUT4、に加え、レプチンやアディポネクチン等のアディポサイトカインを発現していた。さらにGLUT4はインスリン刺激に反応し細胞膜に移動した。また、新規分化方法の開発研究ではPPARγ-siRNAを導入することにより脂肪細胞への分化を抑制し、骨芽細胞へ分化させられることを明らかにした。現在、未分化サルES細胞cDNAライブラリー作製の作成中であるが、約 $1\text{-}2\times10^4$ のサルES細胞から約2μgの良質なtotal RNAが得られることがわかった。

D．考察

サル ES 細胞から内分泌機能を持つ心筋、脂肪細胞を分化させられることが明らかとなった。すなわち機能性を持つ成熟機能細胞への in vitro 分化が可能であることを示した。さらに化学合成した siRNA を使用して分化方向を脂肪細胞から骨芽細胞へと変えられることを明らかにした。今後、この新規細胞分化方法は肥満、骨粗鬆症の治療方法として利用可能であると考えられる。未分化サル ES 細胞 cDNA ライブラリーに関してはフィーダー細胞の混入を最小限にした高品質なライブラリーが作成できると考えられる。

E．結論

サルES細胞から拍動という物理的な機能のみならず、ホルモン分泌能という内分泌機能を持つ心筋、脂肪細胞を分化させられることが明らかとなった。さらに PPARγ-siRNA を使用して脂肪細胞への分化を骨芽細胞へと分化方向を変えられることを明らかにし、細胞の新規分化方法を開発した。また状態の良い未分化サル ES 細胞コロニーを人為的にピックアップすることにより、フィーダー細胞の混入を最小限に抑えた高品質な未分化サル ES 細胞 cDNA ライブラリーを現在作製中である。

F．研究発表

1．論文発表

1. Takada T., Suzuki Y., Kadota N., Kondo Y., and Torii R. Generation of GFP-expressing monkey ES cells. Methods in Molecular Biology. Humana Press Inc. Embryonic Stem Cell Protocols, Turksen K., ed. 2nd ed. 329, 305-312 (2006)
2. Tsukada, H., Takada, T., Shiomi, N., Tani, T., and Torii, R. Acidic fibroblast growth factor promotes hepatic differentiation of monkey embryonic stem cells. In vitro Cell. Dev-An. 42, 83-88 (2006)
3. Abdelalim, E., Takada, T., Toyoda, F., Omatsu-Kanbe, M., Matsuura, H., Tooyama, I., and Torii, R. In vitro expression of natriureic peptides in cardiomyocytes differentiated from monkey embryonic stem cells. Biochem. Biophys. Res. Commun. 340, 689-695 (2006)
4. Yamashita A., Takada, T., Omatsu-Kanbe, M., Nemoto, K., Matsuura, H., Yamamoto G., and Torii R. Monkey embryonic stem cells differentiate into adipocytes in vitro. Cloning Stem Cells 8, 3-9 (2006)
5. Abdelalim, E., Takada, T., Torii, R. and Tooyama, I. Molecular cloning of BNP from heart and its immunohistochemical localization in the hypothalamus of monkey. Peptides 27, 1886-1893 (2006)
6. Yamashita A., Takada, T., Yamamoto G., and Torii R. Stable maintenance of monkey embryonic stem cells in the absence of bFGF. Transplant. Proc. 38, 1614-1615 (2006)
7. Yamashita A., Takada, T., Nemoto, K., Yamamoto G., and Torii R. Transient suppression of PPARγ directed ES cells into an osteoblastic linage. FEBS Lett., 580, 4121-4125 (2006)

2．学会発表

1. 山下晃弘、高田達之、尾松万里子、松浦博、山本学、鳥居隆三、カニクイザルES細胞の脂肪細胞への分化誘導、第５回日本再生医療学会、岡山 2006 (3/8-9)
2. Abdelalim E., Takada T., Toyoda F., Omatsu-Kanbe M., Matsuura H., Tooyama I., and Torii R., Monkey embryonic stem cell-derived cardiomyocytes express natriuretic peptides *in vitro*. 第５回日本再生医療学会、岡山 2006 (3/8-9)
3. 高田 達之、サル ES 細胞と再生医学、第５３回日本実験動物学会、神戸、公開シンポジウム「再生医学の現状と今後の展望」2006 (5/13)
4. 山下晃弘、高田達之、鳥居隆三、サル ES 細胞の骨芽細胞および脂肪細胞への分化誘導、第５３回日本実験動物学会、神戸 2006 (5/13)
5. Abdelalim E., Takada T., Toyoda F., Omatu-Kanbe M., Matsuura H., Tooyama I., and Torii R. Differentiation of Monkey Embryonic Stem Cells into Functional Cardiomyocytes *In Vitro*. 第５３回日本実験動物学会、神戸 2006 (5/13)

6. 高田達之、山下晃弘、鳥居隆三、PPARγ-siRNA を用いたES細胞の骨芽細胞への分化、日本分子生物学会フォーラム、口頭発表採択、名古屋 2006 (12/6-8)

7. Abdelalim E., Takada T., Tooyama I., and Torii R. Expression of ANP and BNP in monkey embryonic stem cell-derived cardiomyocytes. 20th IUBMB International Biochemistry and Molecular biology and 11th FAOBMB Congress Kyoto, Japan June 18-23, 2006

8. Takada T., Nemoto K., Yamashita A., Kobayashi M., Kato M., Kondo Y., and Torii R. siRNA mediated efficient gene silencing of Oct4 and cell differentiation in mouse and monkey ES cells. 4th Annual Meeting of the International Society for Stem Cell Research, Toronto, Ontario, Canada, June 29-July 1, 2006

9. Yamashita A., Takada T., Omatsu-Kanbe M., Nemoto K., Matsuura H., Yamamoto G., and Torii R. Osteoblast and adipocyte differentiation of monkey embryonic stem cells *in vitro*. 4th Annual Meeting of the International Society for Stem Cell Research, Toronto, Ontario, Canada, June 29-July 1, 2006

研究成果の刊行に関する一覧表

1) Naoki Osada, Katsuyuki Hashimoto, Momoki Hirai, Jun Kusuda. Aberrant termination of reproduction-related TMEM30C transcripts in the hominoids. Gene in press (2007)

2) Hisayuki Nomiyama, Kaori Otsuka-Ono, Retsu Miura, Naoki Osada, Keiji Terao, Osamu Yoshie, Jun Kusuda. Identification of a Novel CXCL1-Like Chemokine Gene in Macaques and its Inactivation in Hominids. *J Interferon Cytokine Res.* 27:32-37 (2007).

3) Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, Chou CK, Tsai SF, Wu CI, Shen CK. Rate of Evolution in Brain-Expressed Genes in Humans and Other Primates. PLoS Biol. 2006 Dec 26;5:e13 [Epub ahead of print]

4) Cheong J, Yamada Y, Yamashita R, Irie T, Kanai A, Wakaguri H, Nakai K, Ito T, Saito I, Sugano S, Suzuki Y. Diverse DNA methylation statuses at alternative promoters of human genes in various tissues. DNA Res. 13:155-167 (2006)

5) Takeda J, Suzuki Y, Nakao M, Barrero RA, Koyanagi KO, Jin L, Motono C, Hata H, Isogai T, Nagai K, Otsuki T, Kuryshev V, Shionyu M, Yura K, Go M, Thierry-Mieg J, Thierry-Mieg D, Wiemann S, Nomura N, Sugano S, Gojobori T, Imanishi T. Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. Nucleic Acids Res. 34:3917-3928 (2006)

6) Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K, Sugano S. DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. 34:D86-89 (2006)

7) Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase

8) T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 16: 55-65 (2006)

9) Takada T., Suzuki Y., Kadota N., Kondo Y., and Torii R. Generation of GFP-expressing monkey ES cells. Methods in Molecular Biology. Humana Press Inc. Embryonic Stem Cell Protocols, Turksen K., ed. 2nd ed. 329, 305-312 (2006)

10) Tsukada, H., Takada, T., Shiomi, N., Tani, T., and Torii, R. Acidic fibroblast growth factor promotes hepatic differentiation of monkey embryonic stem cells. In vitro Cell. Dev-An. 42, 83-88 (2006)

11) Abdelalim, E., Takada, T., Toyoda, F., Omatsu-Kanbe, M., Matsuura, H., Tooyama, I., and Torii, R. In vitro expression of natriureic peptides in cardiomyocytes differentiated from monkey embryonic stem cells. Biochem. Biophys. Res. Commun. 340, 689-695 (2006)

12) Yamashita A., Takada, T., Omatsu-Kanbe, M., Nemoto, K., Matsuura, H., Yamamoto G., and Torii R. Monkey embryonic stem cells differentiate into adipocytes in vitro. Cloning Stem Cells 8, 3-9 (2006)

13) Abdelalim, E., Takada, T., Torii, R. and Tooyama, I. Molecular cloning of BNP from heart and its immunohistochemical localization in the hypothalamus of monkey. Peptides 27, 1886-1893 (2006)

14)Yamashita A., Takada, T., Yamamoto G., and Torii R. Stable maintenance of monkey embryonic stem cells in the absence of bFGF. Transplant. Proc. 38, 1614-1615 (2006)

15) Yamashita A., Takada, T., Nemoto, K., Yamamoto G., and Torii R. Transient suppression of PPAR directed ES cells into an osteoblastic linage. FEBS Lett., 580, 4121-4125 (2006)

ARTICLE IN PRESS
+ MODEL
GENE-35703; No of Pages 6

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Gene xx (2007) xxx–xxx

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids

Naoki Osada [a,*], Katsuyuki Hashimoto [a], Momoki Hirai [b], Jun Kusuda [a]

[a] *Division of Biomedical Research Resources, National Institute of Biomedical Innovation, 7-6-8 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan*
[b] *International Research and Educational Institute for Integrated Medical Sciences, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666, Japan*

Received 14 July 2006; received in revised form 28 September 2006; accepted 28 November 2006

Abstract

Finding genetic novelties that may contribute to human-specific physiology and diseases is a key issue of current biomedical studies. *TMEM30C* is a gene containing two transmembrane (TM) domains and homologous to the yeast CDC50 family, which is related to polarized cell division. It is conserved among mammals along with two other paralogs, *TMEM30A* and *TMEM30B*. We found that *TMEM30C* is expressed specifically in the testis of mammals, in contrast to the relatively wide expression distributions of the other paralogs. While macaques expressed two alternative splicing isoforms which include one or two TM domains, humans and chimpanzees predominantly expressed truncated transcripts because of the mutations in the splicing and/or poly(A) signal sites. The major transcript in humans harbored non-stop ORF (open reading frame) while the chimpanzee counterpart encoded a protein with one TM domain. The difference was due to the 1-bp indel upstream of the poly(A) signal site. In addition, both the hominoids expressed minor transcripts encoding short proteins with one TM domain. Phylogenetic analysis has showed the acceleration of amino acid substitution after the human and chimpanzee divergence, which may have been caused by a recent relaxation in functional constraints or positive selection on *TMEM30C*. Elucidating the precise reproductive function of *TMEM30C* in mammals will be important to the foundation of divergence in higher primates at a molecular level.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Primates; Evolution; Transmembrane protein; Reproduction

1. Introduction

Genetic novelties that may contribute to the physiology and diseases of current human populations have fascinated many scientists for many years and have become an important subject of current biomedical studies (Gibbons, 1998; Hacia, 2001; Olson and Varki, 2003; Varki and Altheide, 2005). Chimpanzees (common chimpanzees and bonobos) are the closest relatives of humans, and a draft genome of the common chimpanzee (*Pan troglodytes*), which will be enormously valuable to evolutionary studies, is now publicly available (Chimpanzee Sequencing and Analysis Consortium, 2005). Consequently, extensive research has focused on genetic differentiation among higher primates, mainly based on single nucleotide substitutions and a set of genes has been uncovered in which changes might represent adaptations (Enard et al., 2002b; Clark et al., 2003; Hellmann et al., 2003; Watanabe et al., 2004; Evans et al., 2005; Glazko et al., 2005; Mekel Bobrov et al., 2005; Nielsen et al., 2005).

Undoubtedly, amino acid substitution is not the only factor responsible for the phenotypic diversity of species. In addition to amino acid substitution, divergence at a transcriptional level, such as *cis*-regulated gene expression divergence and the gain-and-loss of genes, is likely to be an important factor in the genomic evolution of organisms to create phenotypic complexity (King and Wilson, 1975; Enard et al., 2002a; Heissig et al., 2005; Marques et al., 2005; Rockman et al., 2005). Although gains-and-losses of genes might cause a larger phenotypic effect than single amino acid substitutions (Olson and Varki, 2003),

Abbreviations: TM, transmembrane; ORF, open reading frame; UTR, untranslated region.

* Corresponding author. Tel.: +81 72 641 9811; fax: +81 72 641 9812.
E-mail address: nosada@nibio.go.jp (N. Osada).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.11.021

ARTICLE IN PRESS

fewer studies have described lineage-specific gains-and-losses of genes among higher primates (Chou et al., 1998; Stedman et al., 2004; Hahn and Lee, 2005; Hayakawa et al., 2005; Wang et al., 2006), with the exception of the frequent births and deaths of genes in large multi-copied gene families (Meyer Olson et al., 2003; Fortna et al., 2004; Gilad et al., 2005; Go et al., 2005). Using the human and chimpanzee genome sequence, we could conduct a genome-wide survey of species-specific pseudogenes, which carry null mutations in the coding region either in the human and chimpanzee (Hahn and Lee, 2005; Wang et al., 2006). However, if the loss of function arose enough recently in the course of evolution, null mutation in the coding region is not a perfect criterion for nonfunctional genes. For example, mutation in a promoter region can lead null expression of the gene without coding disruption. Therefore, to explore lineage-specific transcript structures and expression patterns, valid and detailed experimental evidence is needed.

Recent studies have shown that the rate of transcriptome divergence was high in the testis, which might be involved in many biologically important phenomena such as reproductive isolation (Swanson and Vacquier, 2002). In the course of our transcriptome analysis of the macaque testis (Osada et al., 2005), we found that one of the genes, which showed homology to yeast gene CDC50, was expressed in macaques but not completely expressed in humans. In mammals, three types of CDC50 paralogs have been identified: transmembrane protein 30A (*CDC50A* or *TMEM30A*), 30B (*CDC50B* or *TMEM30B*), and 30C (*CDC50C* or *TMEM30C*). The CDC50 gene product, Cdc50p in yeast is a subunit of phospholipids-transacting P-type ATPase, which has been implicated in the asymmetrical localization of phospholipids within the plasma membrane (Saito et al., 2004) and hence is involved in polarized cell division system. The protein structure of this protein family is highly conserved from yeast to mammals; the structure comprises two transmembrane (TM) regions, with the head and tail sticking out on the cytoplasmic side. Although all three paralogous cDNA clones were present in mice, only *TMEM30A* and *30B* cDNAs have been cloned in humans, and *TMEM30C* has been predicted by only *in silico* analysis (Katoh and Katoh, 2004). Here, we describe that the *TMEM30C* gene is specifically expressed in mammalian testes and the transcript structure is highly diverse among higher primates.

2. Materials and methods

2.1. Sequence analysis

The putative coding sequence of chimpanzee *TMEM30C* was deduced from the chimpanzee draft genome sequence with the exception of the exon 6 sequence, which contained ambiguous base pairs and gaps in the public genome sequence. The sequence of chimpanzee exon 6 (DDBJ/EMBL/Genbank accession number: AB247157) was determined by sequencing the PCR product from genomic DNA, which was extracted from EB-transformed lymphocytes. The *TMEM30C* cDNA sequences of the mouse (AK161475), rat (XM_221533), dog (XM_545073), and bovine (BC111328) were obtained from the public databases. Note that the rat and dog sequences are hypothetical sequences that were predicted from the genome sequence. Human and chimpanzee *TMEM30C* cDNAs were amplified using the 3′-RACE method with an oligo-primer adaptor (GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTT) and forward primers in exon 2. The PCR products were cloned into pUC18 plasmids and the sequences of ≈10 clones from each library were then determined (AB249666; AB250297; AB265818; AB265819). *Macaca fascicularis* cDNA clones, QtsA-12626 (AB070082) and QtsA-16374 (AB070082), were isolated from the oligo-capped cDNA library described previously (Osada et al., 2005). The DNA samples were sequenced using an ABI 310 and 3730 sequencer (Applied Biosystems).

2.2. RT-PCR

The templates of the human and mouse total RNA were purchased from Clontech and Sawady technology (Japan), respectively. Chimpanzee testis sample was collected from an eight-year-old male chimpanzee (*P. troglodytes verus*) which died of natural causes. Total RNA samples of the cynomolgus monkey were obtained as described previously (Osada et al., 2005). One microgram of total mRNA was amplified using a One Step RNA PCR Kit (TakaraBio). The temperature and time schedules were 30 cycles of 94 °C for 20 s, 58 °C/60 °C for 30 s, and 72 °C for 1 min. The primer sequences were shown in Supplemental Table 1.

2.3. Tree construction of mammalian TMEM30C genes

For the phylogenetic analysis, we used 5′-sequences of ORFs (843 bp from the first ATG) that encode TMEM30C in the various mammals described above. The putative human and chimpanzee *TMEM30C* sequences corresponding to the other mammalian cDNAs were extracted from the human (NCBI build 35) and chimpanzee (NCBI build 1) genome sequence, respectively. The nucleotide sequences were aligned taking the translated amino acid sequences as guides and using ClustalW with default parameters (Thompson et al., 1994). The multiple alignments were used to construct a phylogenetic tree using the neighbor-joining method with Kimura's distance (Kimura, 1980; Saitou and Nei, 1987). The alignment and tree construction were performed using the MEGA 3.1 program package (Kumar et al., 2004). For the statistical test for positive selection, we used both PAML (Yang, 1997) and Hyphy (Pond et al., 2005) program packages but both programs failed to find both positively selected lineages and sites. The non-synonymous substitution rate (d_N) and the synonymous substitution rate (d_S) for each lineage were estimated using the maximum likelihood method implemented in PAML (Yang, 1997).

2.4. Tree construction of vertebrate CDC50 homologs

The amino acid sequences were used for the alignment and tree construction of vertebrate TMEM30 proteins and yeast CDC50. The phylogenetic tree was constructed using MEGA 3.1 program with p-distance and neighbor-joining method.

Please cite this article as: Osada, N. et al. Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids. Gene (2007), doi:10.1016/j.gene.2006.11.021

ARTICLE IN PRESS

3. Results and discussion

3.1. TMEM30 protein family in mammals

TMEM30C cDNA had not been identified in human transcripts; we found two clones named QtsA-12626 and QtsA-16374 in a cDNA library from cynomolgus monkey testis (*M. fascicularis*). The cDNAs were derived from seven (variant 1) and eight (variant 2) exons when they were mapped on the human genome sequence, and encoded 292 and 344 amino acids, respectively (Fig. 1). Variant 2 arose from an alternative splicing at a cryptic donor site of exon 7 of variant 1. Transmembrane regions of *TMEM30C* were predicted using the DAS-TMfilter program (Cserzo et al., 2002). The prediction showed that variant 2 has two transmembrane (TM) domains in exons 2 and 8 (2-TM type) but that variant 1 lacks the TM region at the C-terminal (1-TM type). Here, in Fig. 1 and the following results, we shall number the exons and introns using these macaque transcripts. We also found mouse and bovine *TMEM30C* homologous cDNAs in public databases. Mouse *tmem30c* had both 1-TM and 2-TM transcripts, but the 1-TM type *TMEM30C* transcript was not found in bovines.

3.2. Divergence of TMEM30C transcript structures in primates

We cloned cDNAs of *TMEM30C* from the human and chimpanzee testis RNA samples using the 3′-RACE method and determined the cDNA sequences. In both species, majority of the transcripts were spliced at the end of exon 3 but not at the end of exon 4. The human transcript contained a truncated protein-coding sequence with no stop codon in the reading frame and hence was assumed to be not protein-producing (variant 3a, AB249666). On the other hand, the chimpanzee transcript encoded a 157-amino acid protein (variant 3b, AB265818). The difference of coding ability between humans and chimpanzees was due to the 1-bp indel upstream of the poly(A) signal site (Fig. 1). Because the multiple indels were found in the region among higher primates, we were not able to infer whether the common ancestor of humans and chimpanzees had expressed protein-coding transcripts. We also cloned minor transcripts which were not spliced at the end of exon 3 and possessed a poly (A) tail in the following region, yielding 113-amino acid proteins. The minor transcripts have slightly different poly(A) addition sites in the 3′ UTRs (untranslated regions) between humans (variant 4a, AB250297) and chimpanzees (variant 4b, AB265819).

We subsequently surveyed the genomic sequence of exon–intron boundaries of humans, chimpanzees, and rhesus macaques (*Macaca mulatta*) using the public genome sequence databases and confirmed the substitutions by sequencing the human and chimpanzee genomic DNA samples. We found four major substitutions which might be responsible for the hominoid-specific truncation and human-specific non-stop ORF of *TMEM30C* transcripts (Fig. 1). First, the major transcripts of humans and chimpanzees had the 1-bp indel upstream of the poly (A) signal site. Second, poly(A) signal at macaque intron 4 was consensus AATAAA in humans and chimpanzees but GATAAA in macaques, suggesting that the poly(A) signals in humans and

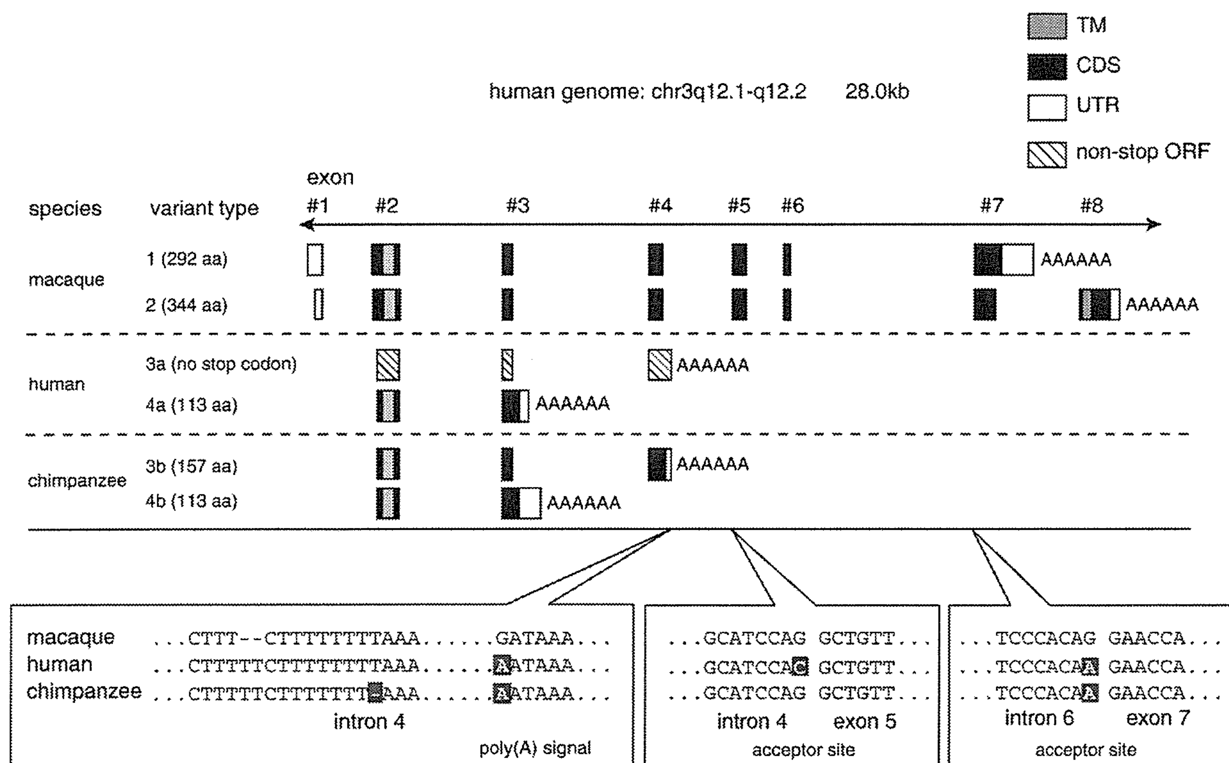

Fig. 1. Schematic presentation of the genomic structure of human and macaque *TMEM30C*. Black, white, shaded, and hatched boxes represent exons in coding regions (CDSs), untranslated regions (UTRs), transmembrane (TM) regions, and non-stop open reading frames (ORFs), respectively. Stretches of A indicate poly(A) tail. As indicated in the lower panel, poly(A) signals and splicing acceptor sites have human- and hominoid-specific substitutions, which are marked by shaded backgrounds. Note that the 5′-end of human *TMEM30C* is not cloned but apparently transcribed.

Please cite this article as: Osada, N. et al. Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids. Gene (2007), doi:10.1016/j.gene.2006.11.021

ARTICLE IN PRESS

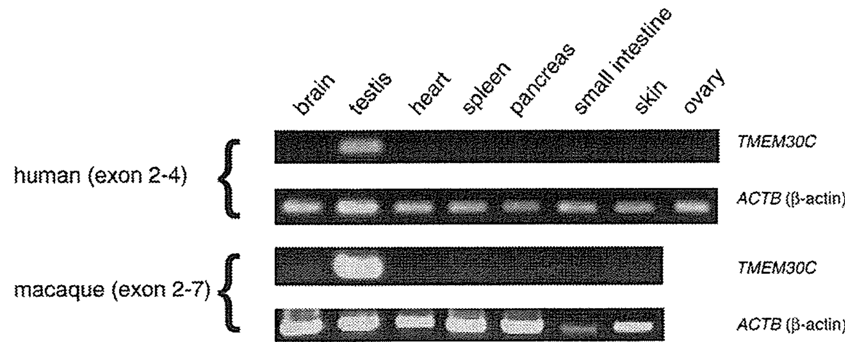

Fig. 2. Tissue distribution of human and macaque *TMEM30C* transcripts. Images of the RT-PCR gels show the amplification of *TMEM30C* and β-actin (control) transcripts from the brain, testis, heart, spleen, pancreas, small intestine, and skin for the human and macaque and ovary for the human.

chimpanzees were stronger than those in macaques. Third, splice acceptor site of exon 5 was substituted from AG to AC in humans while the chimpanzee and macaque have AG. In the human SNP database by HapMap project (Hinds et al., 2005), there was no inference of polymorphisms at the acceptor site (i.e. C was fixed). Fourth, splice acceptor site of exon 7 was AG in macaques but AA in the hominoids.

3.3. Expression pattern of TMEM30C in primates

We performed RT-PCR using several human, macaque and mouse organs to examine the expression pattern of mammalian *TMEM30C* orthologs. As a result, in the three species of mammals that we surveyed, *TMEM30C* was expressed exclusively in the testes (Fig. 2 and Supplemental Fig. 1A). In addition, a very faint amplification was found in the macaque brain.

3.4. Molecular evolution of TMEM30C in mammals

To determine the pattern of molecular evolution of *TMEM30C*, we conducted a molecular phylogenetic analysis of mammalian *TMEM30C*. Mouse, rat, bovine, dog, macaque, chimpanzee, and human coding sequences of *TMEM30C* were obtained by DNA sequencing and from public databases (see Methods section). While the latter half of the human and chimpanzee *TMEM30C* gene is not transcribed, it does not carry any null mutation such as a frameshift or stop codon in the putative coding region. Aligned regions (843 bp from the first ATG) were then used to construct a phylogenetic tree using the neighbor-joining method. The phylogenetic tree and substitution rate for each lineage are presented in Fig. 3. The functional constraint on the gene was inferred using the statistics ω (d_N/d_S). ω is close to 0 when the functional constraint on the gene is strong and approaches 1 when the constraint is weak (i.e. pseudogene when $\omega=1$). The ω for the entire tree was relatively high (0.45) for *TMEM30C*, and ω after the divergence of humans and chimpanzees was extremely high compared with the other mammalian lineages. Indeed, there are no synonymous substitutions between human and chimpanzee *TMEM30C*, but there are four non-synonymous substitutions. Although ω after the human–chimpanzee divergence exceeded 1, it was not significantly deviated from 1 when examined using the likelihood ratio test. Because of the few nucleotide substitutions

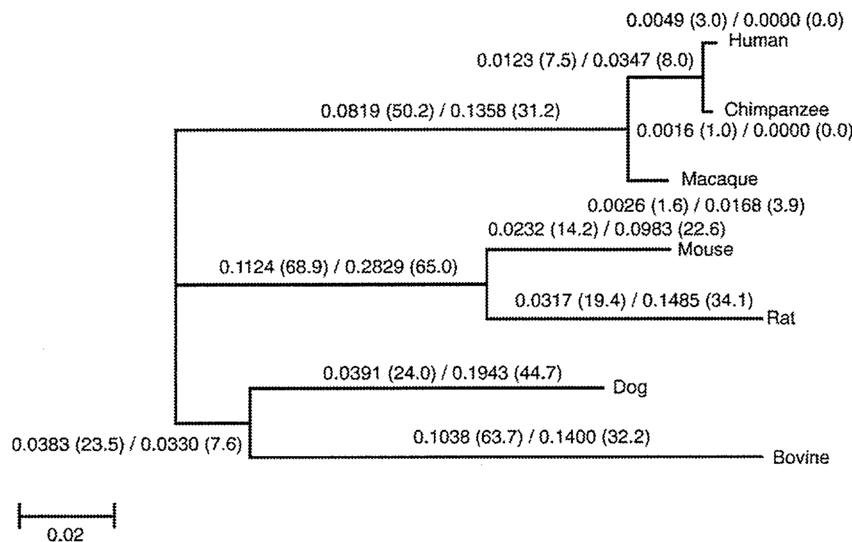

Fig. 3. A phylogenetic tree of mammalian *TMEM30C*. The tree was constructed based on the 843 bp nucleotide sequences from the initial ATG codon. Branch length was estimated using Kimura's 2 parameter method (Kimura, 1980). d_N/d_S for each branch was estimated using the PAML (Yang, 1997) and is presented along the branches. Numbers of substitutions are shown in parentheses.

Please cite this article as: Osada, N. et al. Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids. Gene (2007), doi:10.1016/j.gene.2006.11.021

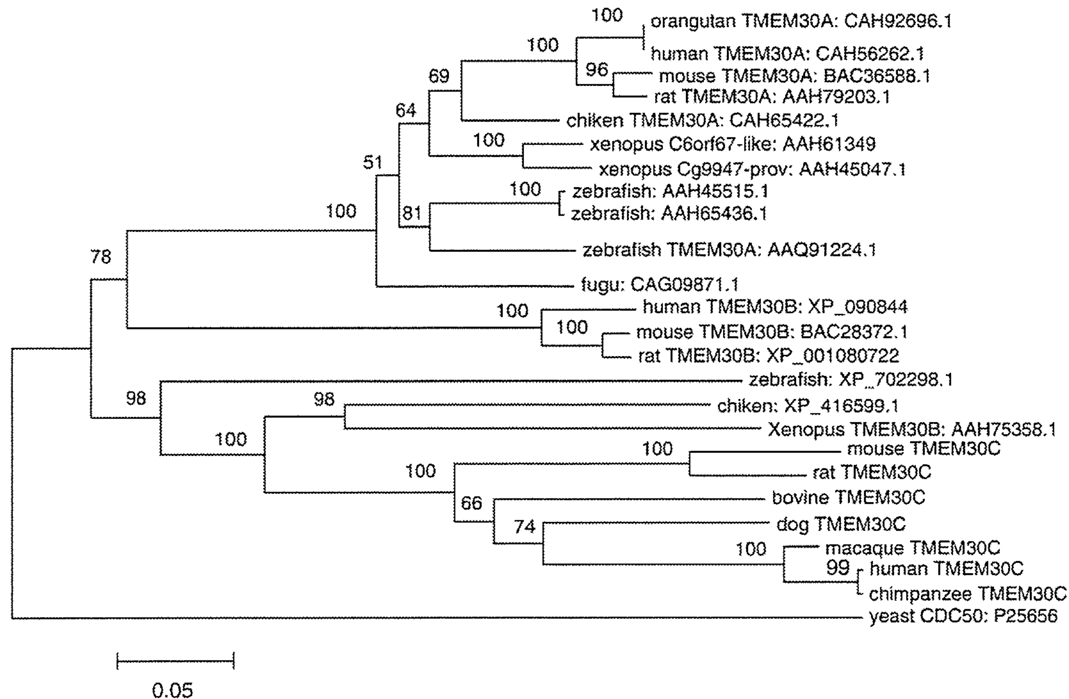

Fig. 4. Neighbor-joining tree of vertebrate TMEM30 proteins. Yeast CDC50 protein sequence was used as an out-group. Bootstrap values (%) by 1000 iteration are indicated upon each branch. DDBJ/EMBL/Genbank accession numbers are shown after each gene name. Gene name was omitted when the gene is a hypothetical (unnamed) gene.

between humans and chimpanzees, we were not able to infer whether the excess of non-synonymous substitutions was due to positive selection or relaxation in functional constraint. However, we suggest that the relaxation in functional constraint is a more reasonable explanation because the majority of the coding regions in the hominoids are not transcribed.

3.5. Evolutionary origin of TMEM30C

TMEM30C is one of three TMEM30 paralogs in mammalian genomes. In order to infer the original function of *TMEM30C*, we surveyed patterns of gene expression and protein evolution in the TMEM30 gene family. First, we performed RT-PCR for *TMEM30A* and *30B* in human tissues. *TMEM30A* and *30B* in mammals have wide ranges of expression patterns (brain, liver, and testis) and only *TMEM30C* showed tissue specific expression patterns (Supplemental Fig. 1B). Next, the public database was searched for TMEM30 protein sequences of various vertebrate species. We aligned the sequences and constructed a phylogenetic tree. The tree indicated that *TMEM30C* arose before the divergence of the teleost fish (Fig. 4). *TMEM30B* is an intronless gene in mice and rats, and the gene in humans apparently carries one intron insertion in the 3′ UTR. Thus the gene is thought to have originated by retrotransposition. Considering that yeast Cdc50p has an indispensable function in polarized cell division, ubiquitously expressed TMEM30A probably retains the original function of CDC50 proteins. Therefore, *TMEM30C* likely originated with *TMEM30A* by gene duplication and later acquired a testis-specific expression pattern. Interestingly, although all mammals possess only one copy of *TMEM30A*, *Xenopus* and zebra fish have multiple copies of TMEM30A-like genes, indicating the convergent expansion of *TMEM30A* genes in amphibians and teleosts.

The function of the TMEM30C protein remains unclear. The TMEM30C protein is probably involved in spermatogenesis in mammals because yeast Cdc50p has been shown to be related to cell division. In this report, we showed the complex transcriptional divergence of *TMEM30C* in primates. Humans and chimpanzees predominantly express non-coding (variant 3a) and 1-TM type (variant 3b) of *TMEM30C* transcripts, respectively. Because the testis is one of the organs that have the widest variety of splicing isoforms (Yeo et al., 2004), other minor transcripts may be hidden at a very low expression level. Actually, we confirmed a weak expression of exon 5 in chimpanzees but not in humans by RT-PCR (data not shown), which might result from the human-specific acceptor site mutation at the boundary of intron 4 and exon 5. However, using several sets of primers, we were not able to identify the expression of exon 5 to 8 in humans and exon 6 to 8 in chimpanzees, suggesting that the 2-TM type *TMEM30C* might become dispensable for hominoids. As shown in Fig. 1, macaques express both 1-TM (variant 1) and 2-TM (variant 2) type *TMEM30C* transcripts by alternative splicing. We also found both types of transcripts in mice, suggesting that the function of 1-TM type *TMEM30C* had already existed before the divergence of mammals. Hence, this phenomenon can be interpreted as a reduction in transcriptional diversity after the divergence of hominoids. We should note, however, that the loss of the 2-TM type *TMEM30C* transcript is not the only factor that might alter the function of human *TMEM30C* as hominoid 1-TM type *TMEM30C* also lost most of

Please cite this article as: Osada, N. et al. Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids. Gene (2007), doi:10.1016/j.gene.2006.11.021

ARTICLE IN PRESS

its extracellular region, possibly resulting in a substantially modified function.

4. Conclusions

In this study, we identified the hominoid-specific truncated transcripts of *TMEM30C*, which is specifically expressed in the testis. From the aspect of species divergence at a transcriptional level, this example is only the tip of the iceberg and further effort is needed to survey the genetic divergence of species more deeply.

Acknowledgements

We thank Ikuo Hayasaka (Sanwa Kagaku), Yutaka Suzuki and Sumio Sugano (The University of Tokyo) for helping with the collection of the chimpanzee cDNA samples. We also thank two anonymous reviewers for helpful suggestions. This study was supported in part by a Health Science Research grant from the Human Genome Program of the Ministry of Health, Labor and Welfare of Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.11.021.

References

Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87.

Chou, H.H., et al., 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo–Pan divergence. Proc. Natl. Acad. Sci. U. S. A. 95, 11751–11756.

Clark, A.G., et al., 2003. Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. Science 302, 1960–1963.

Cserzo, M., Eisenhaber, F., Eisenhaber, B., Simon, I., 2002. On filtering false positive transmembrane protein predictions. Protein Eng. 15, 745–752.

Enard, W., et al., 2002a. Intra- and interspecific variation in primate gene expression patterns. Science 296, 340–343.

Enard, W., et al., 2002b. Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418, 869–872.

Evans, P.D., et al., 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309, 1717–1720.

Fortna, A., et al., 2004. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS. Biol. 2, e207.

Gibbons, A., 1998. Which of our genes makes us human? Science 281, 1432–1434.

Gilad, Y., Man, O., Glusman, G., 2005. A comparison of the human and chimpanzee olfactory receptor gene repertoires. Genome. Res. 15, 224–230.

Glazko, G., Veeramachaneni, V., Nei, M., Makalowski, W., 2005. Eighty percent of proteins are different between humans and chimpanzees. Gene 346, 215–219.

Go, Y., Satta, Y., Takenaka, O., Takahata, N., 2005. Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. Genetics 170, 313–326.

Hacia, J.G., 2001. Genome of the apes. Trends Genet. 17, 637–645.

Hahn, Y., Lee, B., 2005. Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. Bioinformatics 21 (Suppl. 1), i186–i194.

Hayakawa, T., Angata, T., Lewis, A.L., Mikkelsen, T.S., Varki, N.M., Varki, A., 2005. A human-specific gene in microglia. Science 309, 1693.

Heissig, F., Krause, J., Bryk, J., Khaitovich, P., Enard, W., Paabo, S., 2005. Functional analysis of human and chimpanzee promoters. Genome Biol. 6, R57.

Hellmann, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B., Paabo, S., 2003. Selection on human genes as revealed by comparisons to chimpanzee cDNA. Genome Res. 13, 831–837.

Hinds, D.A., et al., 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307, 1072–1079.

Katoh, Y., Katoh, M., 2004. Identification and characterization of CDC50A, CDC50B and CDC50C genes in silico. Oncol. Rep. 12, 939–943.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science 188, 107–116.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A., Kaessmann, H., 2005. Emergence of young human genes after a burst of retroposition in primates. PLoS. Biol. 3, e357.

Mekel Bobrov, N., et al., 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. Science 309, 1720–1722.

Meyer Olson, D., et al., 2003. Analysis of the TCR beta variable gene repertoire in chimpanzees: identification of functional homologs to human pseudogenes. J. Immunol. 170, 4161–4169.

Nielsen, R., et al., 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS. Biol. 3, e170.

Olson, M.V., Varki, A., 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. Nat. Rev., Genet. 4, 20–28.

Osada, N., et al., 2005. Substitution rate and structural divergence of 5′UTR evolution: comparative analysis between human and cynomolgus monkey cDNAs. Mol. Biol. Evol. 22, 1976–1982.

Pond, S.L., Frost, S.D., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.

Rockman, M.V., Hahn, M.W., Soranzo, N., Zimprich, F., Goldstein, D.B., Wray, G.A., 2005. Ancient and recent positive selection transformed opioid *cis*-regulation in humans. PLoS. Biol. 3, e387.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Saito, K., Fujimura Kamada, K., Furuta, N., Kato, U., Umeda, M., Tanaka, K., 2004. Cdc50p, a protein required for polarized growth, associates with the Drs2p P-type ATPase implicated in phospholipid translocation in *Saccharomyces cerevisiae*. Mol. Biol. Cell 15, 3418–3432.

Stedman, H.H., et al., 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428, 415–418.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Varki, A., Altheide, T.K., 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res. 15, 1746–1758.

Wang, X., Grus, W.E., Zhang, J., 2006. Gene losses during human origins. PLoS. Biol. 4, e52.

Watanabe, H., et al., 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 429, 382–388.

Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. Nat. Rev., Genet. 3, 137–144.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yeo, G., Holste, D., Kreiman, G., Burge, C.B., 2004. Variation in alternative splicing across human tissues. Genome Biol. 5, R74.

Please cite this article as: Osada, N. et al. Aberrant termination of reproduction-related *TMEM30C* transcripts in the hominoids. Gene (2007), doi:10.1016/j.gene.2006.11.021

JOURNAL OF INTERFERON & CYTOKINE RESEARCH 27:32–37 (2007)
© Mary Ann Liebert, Inc.
DOI: 10.1089/jir.2007.0099

Identification of a Novel CXCL1-Like Chemokine Gene in Macaques and Its Inactivation in Hominids

HISAYUKI NOMIYAMA,[1] KAORI OTSUKA-ONO,[1] RETSU MIURA,[1] NAOKI OSADA,[2] KEIJI TERAO,[3] OSAMU YOSHIE,[4] and JUN KUSUDA[2]

ABSTRACT

Chemokines are a rapidly evolving cytokine gene family. Because of various genome rearrangements after divergence of primates and rodents, humans and mice have different sets of chemokine genes, with humans having members outnumbering those of mice. Here, we report the occurrence of lineage-specific chemokine gene generation or inactivation events within primates. By using human chemokine sequences as queries, we isolated a novel cynomolgus macaque CXC chemokine cDNA. The encoded chemokine, termed CXCL1L (from CXCL1-like) showed the highest similarity to human CXCL1. A highly homologous gene was also found in the rhesus macaque genome. By comparing the genome organization of the major CXC chemokine clusters among the primates, we found that one copy of the duplicated CXCL1 genes turned into a pseudogene in the hominids, whereas the gene in macaques has been maintained as a functionally active CXCL1L. In addition, cynomolgus macaque was found to contain an additional CXC chemokine highly homologous to CXCL3, termed CXCL3L (from CXCL3-like). These results demonstrate the birth-and-death process of a new gene in association with gene duplication within the primates.

INTRODUCTION

Cynomolgus macaque (*Macaca fascicularis*) and rhesus macaque (*Macaca mulatta*) are closely related old world monkeys commonly used in experimental and toxicologic studies for drug and vaccine development.[1–4] Although both macaques are considered phylogenetically very close to humans, possible genetic differences between macaques and humans that may cause interspecies differences in drug responses and toxicity should be taken into account when the data obtained from macaques are extrapolated to humans. To unveil the genetic differences in primates and also to help identify genes in the human genome, expressed sequence tag (EST) and genome sequencing projects of these macaques are underway.

Chemokines are a large family of cytokines that regulate inflammation, leukocyte trafficking, and immune cell development.[5–7] There are at least 46 chemokine members in humans. Based on the arrangement of the conserved cysteine residues, chemokines are classified into four subfamilies. Two main subfamilies are CXC and CC chemokines, which have the first two conserved cysteines separated by one amino acid or juxtaposed, respectively. Chemokines can also be divided into two functional subgroups. Inflammatory chemokines attract mainly monocytes and neutrophils and mediate innate immunity, whereas homeostatic chemokines are constitutively expressed in organs, such as lymphoid tissues, and are involved in relocation of lymphocytes and dendritic cells (DCs).

The inflammatory CXC and CC chemokines are known to form a large gene cluster.[7] Human CXC and CC inflammatory chemokine gene clusters reside on chromosomes 4 and 17, respectively, and the respective gene clusters in mice are located on chromosomes 5 and 11. Comparison of these gene cluster organizations shows that the chemokine gene content in each cluster is greatly different between human and mouse due to lineage-specific gene duplication or deletion events or both after the divergence of primates and rodents.[8,9] In contrast, the

[1]Department of Molecular Enzymology, Kumamoto University Graduate School of Medical Sciences, Kumamoto 860-8556, Japan.
[2]Division of Biomedical Research Resources, National Institute of Biomedical Innovation, Ibaraki, Osaka 567-0085, Japan.
[3]Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Tsukuba, Ibaraki 305-0843, Japan.
[4]Department of Microbiology, Kinki University School of Medicine, Osaka-Sayama, Osaka 589-8511, Japan.
Sequence data from this paper have been submitted to the GenBank/EBI/DDBJ databases with accession nos. AB262775 (CXCL1), AB262776 (CXCL2), AB262777 (CXCL3), AB262778 (CXCL1L), and AB262779 (CXCL3L).

emergence of noncluster chemokines, most of which are homeostatic chemokines, apparently predated the divergence of primates and rodents.[8,9]

To determine if gene rearrangements within the chemokine clusters occurred even after the diversification of humans and nonhuman primates, we searched the EST and genome databases of nonhuman primates for novel chemokines.

MATERIALS AND METHODS

cDNA cloning

Chemokine cDNAs were cloned by reverse transcriptase-polymerase chain reaction (RT-PCR). Total RNAs were prepared from various tissues of cynomolgus macaque and reverse transcribed. The cDNAs were synthesized using PrimeSTAR HS DNA polymerase (Takara Bio, Kyoto, Japan) and cloned with Mighty Cloning kit (Takara Bio). The primer sequences were designed based on the rhesus macaque genome sequences and were CXCL1, 5′-CTCCAGCTCCTCGCACAG and 5′-AGCCACCAATGAGCTTCTTC; CXCL1L, 5′-AGTTCCCCTGCTCCTCTCAC and 5′-GCCAGTATTTCTGACCAACG; CXCL2, 5′-CCGAAACGCCTGCTGAG and 5′-CTTCAGGAACAGCCACCAAT; CXCL3 and CXCL3L, 5′-TCCCATCCTGCTGAG and 5′-CCGCAGGAAGTGTCAATGT. The cDNAs used in the cDNA cloning as templates are those of liver (CXCL1 and CXCL2), spleen (CXCL1L), and stomach (CXCL3 and CXCL3L). The PCR conditions were 30 cycles at 98°C for 10 sec, 60°C for 5 sec, and 72°C for 1 min.

Tissue expression analysis of CXCL1 and CXCL1L

The prepared cDNAs were amplified with Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). The primer sequences were CXCL1L, 5′-AGGGAATTCACCCCAAGAAC and 5′-GCAAACTCACCTGTTCAGCA; glyceraldehyde 3-phosphate dehydrogenase (GAPDH), 5′-GCCAAGGTCATCCATGACAACTTTGG and 5′-GCCTGCTTCACCACCTTCTTGATGTC. The primers used for CXCL1 were as described.

FIG. 1. Comparison of amino acid sequences of chemokines CXCL1, CXCL1L, CXCL2, CXCL3, and CXCL3L from human, chimpanzee, rhesus macaque, and cynomolgus macaque. Sequences other than those of cynomolgus macaque were taken from Ensembl (*www.ensembl.org/*) or UCSC Genome Browser (*genome.ucsc.edu/*). The amino-terminal sequence of the chimpanzee CXCL1 sequence is still unknown because of a gap in the genome sequence. Conserved four cysteine residues are boxed. ELR (Glu-Leu-Arg) motif is indicated by dots under the cynomolgus macaque CXCL3L sequence. Signal sequences are shown as lowercase letters. hs, human; pt, chimpanzee; mm, rhesus macaque; mf, cynomolgus macaque.

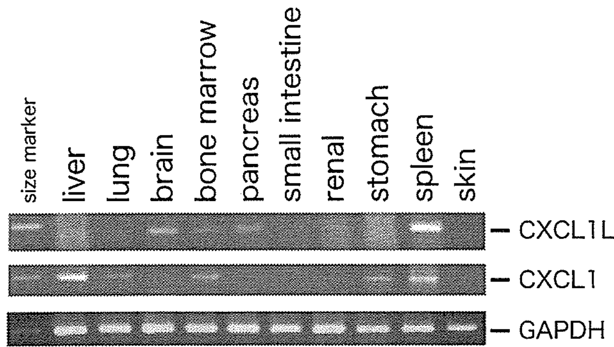

FIG. 2. RT-PCR analyses of CXCL1L and CXCL1 mRNAs in various cynomolgus macaque tissues. cDNAs were prepared from various tissues and amplified by PCR. GAPDH was used as an internal control.

The housekeeping gene GAPDH was used as an internal control. The PCR conditions were 35 cycles at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min. The product sizes were 170 bp (CXCL1L), 398 bp (CXCL1), and 313 bp (GAPDH).

Computational methods

Signal sequences were predicted by SignalP (*www.cbs.dtu.dk/services/SignalP/*). Dot-plot analysis was performed using PipMaker (*pipmaker.bx.psu.edu/pipmaker/*) with the all matches option. For phylogenetic analysis, the chemokine amino acid sequences were aligned with ClustalX (*bips.u-strasbg.fr/fr/Documentation/ClustalX/*). The Neighbor-Joining tree was constructed with PAUP* (*paup.csit.fsu.edu/*) using the protein-Poisson distances, and only >50% bootstrap values are shown at each node (1000 replications). CXCL1P (human and chimpanzee CXCL1 pseudogene, exons 1 plus 2 sequences) and CXCL7P1 (PPBPL1, human, chimpanzee and rhesus macaque CXCL7 pseudogene exon sequences) were also used in the tree construction. The codon frames were inferred from the CXCL1 and CXCL7 genes.

RESULTS AND DISCUSSION

When human chemokine sequences were used as queries, one novel chemokine sequence (clone QnpA-12174) was found in the cynomolgus macaque EST database (Japan National Institute of Biomedical Innovation JCRB Gene Bank, *genebank.nibio.go.jp/*). Because the cDNA was a chimeric and 5′-truncated clone, the rhesus macaque genome database[10] (UCSC Genome Browser Database, *genome.ucsc.edu/*) was searched with the sequence, and the corresponding gene was found to be located in the inflammatory CXC chemokine gene cluster on chromosome 5 (position 55,707,566–55,708,094, January 2006 assembly). Based on this gene sequence, a PCR primer

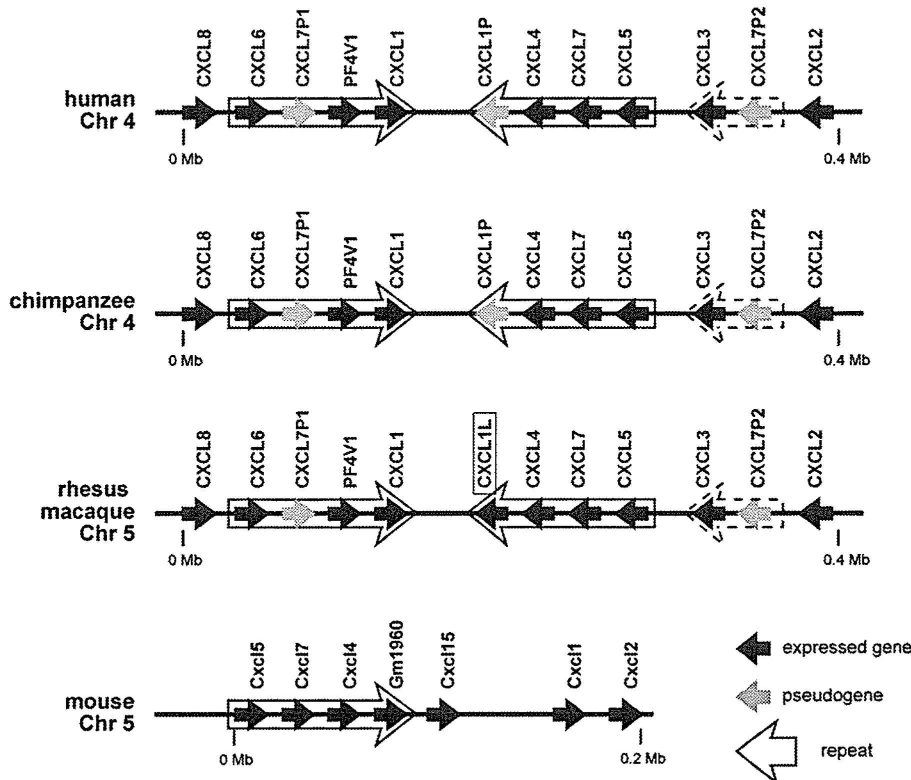

FIG. 3. Genome organization of CXC inflammatory chemokine gene clusters. Genome sequences taken from Ensembl or UCSC were analyzed and are shown schematically. Black and gray arrows denote functionally active gene and pseudogene, respectively, and arrowheads show the transcriptional orientation. Large arrows indicate duplicated regions.

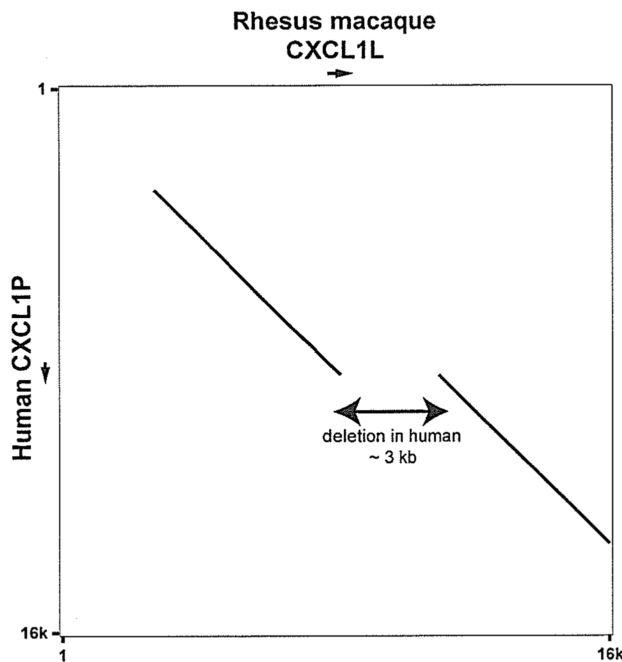

FIG. 4. Dot-plot analysis of the flanking sequences of the rhesus macaque CXCL1L gene and the human CXCL1 pseudogene (CXCL1P). The arrow with two heads shows a deletion of about 3 kb in length seen only in the human genome. The sizes of the genome sequences used were 16 kb for both species.

pair surrounding the coding sequence was prepared. A cDNA clone was isolated by PCR from a cDNA library prepared from cynomolgus macaque spleen. The cDNA encoded a polypeptide of 104 amino acid residues. As the encoded chemokine was highly similar to human CXCL1, CXCL2, and CXCL3, we also isolated cynomolgus macaque CXCL1, CXCL2, and CXCL3 cDNA clones from stomach or liver cDNA libraries for comparison (Fig. 1). The novel chemokine showed 80%, 76%, and 79% similarity to cynomolgus macaque CXCL1, CXCL2, and CXCL3, respectively, and, therefore, was called CXCL1L, from CXCL1-like. The corresponding rhesus macaque CXCL1L was identical to the cynomolgus macaque CXCL1L at the amino acid level but had one synonymous base substitution in the coding region. Interestingly, no orthologous gene for CXCL1L was found in other species, including hominids.

We then examined the expression of CXCL1L in various cynomolgus macaque tissues by semiquantitative RT-PCR. We also examined the expression of CXCL1 for comparison. CXCL1L mRNA was found to be expressed at high levels in spleen and to a lesser extent in brain, bone marrow, and pancreas (Fig. 2). In contrast, CXCL1 mRNA was expressed abundantly in liver, and the expression level in spleen was less than that in liver. Human CXCL1, CXCL2, and CXCL3 genes are likely to be expressed at an extremely low level in spleen according to the expression profiles (*www.ncbi.nlm.nih.gov/UniGene/*), and cynomolgus macaque CXCL1L and CXCL1 may have a unique function in the lymphoid organs.

When we prepared the cynomolgus macaque CXCL3 cDNA from stomach, a cDNA clone closely related to but slightly different from CXCL3 was also isolated. The cDNA encoded a chemokine of 107 amino acid residues that we named CXCL3L (from CXCL3-like) (Fig. 1). CXCL3L was just 5 amino acid residues shorter than CXCL3. Furthermore, although CXCL3L contained 13 base substitutions in the coding region, CXCL3L and CXCL3 differed by only 3 amino acid residues in the signal sequence. No counterpart of the CXCL3L gene was found in rhesus macaque or other primate species.

To obtain clues about the generation mechanism of the CXCL1L gene in macaques, we compared the maps of the inflammatory CXC chemokine gene clusters of human, chimpanzee, rhesus macaque, and mouse obtained from the websites (Ensembl and UCSC) (Fig. 3). The genome sequence of cynomolgus macaque is not available at present. In the primates, a large genome segment including four CXC chemokine genes was duplicated, forming an inverted repeat. Comparison of the maps shows that rhesus macaque CXCL1L gene locus corresponds to those of the CXCL1 pseudogene (CXCL1P) in the human and chimpanzee genomes. The pseudogene in both species contains exons 1 and 2 and the intron between them but lacks the downstream sequence.[11] To examine the genome organization at the nucleotide level, the sequences containing the rhesus macaque CXCL1L and human CXCL1P genes were compared by dot-plot analysis (Fig. 4). Figure 4 clearly shows that a deletion of approximately 3 kb in size is located 3′ downstream of the human CXCL1P gene. This result suggests that the human CXCL1P and rhesus macaque CXCL1L genes arose from a common ancestor gene and that a deletion event in the human genome made the gene inactive after the divergence of hominids and macaques.

Next, a phylogenetic tree was constructed to see the evolutionary relationship of the inflammatory CXC chemokines (Fig. 5). Protein coding sequences were used for the analysis, and the peptide sequences deduced from the human and chimpanzee CXCL1P sequences (exons 1 plus 2) were also included in the analysis for comparison. The tree shows that the macaque CXCL1L and human CXCL1P are indeed closely related, forming a unique branch, whereas the primate CXCL1, CXCL2, and CXCL3 form another relatively independent branch, suggesting the CXCL1L diverged from the common ancestor before CXCL1, CXCL2, and CXCL3 were generated.

Duplication of genes is recognized as the driving force of evolution, and multigene families evolve by a birth-and-death process of duplicated genes.[12] In the CXC chemokine gene cluster, CXCL1 and duplicated CXCL1L in the inverted repeat are both intact in macaques, whereas the latter gene became a pseudogene in hominids. Likewise, the CXCL7 gene in one copy of the repeat turned into a pseudogene CXCL7P1 in the primates (Fig. 3). In addition, another gene duplication appears to have occurred in cynomolgus macaque that generated the CXCL3L gene only in cynomolgus macaque. Given that inactivation of duplicated genes (CXCL1P and CXCL7P1) and emergence of a new gene (CXCL3L) are identified in different primate lineages, the birth-and-death process is still ongoing in the chemokine gene cluster of primates.

As the database search has not identified novel chemokine receptor genes in the macaque genome to date and macaque CXCL1L shows a high sequence similarity to human CXCL1, CXCL2, and CXCL3, the chemokine probably binds the chemokine receptor CXCR2 and has chemotactic activity for neutrophils. CXCL1L indeed contains the Glu-Leu-Arg (ELR)

FIG. 5. Phylogenetic tree of CXC inflammatory chemokines. Circles at the nodes indicate divergence points between primates and rodents, and triangles show the points where segmental duplications have occurred in the primate lineage. Chemokine sequences were taken from Ensembl, UCSC, or Cytokine Family Database (*cytokine.medic.kumamoto-u.ac.jp/*). Mouse Cxcl15 was used as an outgroup. hs, human; pt, chimpanzee; mm, rhesus macaque; mf, cynomolgus macaque; mus, mouse; rn, rat.

motif preceding the first conserved cysteine residue that is essential for receptor binding, neutrophil activation, and angiogenic activity.[13,14] However, because gene duplication also allows one copy of the duplicated genes to acquire a new function,[12] we cannot exclude the possibility that CXCL1L might bind other receptors. Expression of CXCL1L at high levels in macaque spleen also suggests a new function in this lymphoid organ (Fig. 2). Recently, some chemokines were shown to be ligands for both chemokine receptors and those not categorized as chemokine receptors.[15,16] Thus, it will be interesting to see if CXCL1L has receptors and functions differently from CXCL1, CXCL2, and CXCL3.

ACKNOWLEDGMENTS

This work is supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by MEXT. HAITEKU (2002–2006).

REFERENCES

1. Jacqz E, Billante C, Moysan F, Mathieu H. The non-human primate: a possible model for human genetically determined polymorphisms in oxidative drug metabolism. *Mol. Pharmacol.* 1988;34:215–217.
2. Thorgeirsson UP, Dalgard DW, Reeves J, Adamson RH. Tumor incidence in a chemical carcinogenesis study of nonhuman primates. *Regul. Toxicol. Pharmacol.* 1994;19:130–151.
3. Daddario-DiCaprio KM, Geisbert TW, Stroher U, Geisbert JB, Grolla A, Fritz EA, Fernando L, Kagan E, Jahrling PB, Hensley LE, Jones SM, Feldmann H. Post-exposure protection against Marburg haemorrhagic fever with recombinant vesicular stomatitis virus vectors in non-human primates: an efficacy assessment. *Lancet* 2006;367:1399–1404.
4. Misumi S, Nakayama D, Kusaba M, Iiboshi T, Mukai R, Tachibana K, Nakasone T, Umeda M, Shibata H, Endo M, Takamune N, Shoji S. Effects of immunization with CCR5-based cycloimmunogen on simian/HIVSF162P3 challenge. *J. Immunol.* 2006;176:463–471.
5. Moser B, Loetscher P. Lymphocyte traffic control by chemokines. *Nat. Immunol.* 2001;2:123–128.
6. Zlotnik A, Yoshie O. Chemokines: a new classification system and their role in immunity. *Immunity* 2000;12:121–127.
7. Yoshie O, Imai T, Nomiyama H. Chemokines in immunity. *Adv. Immunol.* 2001;78:57–110.
8. Nomiyama H, Mera A, Ohneda O, Miura R, Suda T, Yoshie O. Organization of the chemokine genes in the human and mouse major clusters of CC and CXC chemokines: diversification between the two species. *Genes Immun.* 2001;2:110–113.
9. Nomiyama H, Egami K, Tanase S, Miura R, Hirakawa H, Kuhara S, Ogasawara J, Morishita S, Yoshie O, Kusuda J, Hashimoto K. Comparative DNA sequence analysis of mouse and human CC chemokine gene clusters. *J. Interferon Cytokine Res.* 2003;23:37–45.
10. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ. The UCSC Genome Browser Database. *Nucleic Acids Res.* 2003;31:51–54.
11. Shattuck-Brandt RL, Wood LD, Richmond A. Identification and characterization of an MGSA/GRO pseudogene. *DNA Seq.* 1997;7:379–386.
12. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 2005;39:121–152.
13. Clark-Lewis I, Dewald B, Geiser T, Moser B, Baggiolini M. Platelet factor 4 binds to interleukin 8 receptors and activates neutrophils when its N terminus is modified with Glu-Leu-Arg. *Proc. Natl. Acad. Sci. USA* 1993;90:3574–3577.
14. Strieter RM, Burdick MD, Gomperts BN, Belperio JA, Keane MP. CXC chemokines in angiogenesis. *Cytokine Growth Factor Rev.* 2005;16:593–609.
15. Elagoz A, Henderson D, Babu PS, Salter S, Grahames C, Bowers L, Roy MO, Laplante P, Grazzini E, Ahmad S, Lembo PM. A truncated form of CKbeta8-1 is a potent agonist for human formyl peptide-receptor-like 1 receptor. *Br. J. Pharmacol.* 2004;141:37–46.
16. Nakayama T, Kato Y, Hieshima K, Nagakubo D, Kunori Y, Fujisawa T, Yoshie O. Liver-expressed chemokine/CC chemokine ligand 16 attracts eosinophils by interacting with histamine H4 receptor. *J. Immunol.* 2004;173:2078–2083.

Address reprint requests or correspondence to:
Dr. Hisayuki Nomiyama
Department of Molecular Enzymology
Kumamoto University Graduate School of Medical Sciences
Honjo 1-1-1
Kumamoto 860-8556
Japan

Tel: +81-96-373-5065
Fax: +81-96-373-5066
E-mail: nomiyama@gpo.kumamoto-u.ac.jp

Received 15 June 2006/Accepted 13 July 2006