論文内容の要旨

論文題目　**Database construction and comparative analysis of chimpanzee cDNA sequences**

（チンパンジー cDNA 配列のデータベース構築と比較解析）

氏名　　坂手　龍一

2003 年春にヒトゲノム配列の完全版が公開されるはこびとなった現在、数万個と考えられているヒト全遺伝子のカタログ作成と、その翻訳産物であるタンパク質およびそれらの相互作用といった、ポストゲノミクスの研究が加速している。ひとつの遺伝子配列の一塩基に置換が起こっただけでも、機能や表現型に変化が現れることがあるが、遺伝子配列の相違点について種間で比較解析をおこなうことは、遺伝子機能の解析には非常に有効である。そこで、ヒトに最も近い種であるチンパンジーの遺伝子配列が得られれば、そこから遺伝子機能の比較解析が可能になり、ひいてはヒトをヒトたらしめている遺伝子とその働きを知ることができると考えられる。現在、チンパンジーについてはゲノム配列のシークエンシングが国際プロジェクトとして進行しており、ヒトとの配列の一致度が 98.77%であることが報告されている。しかし、転写・発現している配列についての情報はほとんど無く、公共のデータベースに登録されているのは、グロビンや MHC など、進化系統を研究する対象として個々にシークエンシングされた遺伝子にかたよっているのが現状である。

本研究では、包括的にチンパンジーの遺伝子配列を解析する目的のもと、２頭のチンパンジー(*Pan troglodytes verus*)の脳、皮膚、肝臓の組織を試料として、mRNA のコピーである cDNA をクローン化し、ライブラリーとして大量に保存した。オリゴ・キャッピング法を用いることで、mRNA の 5'端（上流）から 3'端（下流）までの完全長を保存しているクローンに富む cDNA ライブラリーを作製することができた。そこからランダムに選んだク

ローンの 5'端配列部分のシークエンシングをおこなって、400bp 以上の長さの配列（EST; expressed sequence tag, 発現配列断片）を 7,064 配列蓄積した。アノテーション（生物学的意味）づけのため、これらの配列についてヒトの配列データベースに対して BLAST プログラムによる相同性検索をおこなった結果、3,875 配列（1,652 種類）のヒト mRNA と相同な遺伝子配列が含まれていることがわかった。残りの配列のうち 2,107 配列はヒト EST に、746 配列はヒトゲノム配列に一致したが、いずれにも一致しない配列が 85 配列あった。

ヒトと相同な 1,652 遺伝子のうち、1,537 遺伝子(93.0%)は本研究で用いた BLAST $E = e^{-120}$ の条件でヒトの配列と１対１で対応しており、オーソログの関係（２つの遺伝子が、種内での遺伝子重複ではなく、ある共通祖先からの種分化に由来する関係）である可能性が高いと考えられる。全 1,652 遺伝子のうち、脳由来が 928 遺伝子、皮膚由来が 937 遺伝子得られたが、両方から得られたのは 233 遺伝子のみであり、組織特異的な発現傾向を示していると考えられる。両組織で共通に多くクローンが得られたのは、EF1-alpha、tubulin、lactate dehydrogenase、ribosome 関連遺伝子などであった。脳からは neuron-specific protein、synaptosomal-associated protein、myelin 関連遺伝子、皮膚からは vimentin、decorin、keratin 関連遺伝子などが組織特異的に多く得られた。これまでに肝臓から得られたクローンからは、ヒトと相同な遺伝子が 35 種類得られ、albumin や fibrinogen など、脳と皮膚からは得られていない遺伝子が得られた。

遺伝子の塩基配列の種差を解析するには、まず正確度の高い配列データが必要である。クローニングやシークエンシングの過程でエラーが生ずることがあるからである。そこで、ひとつの遺伝子につき３つ以上のクローンの配列が得られている場合に、その複数のクローンの配列からコンセンサス配列（共通配列）を作製して配列比較をおこなった。コンセンサス配列は、複数の配列を ClustalW プログラムでアラインメント（最も共通部分が多くなるように並べる）し、各塩基座位について多数決（50%以上）で塩基を決定することによって作製した。全 1,652 遺伝子のうち 226 遺伝子のコンセンサス配列を得ることができた。

コンセンサス配列を対応するヒトの遺伝子配列と比較して、コンセンサス配列全体(5'-cDNA)、5'端の非翻訳領域(UTR)、コード領域(CDS)、コード領域を翻訳したアミノ酸配列について、一致度(identity)を計算した。CDS では、Li (1993)の方法を用いて同義・非同義置換（アミノ酸の変化をもたらさない・もたらす塩基置換、K_S, K_A）についても計算し、比較検討した。その結果、全体で 99.30%、5'-UTR で 98.79%、CDS で 99.42%、アミノ酸配列では 99.44%、同義置換(K_S)は 1.33%（一致度に変換(K_S*)すると 98.67%）、非同義置換(K_A)は 0.28%（K_S* = 99.72%）という値が得られた。

配列の一致度の比較解析のために対照として、カニクイザル(*Macaca fascicularis*)の 5'-

EST、37,587 配列（国立感染症研究所橋本研究室提供・ 80%が脳組織由来で他に精巣や皮膚などに由来）についても解析した。チンパンジーの場合と同様に、BLAST 相同性検索によるヒトと相同な遺伝子の同定およびコンセンサス配列の作製をおこない、ヒトとの配列の一致度を計算した。チンパンジーとカニクイザルで共通に得られた 133 遺伝子のコンセンサス配列について結果を得た。チンパンジー、カニクイザルの順に、全体(5'-cDNA)で99.43%, 97.43%、5'-UTR で98.76%, 93.68%、CDS で99.57%, 98.04%、アミノ酸部位では99.70%, 98.88%、同義置換(K_S)は 1.22%, 5.47%（K_S* = 98.78%, 94.53%）、非同義置換(K_A)は 0.14%, 0.53%（K_A* = 99.86%, 99.47%）という値が得られた（図）。

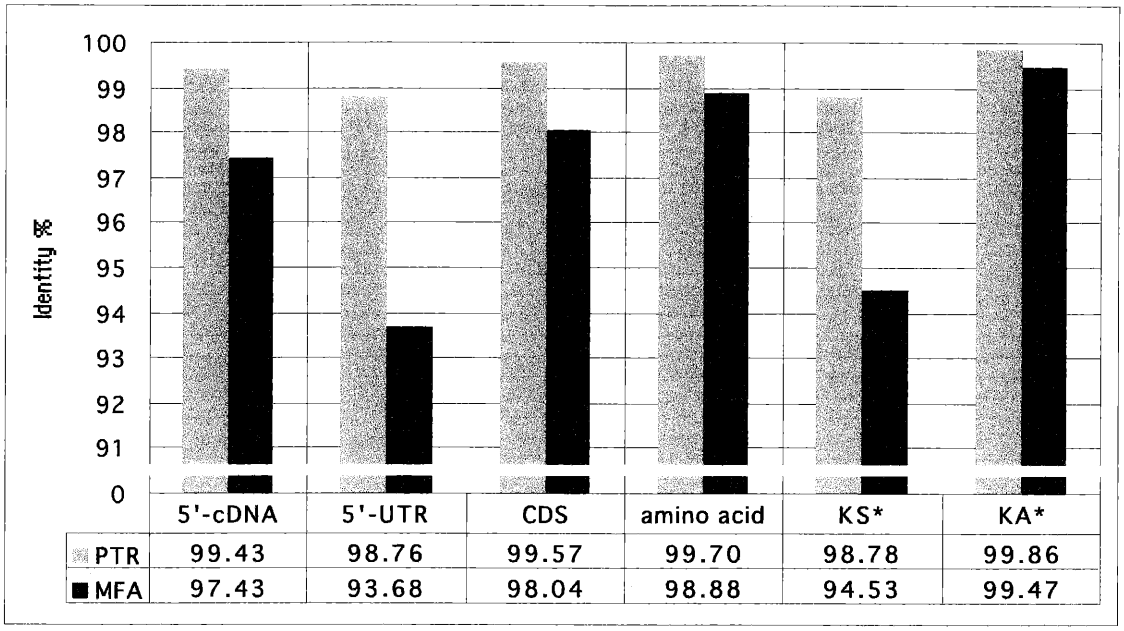
一般に、遺伝子(mRNA)配列では配列部位ごとに種間での保存度に違いがある。本研究でもヒトとの一致度は 5'-UTR では低く、CDS では高く、アミノ酸配列ではさらに高くなっている。アミノ酸の置換をおこす非同義置換は同義置換と比べて全体では約 1/5～1/10 である。同義置換は中立突然変異の蓄積を反映するため、種間系統距離を計ることができると考えられている。単純に見積もっても約 4.7 倍カニクイザルのほうがチンパンジーより変異度が大きい。これは本研究で初めて多数の核遺伝子の配列を比較した結果であり、ヒトとの分岐年代がチンパンジーでは約 550 万年、カニクイザルが約 2,500 万年とされている系統関係の値を支持するひとつの基準となるだろう。アミノ酸をコードしない 5'-UTR 配列の一致度は、本研究の計算結果では同義置換座位の一致度と差が無く、配列の保存性に関わるバイアスを確認できなかった。

本研究で計算した、種間での配列の一致度は、同じ遺伝子の同じ配列部位を比べて見つかった塩基置換数に基づいている。計算の過程で、塩基・アミノ酸単位の欠失や、数十塩基にわたって種間で大きく異なる配列があることが確認できた。これらはゲノム配列の違いだけでなく、オルターナティブ・スプライシング（ゲノム上の一遺伝子座位からエキソンの組み合わせにより複数パターンの mRNA が転写されること）などの転写後修飾を含むと考えられる。また、チンパンジーはヒトと比較して種内変異が大きいとされているが、本研究での同一遺伝子の複数クローン配列から、一塩基多型(SNP)なども確認できた。今後、種間で遺伝子配列のより詳細な比較解析をおこなうには、このような転写後修飾や種内変異を考慮して解析していく方法論を検討する必要があるだろう。

チンパンジーの大規模な遺伝子配列のデータおよびその解析はこれまでになく、ヒトに近縁な霊長類の遺伝子配列の解析は、ヒトの進化に関連する遺伝子や疾患遺伝子の探索・機能解析など、様々な側面から有効に活用することができると考えられる。cDNA ライブラリーおよびその配列データの蓄積はポストゲノミクスの研究にとっても、非常な貴重な試料となるだろう。

なお、本研究で強調すべき点のひとつに、コンセンサス配列の作製や一致度の計算をは

じめとする解析処理は、独自のプログラムを作成しておこなったことが挙げられる。本研究で得られた配列データおよび解析結果についてはデータベースを構築しており、インターネットを通じてアクセスすることができる（http://www.prigen.org/）。そこではチンパンジーのヒトと相同な遺伝子について、遺伝子名のキーワード検索や、BLAST による遺伝子配列の相同性検索をおこなうことができる。



	5'-cDNA	5'-UTR	CDS	amino acid	KS*	KA*
PTR	99.43	98.76	99.57	99.70	98.78	99.86
MFA	97.43	93.68	98.04	98.88	94.53	99.47

図　チンパンジー(PTR)とカニクイザル(MFA)の共通の 133 遺伝子配列のヒトとの一致度(identity)。UTR: untranslated region, CDS: coding sequence, KS* = 1 – KS, KA* = 1 – KA (%, KS: synonymous substitution, KA: nonsynonymous substitution)