

学術俯瞰講義

「いのち」のシステムを解き明かす

—急展開する生命科学—

平成21年5月13日

「ゲノム、情報、進化」シリーズ第4回

ゲノムから生命システムの理解へ

情報・システム研究機構ライフサイエンス統合データベースセンター

東京大学大学院新領域創成科学研究科情報生命科学専攻

国立遺伝学研究所生命情報・DDBJ研究センター

高木利久

†: このマークが付してある著作物は、第三者が有する著作物ですので、同著作物の再使用、同著作物の二次的著作物の創作等については、著作権者より直接使用許諾を得る必要があります。

前回までの講義(3回分)のおさらい

- 高速なコンピュータ、アルゴリズムを使ったゲノムの配列(ATCGの並び)の決定、解析により
 - どのような遺伝子がコードされているか？
 - 生物に共通な遺伝子は？固有な遺伝子は？
 - それらの遺伝子は進化のプロセスでどのようにして獲得されてきたか？
 - これまで提唱されてきた進化の仮説の妥当性は？
 - 人類の多様性は？、どうやって進化してきたか？

など、さまざまなことが分かってきた

これで生命が分かったと言えるか？

- ゲノムは生命の設計図、レシピ
 - では、ゲノムが分かれば分かったのか？
 - 病気の予防、診断、治療、創薬はできるのか？
 - 美味しくて寒さや病害虫に強いお米は作れる？
-
- 答えはノー
 - なぜなら、生命の設計図、レシピを読み解くだけの知識を持ってないから
 - 遺伝子は複雑に絡み合って機能を発揮しているから、すなわちシステムだから

ゲノムは生命の設計図

DNA

GCTACTAGGCGCAGCGCATTGATCAGCCATGGAAA

遺伝子

↓ 転写

mRNA

ATG	GCT	TGG	GAC	AGT	CTA
-----	-----	-----	-----	-----	-----

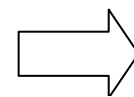
↓ 翻訳

アミノ酸配列



↓ 折り畳み

タンパク質



生物の形や働き
相互作用

本日の講義の内容

- 前回までは分かった、分かりそう
- 今回はこんなことも分からない、こんな研究必要
- 研究の最前線で研究者がもがいている様子
- ポストゲノム(シーケンス)の動向
 - システムの理解
 - オーム、オーミクス、システム生物学
 - データベース、統合化
 - オントロジー、テキストマイニング
- ゲノム研究、ゲノムを基盤とした研究の広がり
- 生命研究のパラダイムシフト
- 情報生命科学者の必要性(皆さんへの期待)

生命システムを理解すること

- システムとは？

- 日本語にうまく訳せない
- 便利な言葉
- 分かったようで分らない言葉
- 分野によっても人によって微妙に定義が違う

- ここでは、システムを、いろいろな部品が組み合わさってできていて、全体としては(部品の寄せ集めからは想像できないような)複雑な働きをするものと定義
- 生命はDNA(遺伝子)、RNA、タンパク質、化合物、などの部品からなるシステム

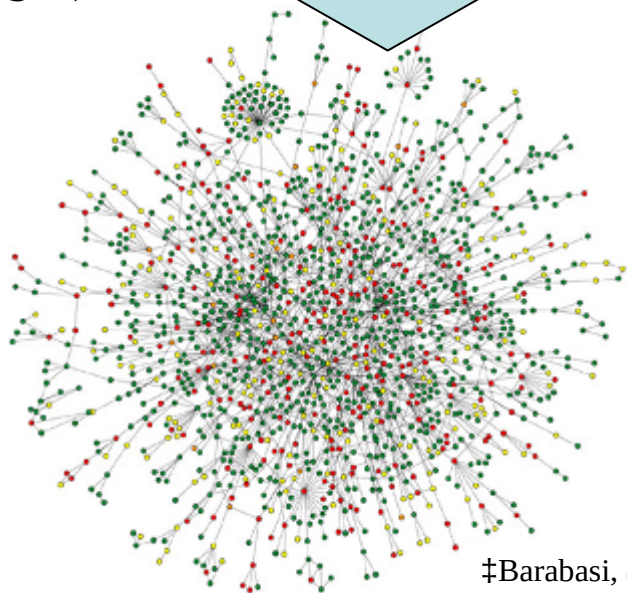
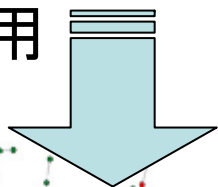
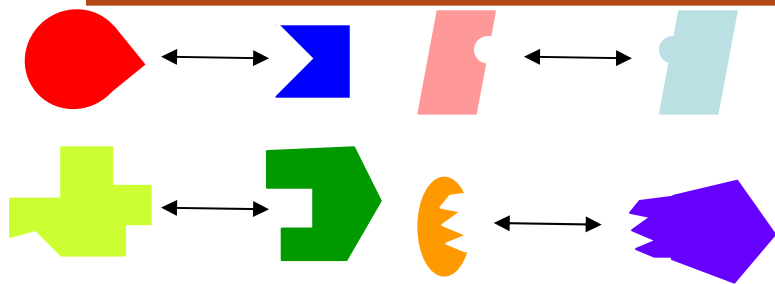
さまざまな理解のレベル

- 部品の列挙、カタログ化する
- 個々の部品の形が分かる
- 部品の使われる場所や時期が分かる
- 個々の部品の働きが分かる
- 部品のつながり、ネットワークを知る
- サブネットワークの働き、性質を知る
- システムの振る舞いが予測可能
- システムの振る舞いが制御可能
- 同じシステムを作ってみることができる

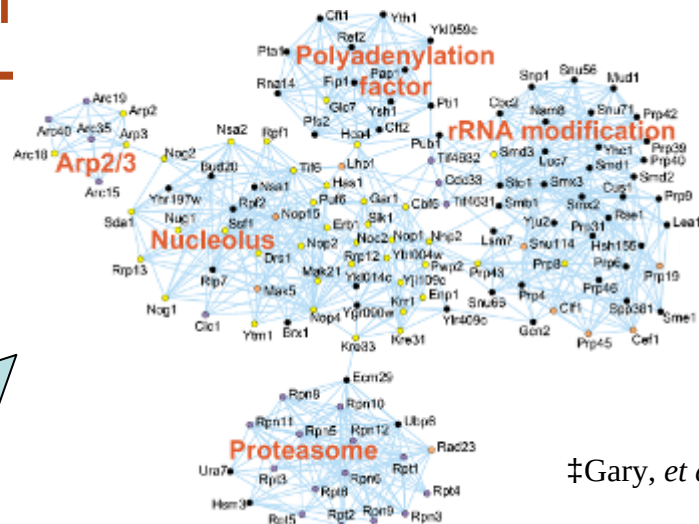
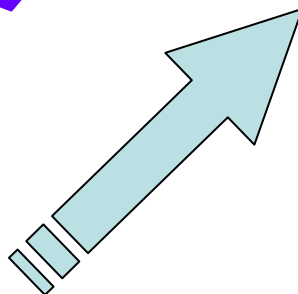
オーム、オーミクス、システム生物学

- 部品の同定、枚挙、カタログ化する ゲノム、ゲノミクス
- 個々の部品の形が分かる プロテオーム、プロテオミクス
- 部品の使われる場所が分かる トランスクリプトーム
- 個々の部品の働きが分かる 機能ゲノミクス
- 部品のつながり、ネットワークを知る インタラクトーム
- サブネットワークの働き、性質を知る ネットワーク生物学
- システムの振る舞いが予測可能 システム生物学
- システムの振る舞いが制御可能 システム生物学
- 同じシステムを作ってみることができる 合成生物学

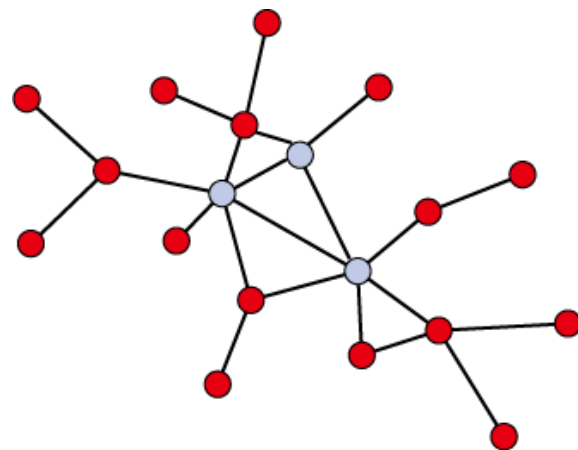
相互作用ネットワークの解析



相互作用のネットワーク



機能モジュールの同定



ネットワークの性質の解析

ネットワークに共通な性質解明



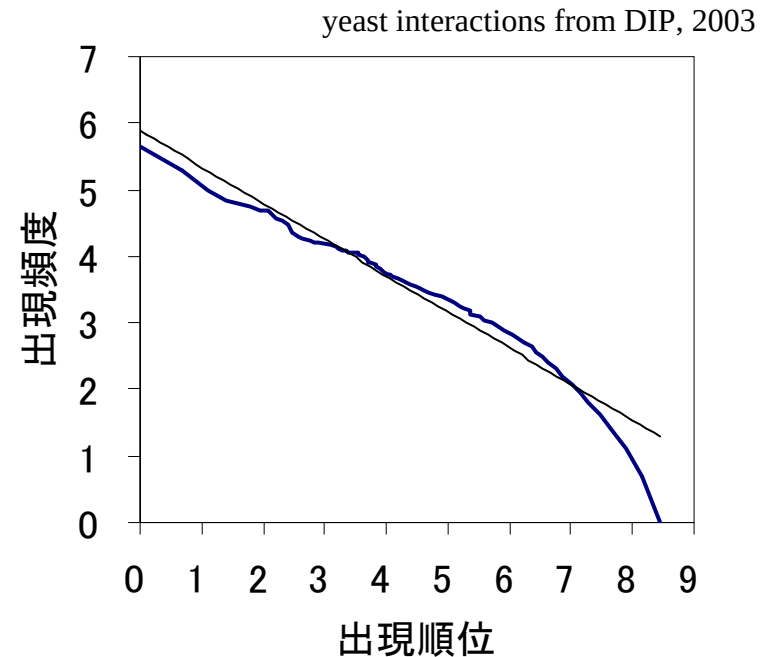
- 物理学者 バラバシ
- ネットワークに共通の性質
 - タンパク質のネットワーク
 - 代謝のネットワーク
 - コンピュータのネットワーク
 - 人間関係のネットワーク
 - スモールワールド
 - スケールフリー性

‡NHK出版 2002

ネットワーク生物学の例

- Zipfの法則
- 英単語の出現頻度
 - the 10%
 - of 5%
 - and 3%
 - to 2%

遺伝子名	相互作用数
JSN1	282
SRP1	197
NUP116	147
⋮	⋮
PBN1	1



分かりたいこと

- 構造と機能の関係
- 遺伝子型(genotype)と表現型(phenotype)の関係
- 1対1からN対Nの関係の理解へ
- Nの要素がどういう関係(システム)か？
- これらをそれぞれの
生命階層(細胞、器官、個体、集団)において
知りたい

生命研究におけるゲノムの意義

- ゲノムは部品の枚挙
- 有限性、網羅性→探索空間の絞り込み、削減
- ゲノム情報、ゲノム解析技術を基盤とした
さまざまな網羅的計測（例：DNAチップ）が可能
- ゲノムにより普遍性と多様性を一緒に、統一的
に研究可能
- ゲノムと他のオームは違うことに注意
 - ポスト・ポスト・ゲノム（シーケンス）はまたゲノム（シーケンス）

ゲノム解析の現状

Contact: Genomesonline	Last Update: May 11, 2009	Location www.genomesonline.org
992 Published Complete Genomes	Search GOLD : 4807 genome projects	167 Metagenomes
96 Archaeal Ongoing Genomes	2523 Bacterial Ongoing Genomes	1029 Eukaryotic Ongoing Genomes

GOLD: Genome Online Database v2.0より

ゲノムサイズと遺伝子数

	サイズ(bp)	遺伝子数
マイコプラズマ	60万	500個
大腸菌	460万	4,400個
出芽酵母	1200万	6,100個
分裂酵母	1300万	4,900個
粘菌	3400万	16,000個
線虫	1億	20,000個
ハエ	1.3億	14,000個
フグ	3.4億	28,000個
ニワトリ	11億	25,000個
マウス	26億	19,000個
ヒト	29億	22,000個

遺伝子数とシステムの複雑さの乖離

- ゲノム研究がもたらした大きな驚き、謎
 - これを理解するために
 - 遺伝子領域の読み替えによる多様なタンパク質産出
 - 遺伝子の転写調節の仕組み
 - 複雑な分子間相互作用
 - タンパク質非コードRNA(機能性RNA)
 - エピゲノム
- などに関する研究が展開されれている

システムの理解に向けたアプローチ

- 構造データの収集

- 部品やそのつながりの情報を網羅的に収集
- ハイスループットで高精度な測定装置の出現

- 機能データの収集

- RNA干渉などにより遺伝子機能を阻害して表現型見る
- 論文から過去の知見を集める

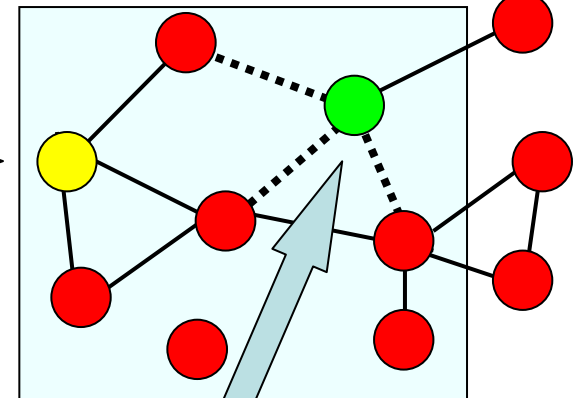
- 構造データと機能データの突き合わせ

- データベース上ですべての情報を整理、統合
- ネットワークを仲介にして

システムの理解に向けて

機能モジュールの同定

相互作用への
マッピング



コンピュータによる
隠れた関係の発見

翅の支脈形成
機能の理解

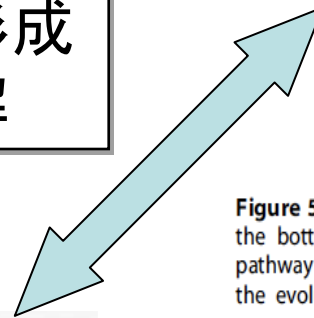


Figure 5. Modes of pathway acquisition. Each sequential line from the bottom left corner to the top right represents how rapidly a pathway was acquired. A segment corresponds to one branch during the evolutionary period when the pathway was acquired. Its slope represents the proportion of acquired genes divided by the branch length derived from the linearized phylogenetic tree (see Methods). Segments were sorted in descending order of their slopes to visualize how strongly the acquisition was biased. If the gradual acquisition scenario holds true, sequential lines will approach the 45° line, whereas the rapid acquisition scenario will produce lines that are strongly convex upward. To permit comparison, representative lines for the present and the randomized data are shown as gray dashed lines. doi:10.1371/journal.pgen.1000402.g005

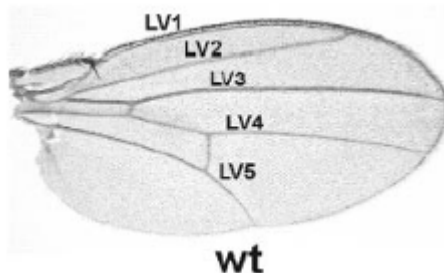
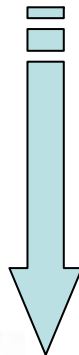
so be
uch as
genes
effects
refere,
ld hold

ce, we
resents
its the
thway
: rapid
pport-

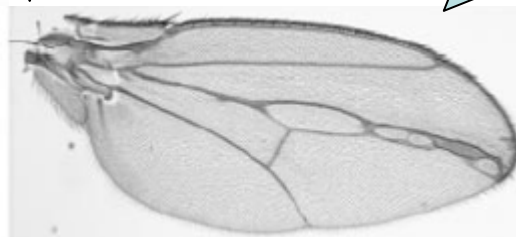
as described in the previous section (p. 5000). This rapid gene gain scenario is consistent with the highly heterogeneous modes of gene-content evolution; that is, genomes change drastically by sometimes expanding or shrinking quickly [13].



遺伝子破壊実験



wt

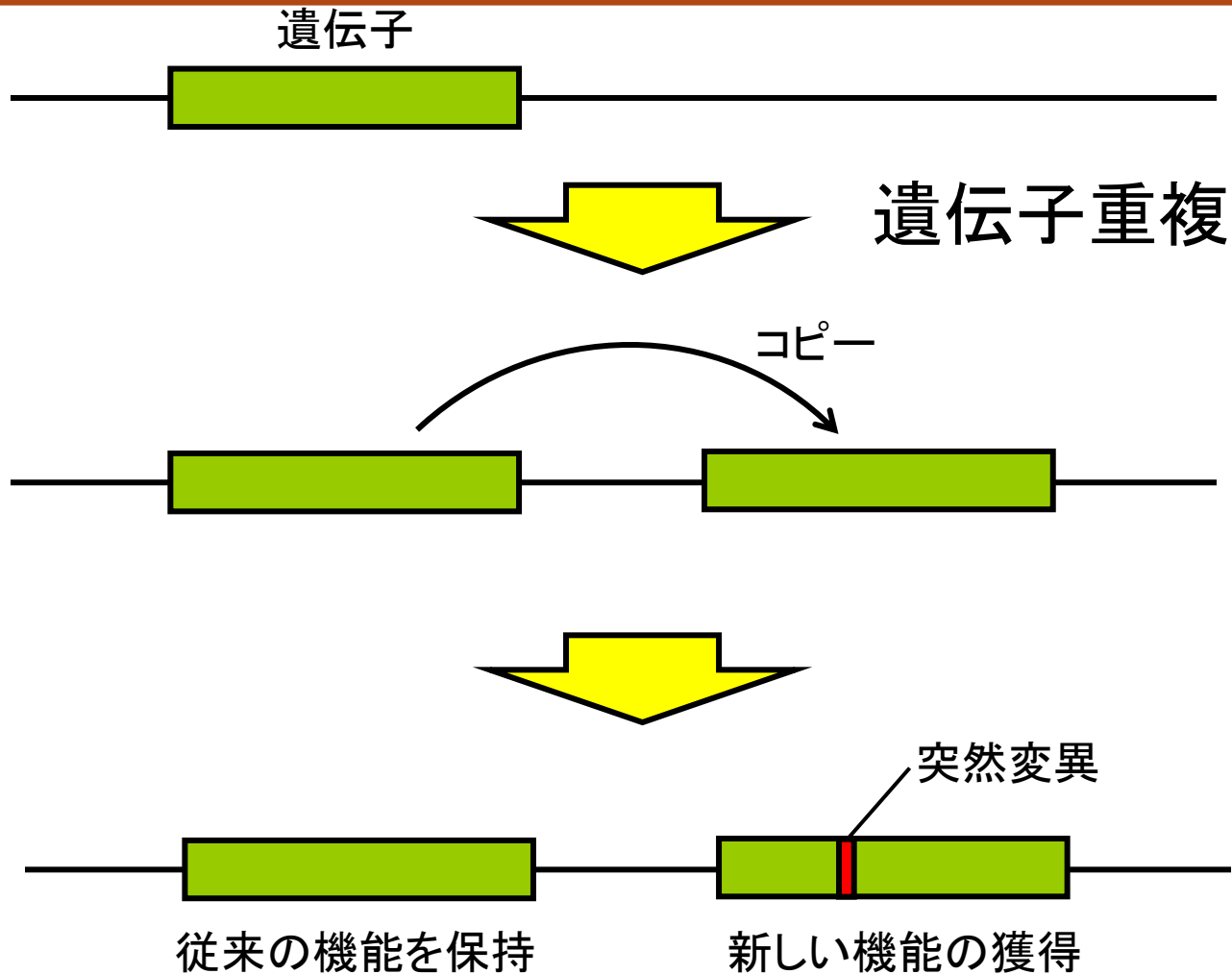


‡Ascano et al., 2004

生命システムの進化解析

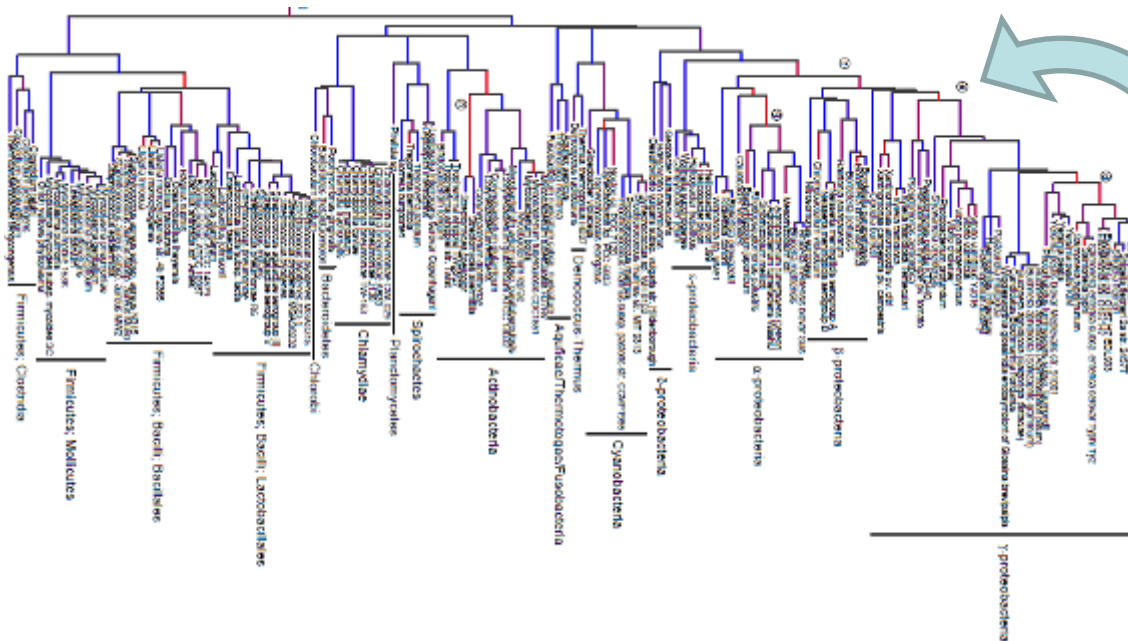
- “*Nothing in Biology Makes Sense Except in the Light of Evolution*” Dobzhansky(1973)
- ゲノムの進化解析からシステムの進化解析へ
 - 複雑な生命システムはどのように構築されてきたか
 - その背後にはどのような法則性が存在するか
 - 逆に、各生物種固有の表現型に対応する機能モジュール等を発見することはできるか

ゲノムは進化の産物:ゲノムの変化が進化の原動力



ゲノムの進化解析

- ゲノム規模かつ多くの生物種を用いた進化解析が可能に
 - 今年、解読済みゲノムは1000の大台に
- 今後もゲノム配列情報は爆発的に蓄積されていく
- さらに遺伝子発現制御ネットワーク等の情報も

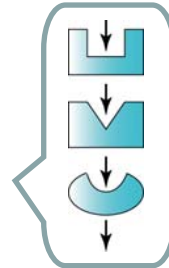


ゲノムの
進化過程
の再構築

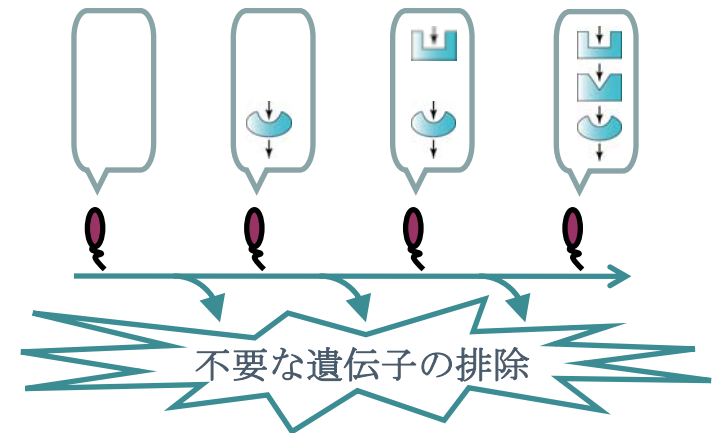
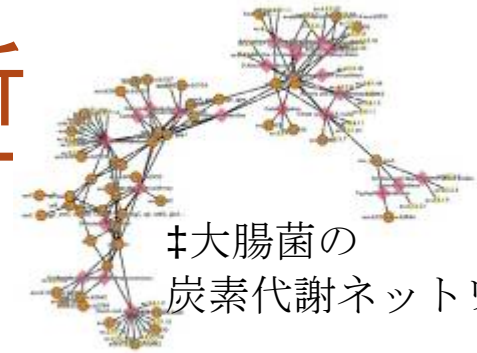
生命システム進化解析

- 多くの遺伝子／タンパク質が複雑に協調して働くネットワークは構成要素がそろって始めて役に立つ

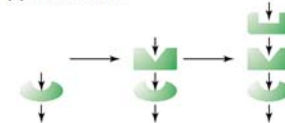
- 例：代謝ネットワークにおける連続する酵素反応



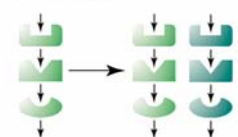
- その構築メカニズムは？
 - 一つ一つ遺伝子を獲得するのでは難しい！
 - 理論的なモデル構築，議論が先行
- ゲノムの進化過程と突き合わせすることで、そのメカニズムを実際に明らかに



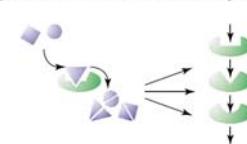
(b) Retro-evolution



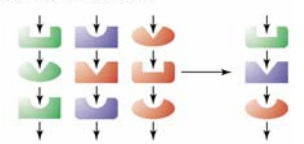
(d) Pathway duplication



(c) Specialization of a multifunctional enzyme

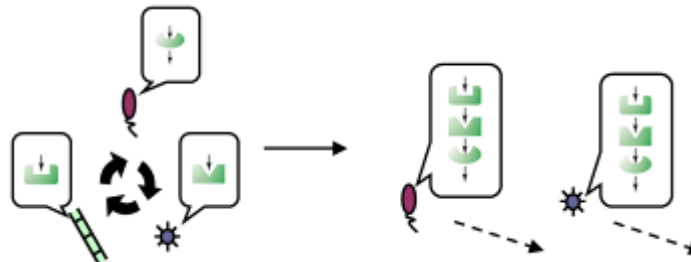
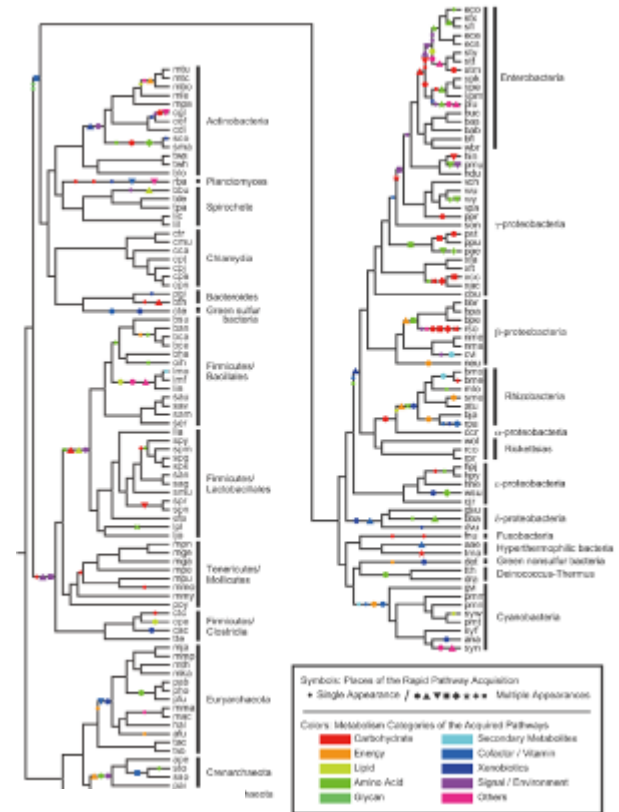


(e) Enzyme recruitment



ゲノムと生命システムの進化解析

- 160種の原核生物ゲノムを用いて過去の生物種のゲノムを推定
 - 代謝DBと組み合わせ、代謝ネットワーク進化過程をゲノム規模かつ生命の樹規模で再現
- 新規モジュールの”発明”が系統樹上で同じ時期、類似環境でなされる例
 - 同一環境の原核生物同士が遺伝子を提供し”協力”することで、ネットワークの進化が達成される可能性



データベースの必要性

- 多種多様で膨大なデータ、知識の整理必要
- 高速なコンピュータ、アルゴリズム
- 統合化されたデータベース
- データベース作りは作業ではない
- 高度に知的な活動
- これこそ生命研究そのもの？

- データベース生物学へ

複雑な知識をコンピュータでどう表現？

- 分子レベルでの実体の表現
 - ゲノム配列、遺伝子 → ATGCの並び
 - タンパク質 → アミノ酸配列、原子の座標配置
- 分子間の関係や挙動の表現
 - 分子間相互作用 → 2項関係
 - 遺伝子発現情報 → 発現強度(数値リスト)
 - 上記データ取得の文脈 → オントロジー
- 機能、表現型の表現
 - パスウェイ、ネットワーク → グラフ
 - 表現型 → 画像、動画、言葉
 - 概念、機能とその階層 → オントロジー
 - 動的な挙動 → 数理モデル、シミュレータ

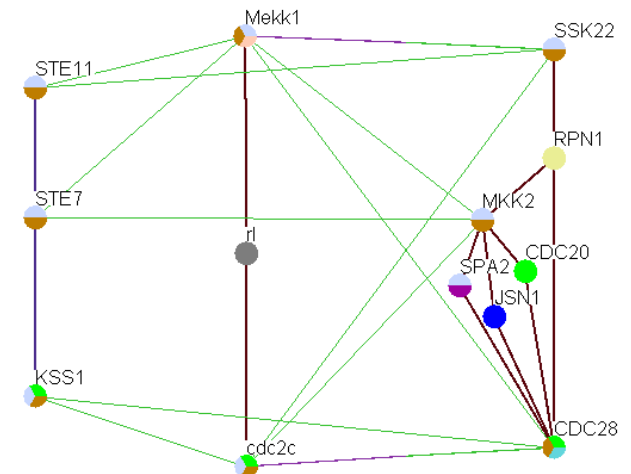
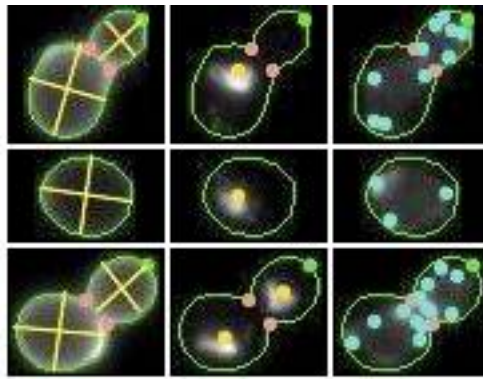
生命研究ではさまざまな比較が重要

ATGCCTTGG-TATAATCCTATGA--GTATCG...

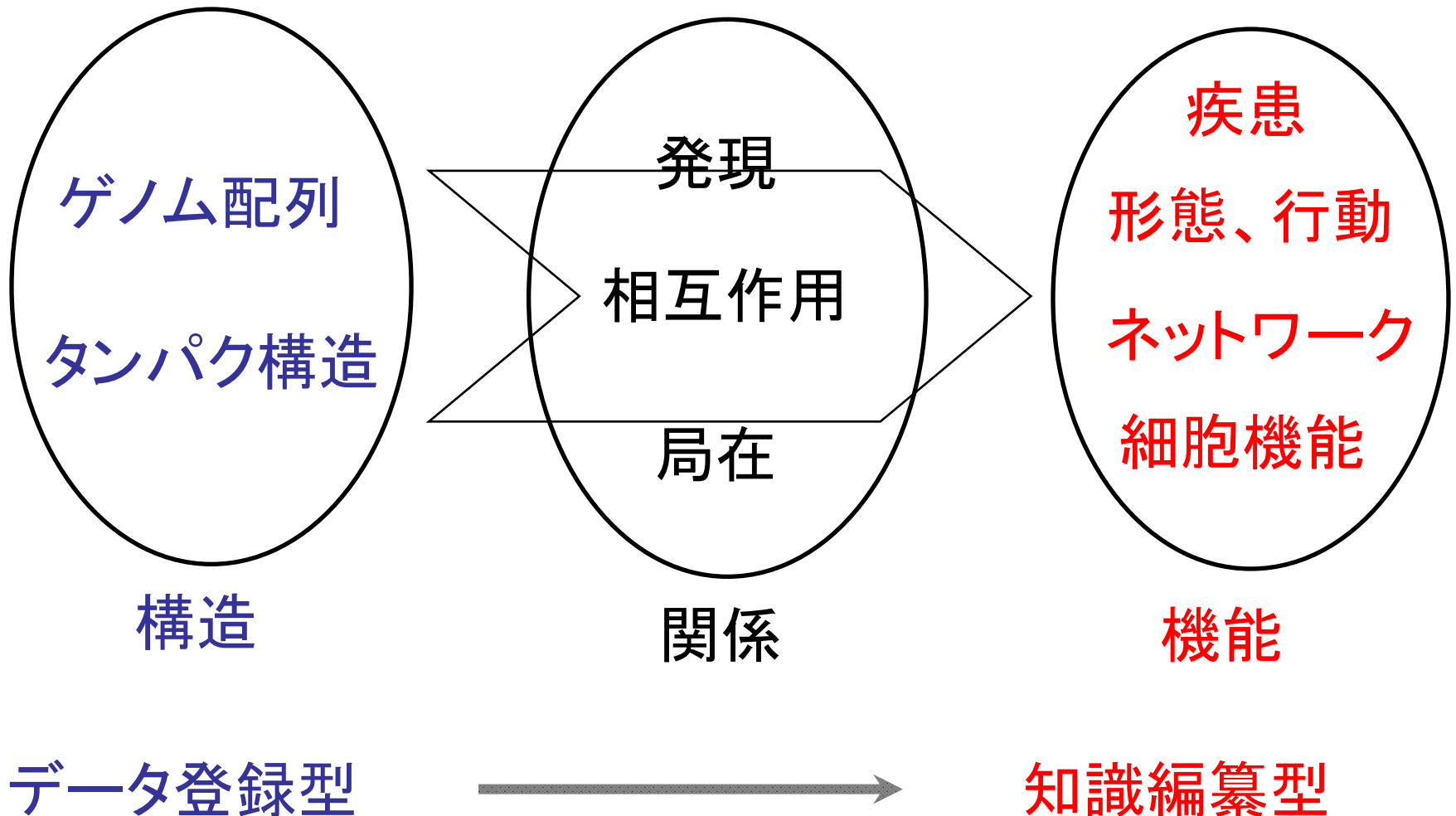
: : : : : : : : :

AACCCT-GGGTA-AAT-TTTGAAAGTTTCGG...

	肝臓	心臓	脳	肺	腎臓	胃	腸
遺伝子A	300	30	0	5	20	10	300
遺伝子B	0	0	10	10	200	20	30
遺伝子C	0	0	500	0	0	0	0
遺伝子D	400	20	0	10	20	10	300
遺伝子E	0	300	0	20	0	0	0



データベース構築の傾向



オントロジーとは

- 存在論を意味する哲学用語
- コンピュータ科学で知識処理研究として
- 生命科学では
 - 遺伝子名・タンパク質名の辞書(別名、略語も)
 - 機能を表す用語の統一(生物種横断的)
 - さまざまな概念や用語の階層関係
 - 概念間の関係の記述法
 - 統一的なデータベースフォーマット
- 機能DB、テキストマイニング、DB統合に不可欠
- ゲノムは生物種横断的、共通性と多様性の研究可能

ゲノム研究におけるオントロジー

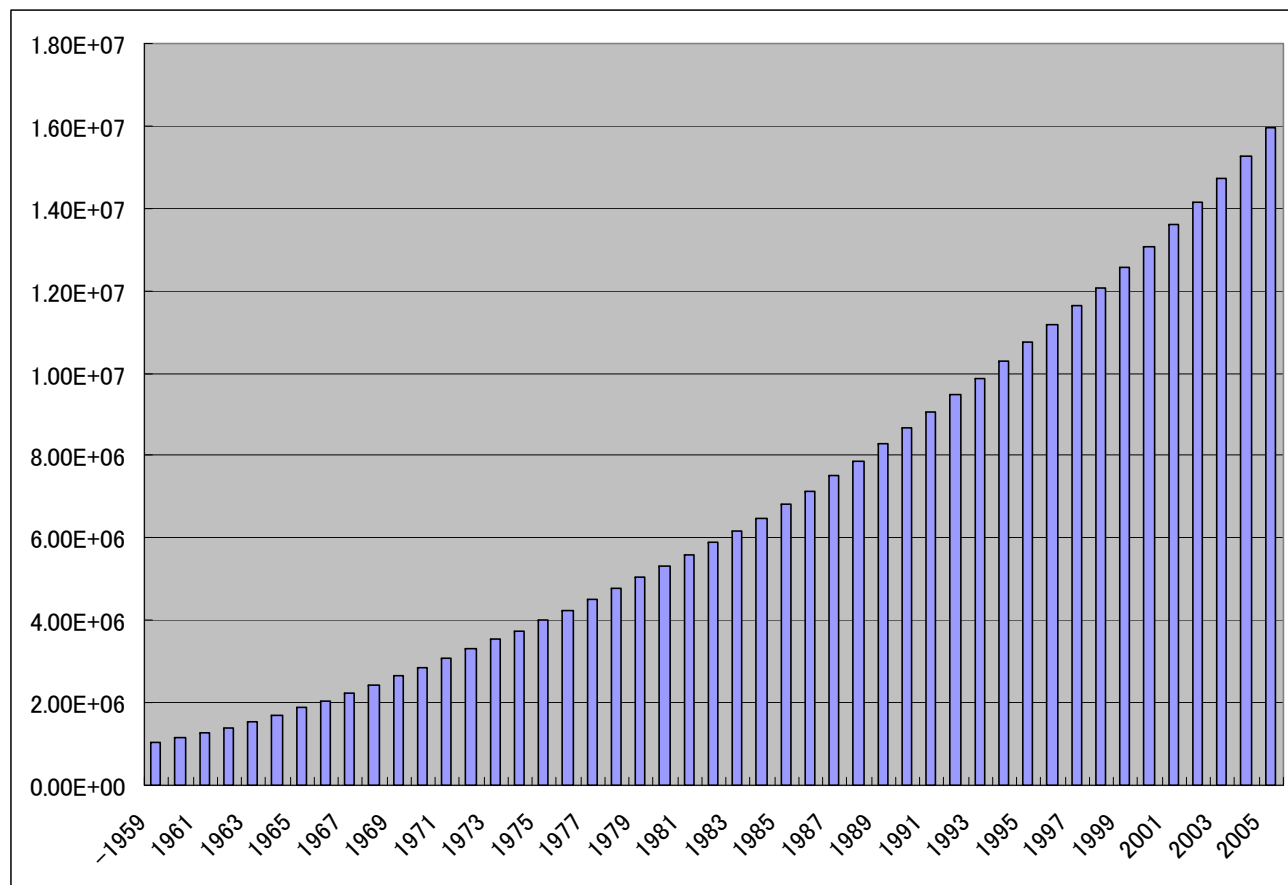
- 生物学では、生物種や生命現象ごとに研究コミュニティが形成されてきた
- 同じ分子、概念、機能でもコミュニティによって用語が異なること(その逆も)がしばしば
例: 遺伝子の定義がDBで違う
- 生物種、生命現象横断的なゲノム研究では比較や統合が必要となり、それらの用語の対応関係を整理し、統一的な語彙を使うようになってきた(2000年頃より)

遺伝子名のバリエーションの例

ID		
● GHS012062	Symbol	MAPK1
	Full name	mitogen-activated protein kinase 1
	Synonym(s)	ERK
		ERK2
		ERT1
		EXTRACELLULAR SIGNAL-REGULATED KINASE 2
		EXTRACELLULAR SIGNAL-REGULATED KINASE II <i>[manual]</i>
		MAPK2
		p38
		p40
		p41
		p41mapk
		p42MAPK
		p42MAPK1 <i>[PUBMED:2813464]</i>
		PRKM1
		PRKM2
		protein kinase, mitogen-activated 1 (MAP kinase 1; p40, p41)
		PROTEIN KINASE, MITOGEN-ACTIVATED, 1
		PROTEIN KINASE, MITOGEN-ACTIVATED, 2
		PROTEIN KINASE, MITOGEN-ACTIVATED, I <i>[manual]</i>
		PROTEIN KINASE, MITOGEN-ACTIVATED, II <i>[manual]</i>
		PROTEIN TYROSINE KINASE ERK2

論文数の増大

文献DB MEDLINEに収録されている文献(アブストラクト)数



1960年
105万件
↓
2008年
1,800万件

オープンアクセス化の動きも

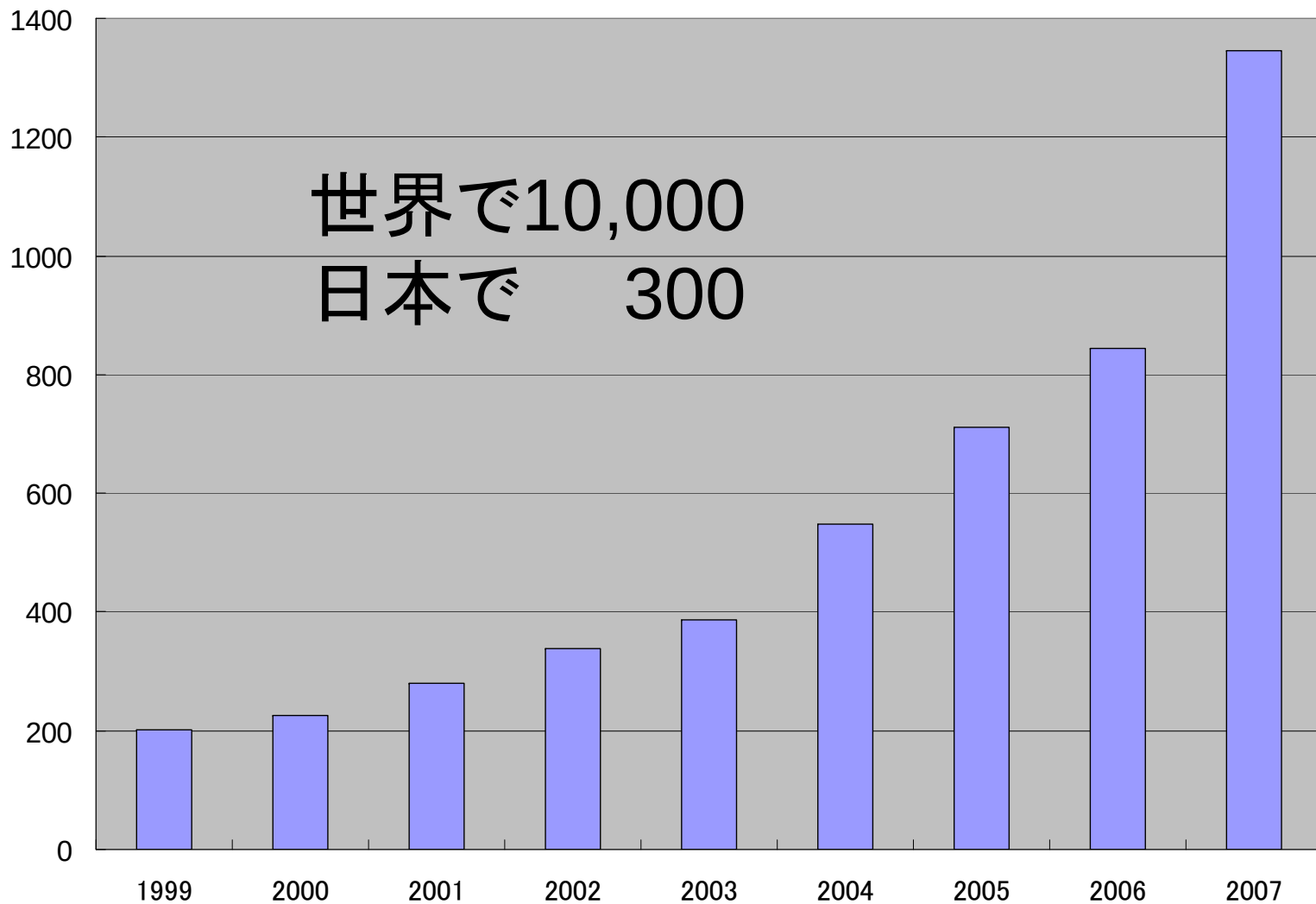
テキストマイニング

Many biological functions, from energy metabolism to antibiotic resistance, are carried out by biological pathways that require a number of cooperatively functioning genes. Hence, underlying mechanisms in the evolution of biological pathways are of particular interest. However, compared to the evolution of individual genes, which has been well studied, the evolution of biological pathways is far less understood. In this study, we used the abundant genome sequences available today and a novel algorithm we recently developed to trace the evolutionary history of prokaryotic metabolic pathways and to analyze how these pathways emerged. We found that the pathways have experienced significantly rapid acquisition, which would play a key role in eliminating the difficulty in holding genes during the course of pathway evolution. In addition, the emergence of novel pathways was suggested to have occurred more contemporaneously than expected across different phylogenetic clades. Based on these observations, we propose that novel pathway evolution can be facilitated by bidirectional horizontal gene transfers in prokaryotic communities. This simple model may approach the question of how biological pathways requiring a number of cooperatively functioning genes can be obtained and are the core event within the evolution of biological pathways in prokaryotes.

自動抽出

分子間相互作用
疾患と遺伝子の関係
遺伝子発現制御
細胞内局在
生体内パスウェイ
タンパク質の機能
機能や概念の階層
実験条件

データベースの個数(NAR誌より)

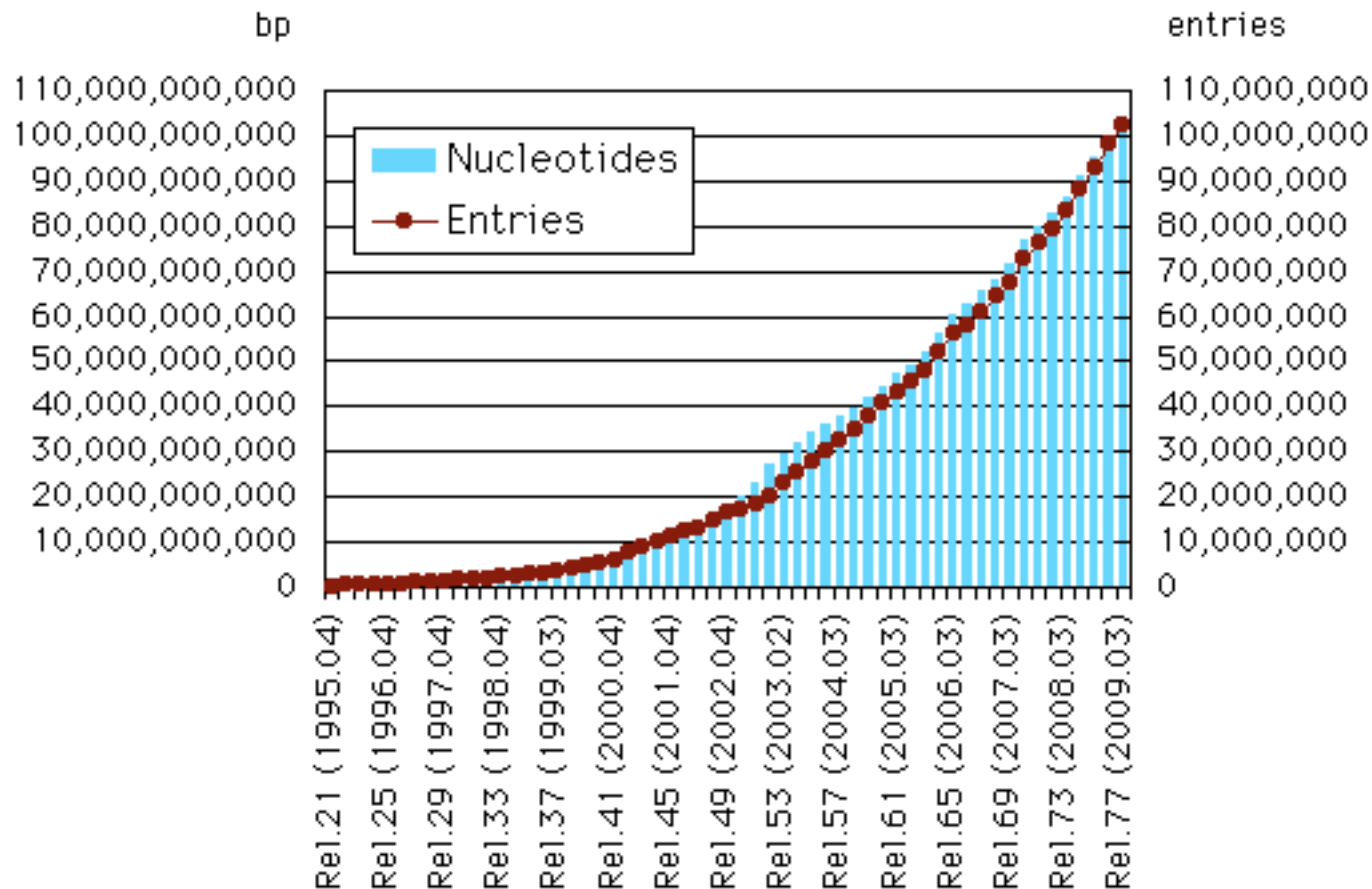


研究のパラダイムシフトと データ・知識の共有、統合化の重要性

- ゲノム研究の進展
- ハイスループットで高精度の測定装置の出現
- 仮説駆動型からデータ駆動型へ
- 法人化、バイドール法→データの囲い込み
- データ爆発、知識爆発、データベース爆発
- 分野の細分化、データ・知識の断片化・分断
- システムの理解には共有・統合化が必要

DDBJにおける塩基数の伸び

DDBJ/EMBL/GenBank database growth



‡ * Note : CON and TPA divisions are not counted in the Release statistic.

ペタバイト時代の科学

著作権の都合上、
ここに挿入されていた図は
削除いたしました。

nature
2008年9月4日号



ポータル

[生命科学系 データベース カタログ](#)  

[生命科学系 学協会カタログ](#)  

[ゲノム・ポストゲノム主要プロジェクト一覧](#) 

[生物アイコン](#)  

[WingPro \(JSTのDBポータル\)](#)  

[Webリソースポータルサイト](#) (JST解析ツールポータル)



検索

[\(旧\)生命科学データベース横断検索](#)  

[蛋白質核酸酵素 全文検索](#) 

[文科省「ゲノム」研究報告書 全文検索](#) 

[学会要旨統合検索](#) 

[新聞記事検索](#) 

[OReFiL \(オンラインリソースファインダー\)](#)  

[Allie \(略語の正式名称を検索\)](#)  



データベース

[DNAデータベース総覧と検索](#)  
(DDBJ/EMBL/GenBank)

[遺伝子発現バンク\(GEO\)目次](#)  

[かずさアノテーション & Navigation](#) (かずさDNA研究所) 

[ゲノムネット医薬品データベース](#) (京大)

[統合医科学データベース](#) (東京医科歯科大グループ)

[疾患解析から医療応用を実現するDB開発](#) (東大グループ)



アーカイブ

[生命科学系データベースアーカイブサービス](#) 
[トレースアーカイブ](#) (遺伝研 DDBJ)



ツール & 解析サービス

[アナトモグラフィー/BodyParts3D](#)  

[Wired-Marker](#)  



基盤技術開発

[共通基盤技術開発の概要](#)

[TogoDB](#) (誰でもデータベースが構築できる) 

[TogoWS](#) (ウェブサービスの標準化)  

[OpenID 認証システム](#) 

[統合DB情報基盤サイト](#) (CBRC) 

<http://lifesciencedb.jp/>



教材・人材育成

[統合TV](#) (DBやツールの動画教材)  

[MotDB](#) (教育・人材育成のサイト)



統合DB事業

[文科省 統合データベース整備事業サイト](#)

[国内データベースの統合\(受入れ\)事業](#)

[H18年度成果公開サイト](#)

まとめ

- ゲノム研究の広がり、応用面も
- システムの理解へのさまざまなアプローチ
- コンピュータ、データベースが重要
- さまざまな分野との融合
- がん、脳、ナノテク、物理学、複雑系科学、化学、生命倫理、、、
- 研究のパラダイムシフト
- データを出すより、解析することが重要
- 新しい理論、方法論が必要