

## ゲノムから読み解く生命システム——比較ゲノムからのアプローチ

監修 藤山秋佐夫  
(国立情報学研究所)

連載 第3回

## 脊椎動物の比較ゲノム: 遺伝子間領域の比較解析

隅山健太 斎藤成也  
Kenta Sumiyama, Naruya Saitou

最近の脊椎動物のゲノム解析の進行により複数生物種間での比較が容易にできるようになり、タンパク質をコードする領域以外の遺伝子間領域にも進化的に保存されている機能領域が存在することが明らかになってきた。ゲノム重複

などの影響を考え、適切な種を組み合わせることで、より多くの機能領域を発見することができる。さらに、複数生物種の比較を統合することで、ペアワイズ比較よりも詳細な機能領域のマッピングが可能になる。

## はじめに

脊椎動物のゲノムは、大半が“がらくたDNA (junk DNA)”と呼ばれる機能のない領域であり、遺伝子をコードする領域はわずかに過ぎないとされている。ヒトゲノムを例にとると、タンパク質をコードする遺伝子の数はたかだか25000個であると推定されているが<sup>1)</sup>、1遺伝子当たりのエクソンの総塩基数を3000個とすると、全遺伝子では75Mbとなる。これは3000Mbと推定される全ゲノムのわずか2.5%を占めるにすぎない。ヒトゲノムが特殊なのではなく、むしろ哺乳類ゲノムの典型だと考えられる。

では、ゲノムの残りの領域はすべてが機能を持たない、文字通りがらくたの領域なのだろうか？ ヒトゲノムの場合、膨大な直列反復配列領域や、散在反復配列のSINEとLINEが過半数を占めており、これらは特に積極的に機能を持たないがらくたであるようだが、それ以外のDNA領域には、単純な繰返しではないユニーク配列がたくさんある。この中には、遺伝子重複のなれの果てである偽遺伝子がたくさん含まれているが、それら以外にも、由来の不明な配列が多数存在する。それらの塩基配列の中には、その生物を成り立たせるのに重要な機能を有するものが数多く存在することが、近年少しずつわかってきた。これには、進化的に近縁な生物種間のゲノム配列の大規模比較が可能になり、高解像度で進化的な保存領域が同定できるようになってきたことが大きく寄与している。

本稿では、これからさらに急速に増大すると予想される、脊椎動物の近縁種ゲノムの比較解析の中でも、非コード領域の解析を中心とする脊椎動物のゲノム比較をどのように行うのか、また実際にどのようなことが見えるのかについて、筆者らの研究例を交えながら解説する。

## I. 比較ゲノムから明らかになる遺伝子間の“進化的保存領域”

タンパク質をコードする遺伝子領域からはmRNAが発現されるので、発現したmRNAの塩基配列の全長を大規模に決定すれば、ゲノム配列中のどの部分がエクソンとなり、どの部分がイントロンとして切り出されるのが明確にわかる。また、5'RACEや3'RACEにより、転写開始点、および終結点についても知見を得ることができる。さらに、このような既知の遺伝子構造のパターンを学習し、未知の配列中の遺伝子予測を行うことも盛んに行われている。

これに加え、ヒトゲノム配列やマウスなどのモデル生物ゲノム配列決定終了後に次々とモデル生物以外の近縁種ゲノム配列が決定されつつあり、その比較ゲノム手法により、“進化的に保存された領域”を機能部位の探索に用いることが可能になった。この進化的保存性を指標にした機能部位の探索法は、機能そのものが何であるかを知らずに機能部位を同定することができる、非常に強力な方法である。さらに、発現データやコンピュータによる予測などの独立した手法との組み合わせにより、信頼性の高い遺伝子発見ができるようになった。

機能の内容にかかわらず、機能の重要性に応じて進化的に保存されるというその特性から、タンパク質コード領域ばかりでなく、イントロンや転写されていない遺伝子間領域にも、機能的に重要な多数の進化的保存領域が見いだされることが近年わかってきた。それらの中には、超保存配列<sup>2)</sup>と呼ばれる、通常のタンパク質コード領域よりもはるかに保存度が高い領域も見いだされる。こうした遺伝子間の高保存領域は、おそらく遺伝子の発現調節を司る機能領域(エンハンサーや、オルタナティブプロモーターなど)や、



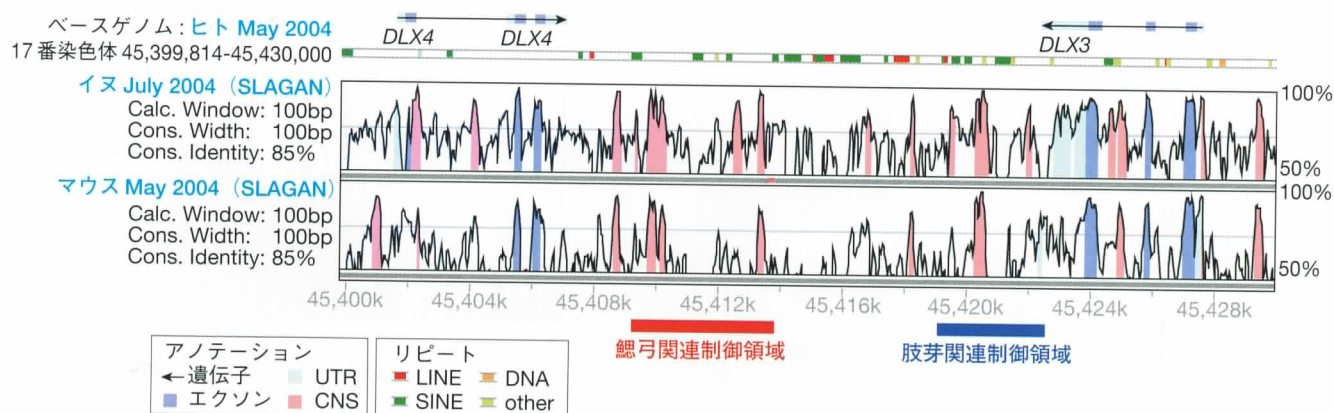


図1. VISTAを用いたヒト、イヌ、マウスのゲノム配列の比較の例

ウィンドウのサイズは100bp. マウスの発生において鰓弓と肢芽特異的エンハンサー活性を示す領域を図中に重ねて示した. UTR; untranslated region, CNS; conserved nucleotide sequence.

タンパク質をコードしないRNA遺伝子などであり、中立進化する周辺領域の配列に比較して進化的制約度が高いために、保存領域として観察できるのだと予想される. 特に高い保存度を示す領域は、初期の発生など進化的な制約が非常に強い生命現象に関わる遺伝子の発現制御などに関連しているのかもしれない<sup>3)</sup>.

## II. 遺伝子間領域の保存領域を解析するための方法

ここでは、長大なゲノム配列の比較解析について説明する. 種間でゲノム配列を比較するには、適切な整列(アライメント)を行うソフトウェアが必要である. 塩基配列の整列をどれだけ正確にできるかが、後の解析すべてに大きな影響を与えるため、最も重要なステップと言ってもよい. さらに、整列結果を図示する手段も必要である. 広く使われている定番のソフトウェアには、PipMaker<sup>4)</sup>とVISTA<sup>5)</sup>がある. 前者は局所的整列の集合体として全体の相似度を表現する方法であり、相同性(ホモロジー)の程度が直感的に理解しやすく、逆位や重複などがある場合にも使いやすい. 後者は固定した塩基長のウィンドウを配列全体にずらして当てはめていったときの各所のスコアを表示する方法であり、指定したウィンドウサイズに依存して結果が変化するという難しさはあるものの、原理的にシンプルで昔からよく使われている. どちらも有用であり、それぞれの長所がある. どちらを使うかは利用者の目的や好みによって選択すればよい. 図1にVISTAの解析結果を示しているが、まったく同じ領域をPipMakerで解析した図を発表している<sup>6)</sup>ので興味のある方はその違いを見比べて頂きたい.

## III. 比較に用いる生物種はどのように選ぶべきか

種間で保存されている機能部位を探索する目的でゲノム

配列を比較する場合、比較に用いる生物種の選択が非常に重要である. 第一に、ゲノム重複などによる遺伝子重複に注意しなければならない. 脊椎動物では、祖先の遺伝子数をNとすると、哺乳類に至るまでに2回のゲノム重複を経験し4Nになっていると考えられている. 硬骨魚類の系統ではさらにもう1回ゲノム重複が起こり、8Nとなっている(図2).

ゲノム重複が起こると一時的に進化的な制約が緩んで再編成が起こるためなのか、検出できる保存領域が減少することが知られている<sup>7)</sup>. ゲノム重複を1回多く経験した硬骨魚類のゼブラフィッシュと哺乳類間の比較よりも、系統独自のゲノム重複が起きていないサメと哺乳類のほうが、遺伝的により遠縁の関係にあるにもかかわらず、むしろ保存領域が多く見いだされるのである. このため、可能であれば、ゲノム重複を超えての比較を避けて、ゲノムの倍数性が同じ種同士を比較するほうが、より効率的に保存領域が解析できる確率が高くなる.

次に、比較に用いる生物種があまりに遠い関係であれば、ほとんどのゲノム領域にわたって整列が不可能であることがある. 例えば、ショウジョウバエとマウスのゲノム配列を比較しても、遺伝子間領域にはほとんど進化的な保存領域を見いだすことはできない. これに対して脊椎動物種間であれば、例えばフグとヒトのゲノム配列比較を行って遺伝子発現調節領域を探索することはずっと現実的になる. ただし、遺伝距離が離れていることに加え、先に述べたゲノム重複の影響もあって、進化的保存領域として検出されるのは、機能的に極端に重要でほとんど塩基配列変化が許されない領域だけである. フグ・ヒトの間でこのような保存領域が1400程度見つかっている<sup>8)</sup>. 比較する脊椎動物がさらに近縁種同士になっていけば、より多くの保存部位が見いだされるようになる.

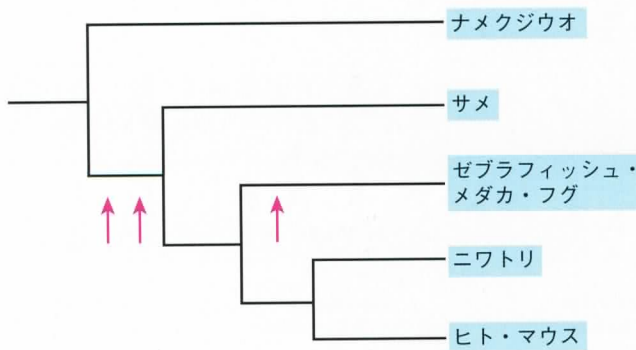


図2. 脊椎動物の進化過程におけるゲノム重複時期のモデル  
矢印は、ゲノム重複による遺伝子コピー数の増大を示す。

図1は哺乳類種間での比較で、ヒト、マウス、イヌの *DLX3/4* 遺伝子クラスターのゲノム配列を VISTA で解析したものである(隅山;未発表)。*DLX3/4* 遺伝子のエクソンが高いスコアを出しているのは当然として、そのほかにイントロンや、*DLX* 遺伝子3'側遺伝子間領域に非常に多くの保存領域を示すピークが存在することがわかる。実際この遺伝子間の保存領域は、発生過程の鰓弓や肢芽での発現制御をする機能領域であることが明らかになっている<sup>6), 9)</sup>。

このように、ヒトとマウスやイヌといった、哺乳類の中でも系統的に大きく異なる生物種間のゲノム比較は非常に有用であるが、同時に2種だけの比較には限界もある。種間の保存度を指標にするこのような方法では、比較する領域内に種が分岐してから現在に至るまでに十分な量の塩基置換などの突然変異が蓄積していることが前提である。あまりにも小さな領域の比較を行おうとすると、その範囲に十分な塩基置換が期待できず、保存しているのかどうかかわからない。また、もし比較する2種が最近に分岐した生物であれば、機能のない領域であっても種間の違いはもともと非常に少なく、たとえ広い領域をとっても保存領域をあぶり出すこと自体が困難になる。

例えば、マウスとラットとの間での保存領域を探す、ということは、ゲノム配列を用いることができるにもかかわらず、両者が進化的に近縁なので、おそらくかなり難しい作業になる。この問題を解決するための方法が、多種間でのゲノム配列比較である。

#### IV. 近縁多種間ゲノム比較による機能領域高精度マッピング

VISTA で描いた図1をもう一度見て頂きたい。ヒトとイヌ、ヒトとマウスのペアワイズ比較が示されているが、よく見ると両方の同じ位置のピークの高さが異なっている場合

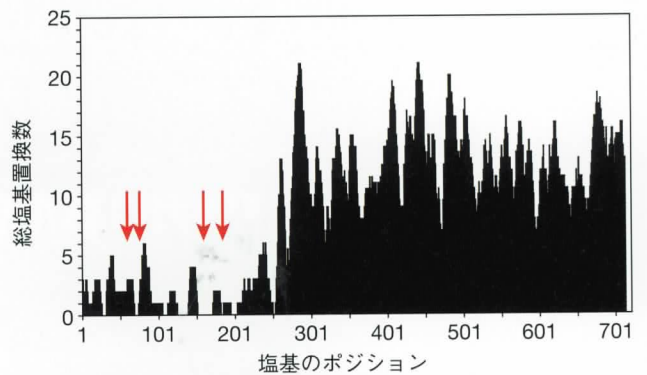


図3. *HOXC8* 遺伝子の早期エンハンサーにおける8哺乳類種を用いた進化保存モチーフの解析

矢印は実験で転写因子が結合することが示された場所を示す。ウィンドウサイズは10bp。

があることがわかりいただけると思う。これは種に特異的な変化である可能性もあるが、大部分は遺伝的な距離が十分でないため、偶然に置換数がばらつくことも影響している。このような2種だけの解析では、ウィンドウの幅が100bpであり、これよりも小さな保存領域を検出することは難しいが、かといってウィンドウのサイズを小さくすると、今度は偶然に種間で一致してしまうような擬陽性のピークが多数生じてしまうことになり、検出の信頼度が落ちてしまうことになる。

そこで、より多くの種間比較を行い、各比較の合計塩基置換数の期待値を上げることによって、信頼できるウィンドウサイズを小さくしていくことができる。近縁種間のペアワイズ比較では限られた情報しか得ることができないが、生物種を増やして近縁・多種間で比較を行いそのデータを統合すると、非常に高精度で緻密な進化的保存領域をマップすることができるようになるのである。

具体的な例として、図3に哺乳類 *HOXC8* 遺伝子の制御領域の近縁多種間比較の結果を示した。近縁多種間での比較により、転写因子が結合する配列のサイズに相当する10bp以下の保存配列を検出することができた<sup>10)</sup>。この検出結果は、生化学的な転写因子の結合実験の結果ともよく一致するものであった。

このような手法は、系統フットプリント法 (phylogenetic footprinting) とも呼ばれる。この呼び名の提唱者でもあるモリス・グッドマンらは、合計の進化時間が2億5千万年以上になるような種の組み合わせが望ましいと述べている<sup>11)</sup>。また、一応の目安として、統計的に有意に進化的保存があることを示すためには、統合された多種間比較ウィンドウ内に10以上の塩基置換が期待できるようなウィンドウサイズを設定するべきだろう。



## おわりに：今後の展望

ゲノムデータの蓄積により、その機能がまだ不明な物も含めて、タンパク質をコードする領域以外の機能部位を、近縁多種間比較により詳細に探索することが可能になってきた。これからはこのような機能未知のデータが大量に蓄積されていく時代になっていく。その機能を知るためのハイスループットの実験系が必要になるだろう。それによってグローバルな転写制御機構などの遺伝子発現プログラムの解明などが劇的に進むことが予想される。こうした知識の蓄積によってゲノム進化が生物進化に果たす役割が分子レベルで理解できるようになるだろう。

隅山健太 国立遺伝学研究所 集団遺伝研究部門

E-mail : ksumiyam@genes.nig.ac.jp

1991年東京大学理学部生物学科卒業、1996年同大学大学院修了、博士(理学)。Yale大学Frank Ruddle研究室で5年間ポストドク後、現所属、助手。発生進化遺伝学を研究。

斎藤成也 国立遺伝学研究所 集団遺伝研究部門

E-mail : nsaitou@genes.nig.ac.jp

1979年東京大学理学部卒業、1986年テキサス大学ヒューストン校Ph.D.。現在、国立遺伝学研究所集団遺伝研究部門および総合研究大学院大学教授。人類進化を中心とするゲノム進化を研究。

## 文献

- 1) International Human Genome Sequencing Consortium: Nature (2004) 408: 708-713
- 2) Bejerano G, et al: Science (2004) 304: 1321-1325
- 3) Sandelin A, et al: BMC Genomics (2004) 5: 99
- 4) Schwartz S, et al: Genome Res (2000) 10: 577-586  
<http://pipmaker.bx.psu.edu/pipmaker/>
- 5) Frazer KA, et al: Nucleic Acids Res (2004) 32 (Web Server issue): W273-279  
<http://genome.lbl.gov/vista/index.shtml>
- 6) Sumiyama K, et al: Proc Natl Acad Sci USA (2002) 99: 780-785
- 7) Chiu CH, et al: Genome Res (2004) 14: 11-17
- 8) Woolfe A, et al: PLoS Biol (2005) 3: e7
- 9) Sumiyama K, et al: Proc Natl Acad Sci USA (2003) 100: 4030-4034
- 10) Sumiyama K, et al: Genomics (2001) 71: 260-262
- 11) Gumucio DL, et al: Mol Phylogenet Evol (1996) 5: 18-32

## 用語解説

### 遺伝子間領域

ゲノム中のタンパク質をコードする遺伝子構造でない領域、またはエクソンでない領域。タンパク質をコードしないRNAや、遺伝子の発現調節に関わるエンハンサーやサイレンサー、さらに大規模なクロマチン構造の制御に関わる因子などがあると期待されるが、機能の解析が難しいため全容は明らかになっていない。

### がらくたDNA (junk DNA)

単純反復配列など特に機能的な意味を持たないと考えられている、現在のところ機能不明の配列。高等脊椎動物ではゲノム中でかなりの割合を占める。

### エンハンサー

遺伝子発現を増強させる作用を持つシスに働く因子。組織特異的な発現などを制御する。

### オルタナティブプロモーター

- 同一遺伝子のプロモーターが複数存在し、組織特異性などにより選択的に使われているもの。エンハンサーの多様性ととも、高等脊椎動物の遺伝子機能の多様性を生み出す機構の1つと考えられる。

### ゲノム重複

- 生物のゲノムが倍加することにより遺伝子重複を起こすこと。遺伝子の重複により各遺伝子にかかる機能的な制約による淘汰圧が緩むために遺伝子の進化速度が上がり、新しい進化を起こすきっかけになる可能性がある。