



平成22年度 第1回データベース講習会
「創薬研究における統合データベースの活用」

比較ゲノムデータベースEvola, G-compassの活用法

<http://hinv.jp/evola/>

<http://hinv.jp/g-compass/>

産業技術総合研究所 バイオメディシナル情報研究センター

原 雄一郎

2010年9月14日(水) 於: 産業技術総合研究所臨海副都心センター

本日の概要

H-InvDBが提供する2つの比較ゲノムデータベースの紹介およびデモ



Evola

分子進化データベース



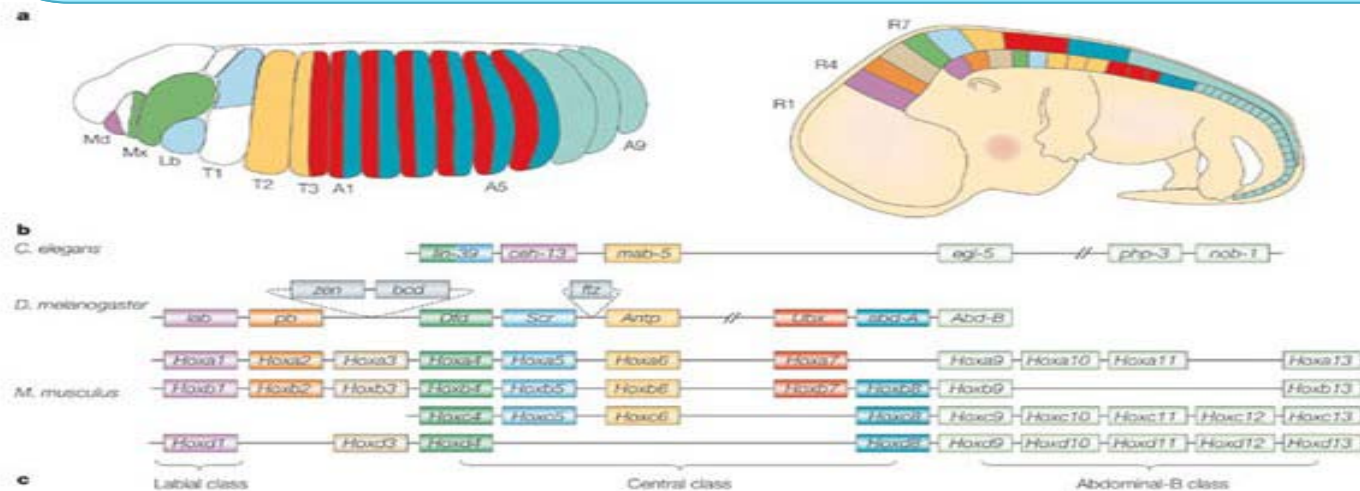
ゲノムアラインメントブラウザ

比較ゲノム・・・異種間あるいは同種間におけるゲノム配列や構造の網羅的な比較解析を比較ゲノムという。

ゲノムの注釈づけ(アノテーション)や、進化学的観点に基づく生命現象の解明への基礎データとして用いられる。

進化学から見たモデル生物

- モデル生物を用いた実験がヒト遺伝子の機能解析の代替となり得るわけ
 - ヒトとモデル生物は共通祖先から進化した
 - ヒトとモデル生物は相同なゲノム・遺伝子セットを持つ
 - ヒトとモデル生物で共通する形質は共通祖先から進化的に保存されてきた可能性が高い
- ヒト遺伝子とモデル生物の相同遺伝子は、相同な機能を持つと考えられる

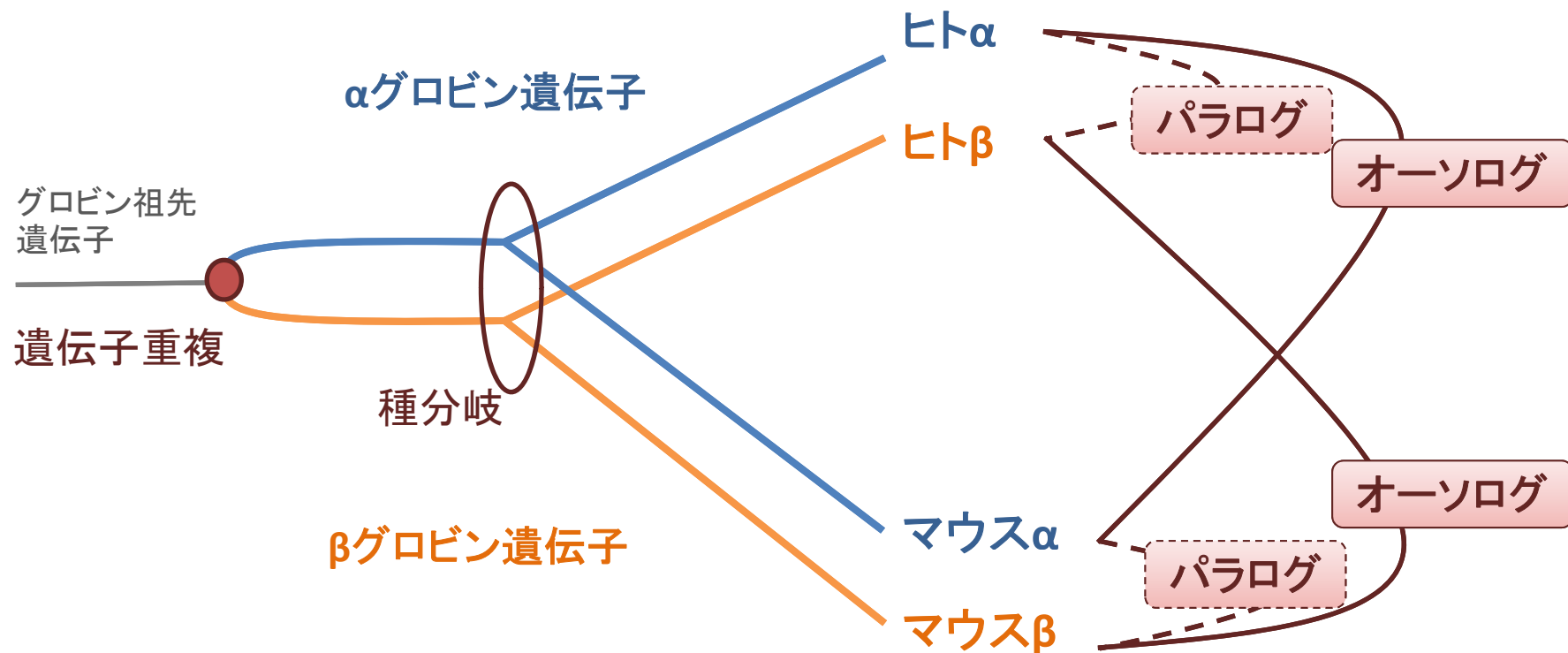


Pearson *et al.*,
Nat.Revi.Genet., 2005

相同遺伝子: オーソログとパラログ

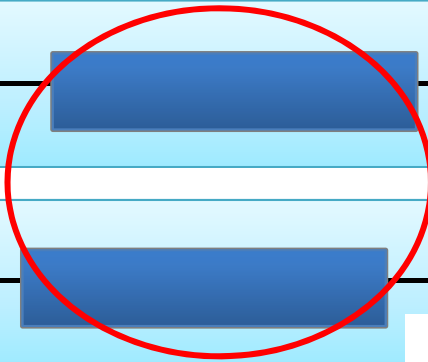
オーソログ: 種分岐によって分かれた相同遺伝子

パラログ: 遺伝子重複によって分かれた相同遺伝子



Evola, G-compassが対象とするデータ

ヒト



マウス



Evola

オーソログ群を中心とした遺伝子ファミリー

H-InvDB Annotated Human Gene Database

Evola

Home Quick guide Nav

Human (Homo sapiens)
HIT000000011 (AB002003) ZFYVE16 1 1539 aa
Zinc finger FYVE domain-containing protein 16 (Endofin) (Endosome-associated FYVE domain protein).
Category 1: Identical to known human protein Q7Z3T8
Alignment Locus maps

Gene family/group
Chimp View Macaque View Mouse View Rat View

Ortholog 2
Similarly Search Tool

Species	Accession	Similarity	Length
Chimpanzee (Pan)	XM_001136140	98.38%	1539 aa*
Macaque (Macaca)	ENSMMUT00000032568	94.95%	1491 aa*
Mouse (Mus)	BC042669	73.09%	1546 aa*
Rat (Rattus)	XM_001058708	72.00%	1557 aa*
Dog (Canis)	ENSCAFT00000014049	82.71%	1550 aa*
Horse (Equus)	XM_001504645	85.03%	1550 aa*
Cow (Bos)	XM_866874	83.67%	1549 aa*
Zilbarfish (Danio)	XM_001920883	40.42%	1561 aa*

* Calculated by FASTA alignment, overriding internal stop codon or frameshift.

Alignment 3
Amino acid
HIT000000011
XM_001136140
ENSMMUT00000032568
BC042669
XM_001058708
ENSCAFT00000014049
XM_001504645
XM_866874
XM_001920883
HIT000000011
XM_001136140
ENSMMUT00000032568
BC042669
XM_001058708
ENSCAFT00000014049
XM_001504645
XM_866874
XM_001920883

Ortholog 2
Similarly Search Tool

Data download
Sequence (fasta) Transcript, Protein
Gene family tree (phb**) Neighbor joining (NJ) [View by ATV] (requires Java)
* Available for alignable sequences.
** View the phb file using Nplot or TreeView.

Gene ontology (GO)
Molecular function zinc ion binding GO:0008270

InterPro

Accession number	Short name	Sequence region (aa)
IPR000306	Zinc finger, FYVE-type	739-806, 742-806, 747-805

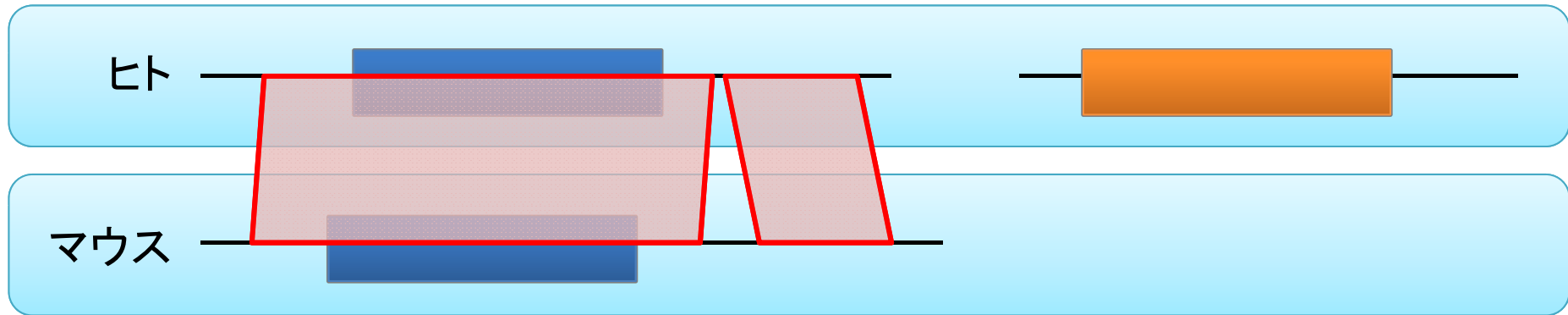
ATV | A Tree Viewer
File Search Tools View as Text Display Options Help

- phyloaram
- dynamic hiding
- show internal data
- show node names
- show taxonomy
- show sequence acc
- show annotation
- show binary characters
- show binary character counts
- show domain architectures
- branch length values
- show support values
- show event
- taxonomy colorization
- color branches

click on node to:
display node data
Show Properties
zoom in X zoom out X
zoom in Y zoom out Y
show whole
return to super tree
order subtrees
uncollapse all

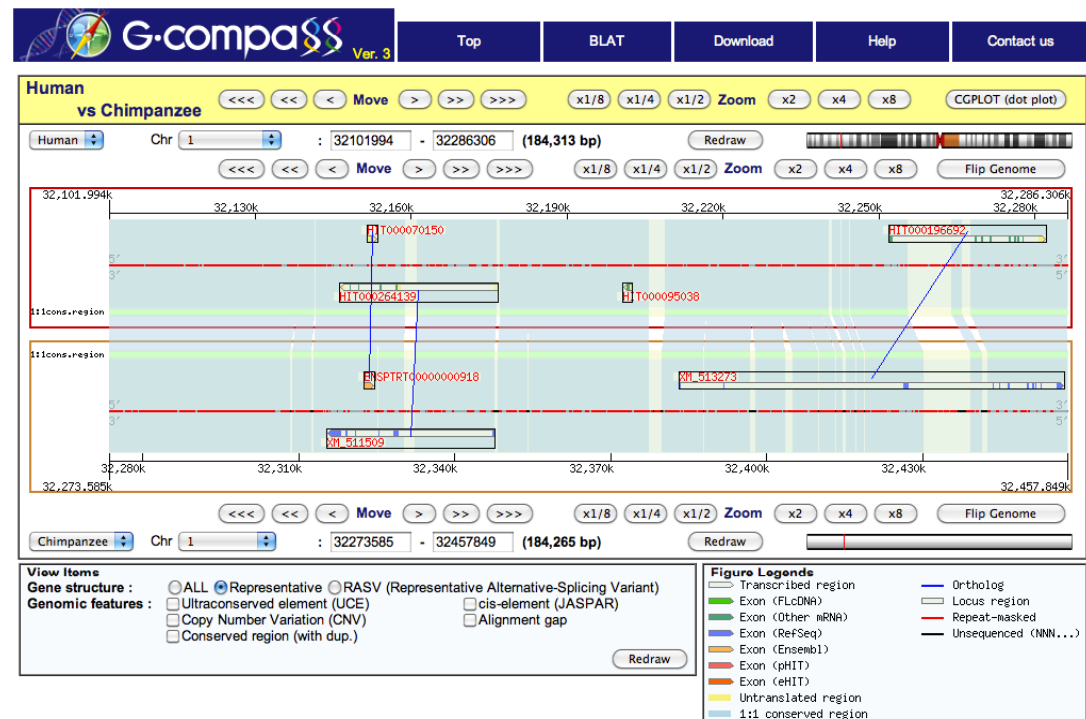
```
graph TD
    Node1[1000 XM_001058708_Rattus] --- Node2[1000 BC042669_Mus]
    Node2 --- Node3[489 XM_001504645_Equus]
    Node3 --- Node4[1000 XM_866874_Bos]
    Node4 --- Node5[1000 ENSCAFT00000014049_Canis]
    Node5 --- Node6[1000 HIT000000011_Homo]
    Node6 --- Node7[1000 XM_001136140_Pan]
    Node7 --- Node8[1000 ENSMMUT00000032568_Macaca]
    Node8 --- Node9[1000 XM_001920883_Danio]
    Node9 --- Node10[1000 XM_001544468_Danio]
    Node10 --- Node11[1000 ENSGAL700000017268_Gallus]
    Node11 --- Node12[1000 ENSMAGT000000106655_Mus]
    Node12 --- Node13[1000 XM_001060706_Rattus]
    Node13 --- Node14[1000 HIT000000011_Homo]
    Node14 --- Node15[673 1600-XR_013958_Macaca]
    Node15 --- Node16[1000 ENSPFR00000045158_Pan]
    Node16 --- Node17[866 ENSGCA700000007948_Equus]
    Node17 --- Node18[995 XM_865952_Bos]
    Node18 --- Node19[721 XM_854135_Canis]
```

Evola, G-compassが対象とするデータ



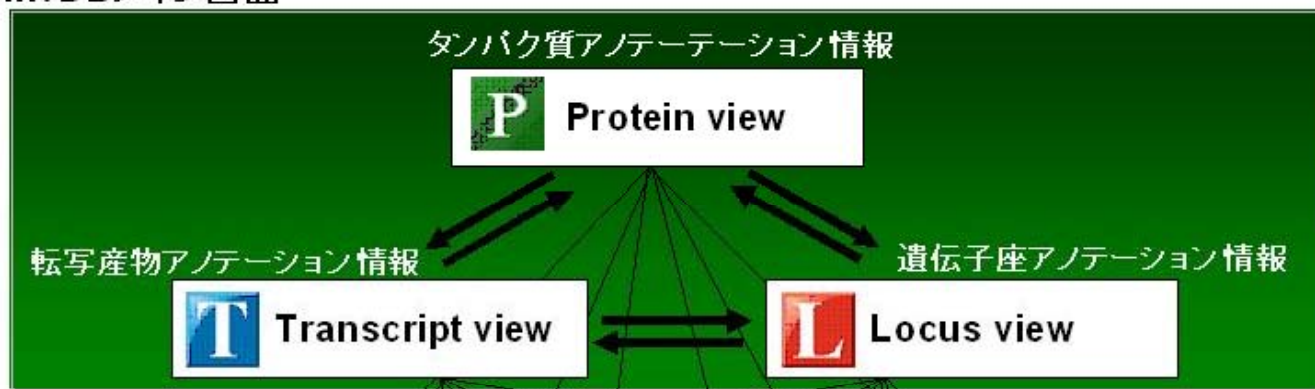
G-compass

ヒト-他生物間における
オーソログ領域の
ゲノムアラインメント

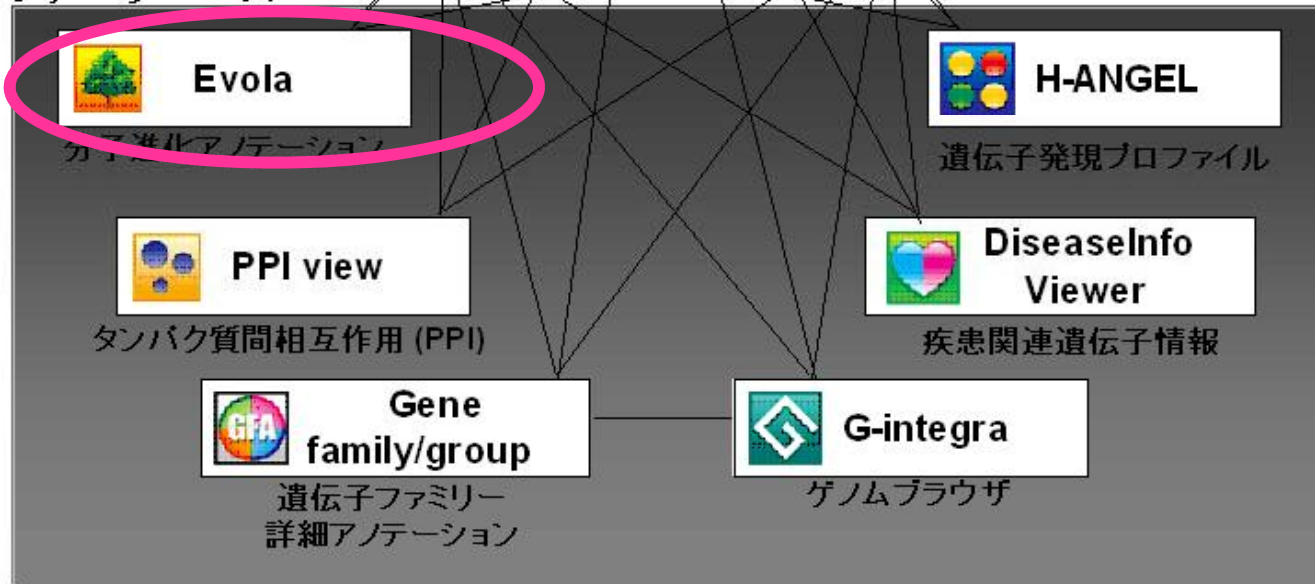


H-InvDBにおけるEvola, G-compassの関係

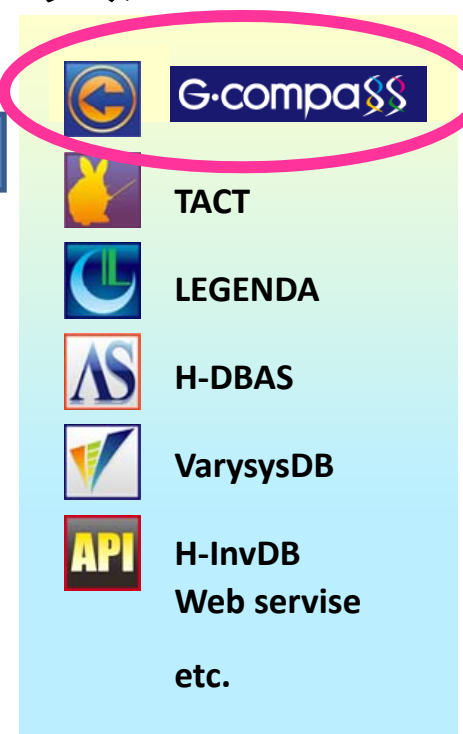
H-InvDBメイン画面



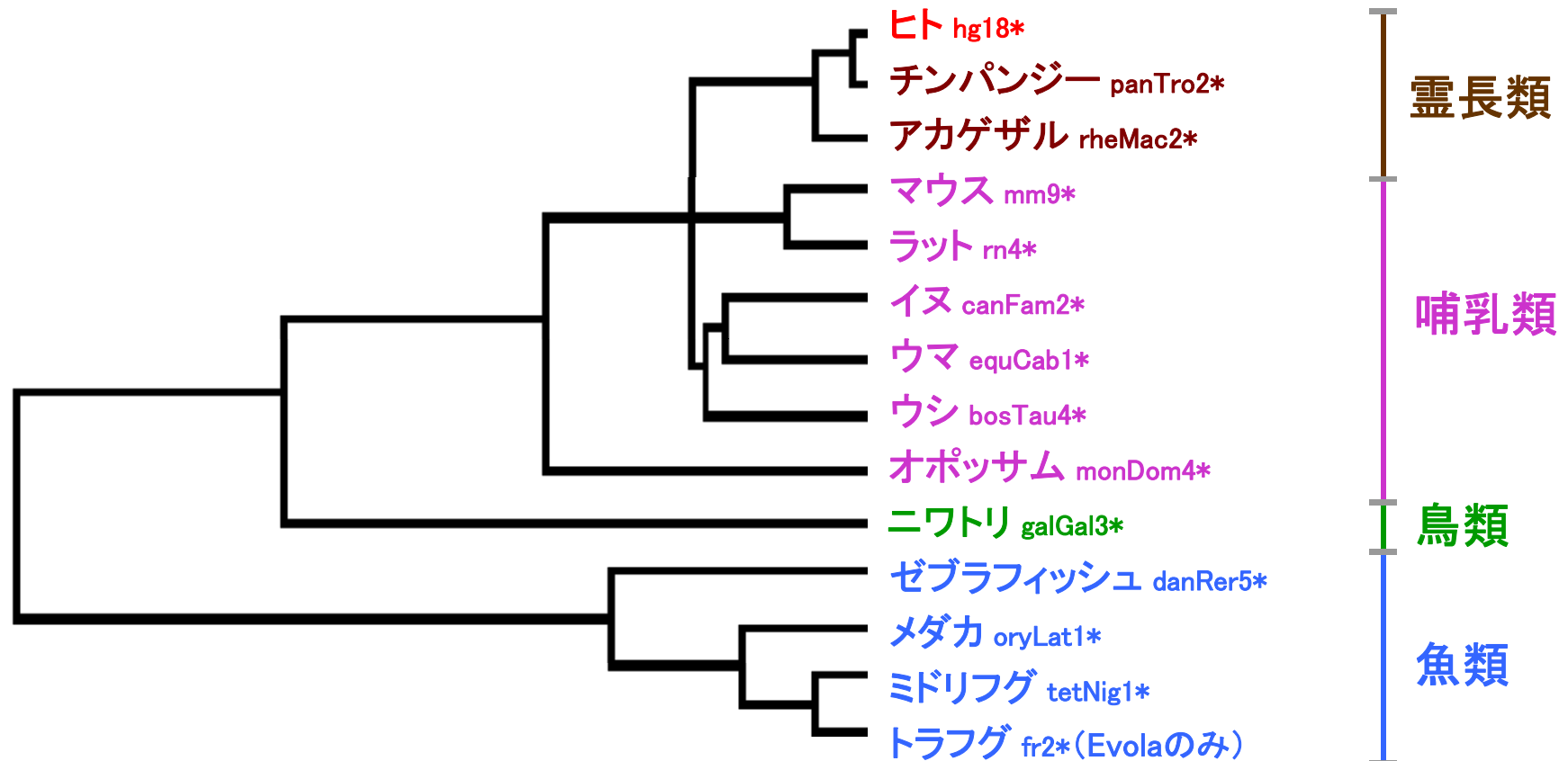
サブデータベース



サテライトデータベース・ツール



Evola, G-compassに用いている生物種



* UCSCゲノム <http://genome.ucsc.edu/>

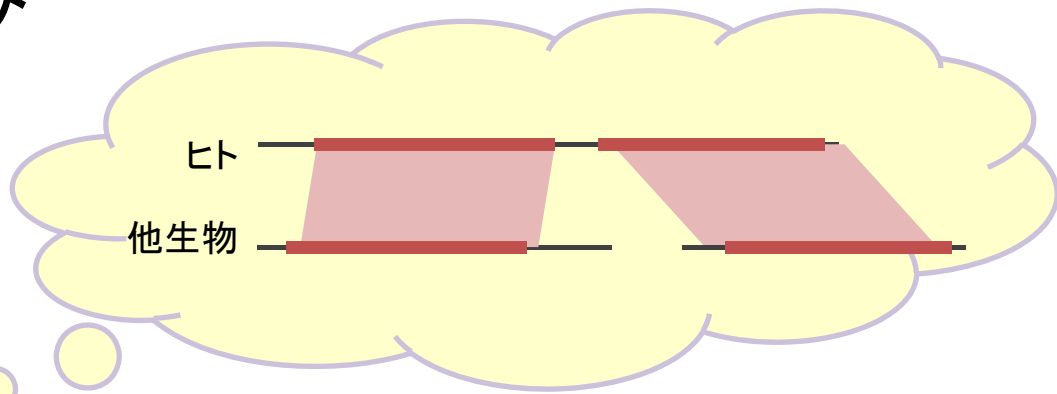
ヒトゲノムおよび13種の脊椎動物ゲノムを使用

データ作成の枠組み

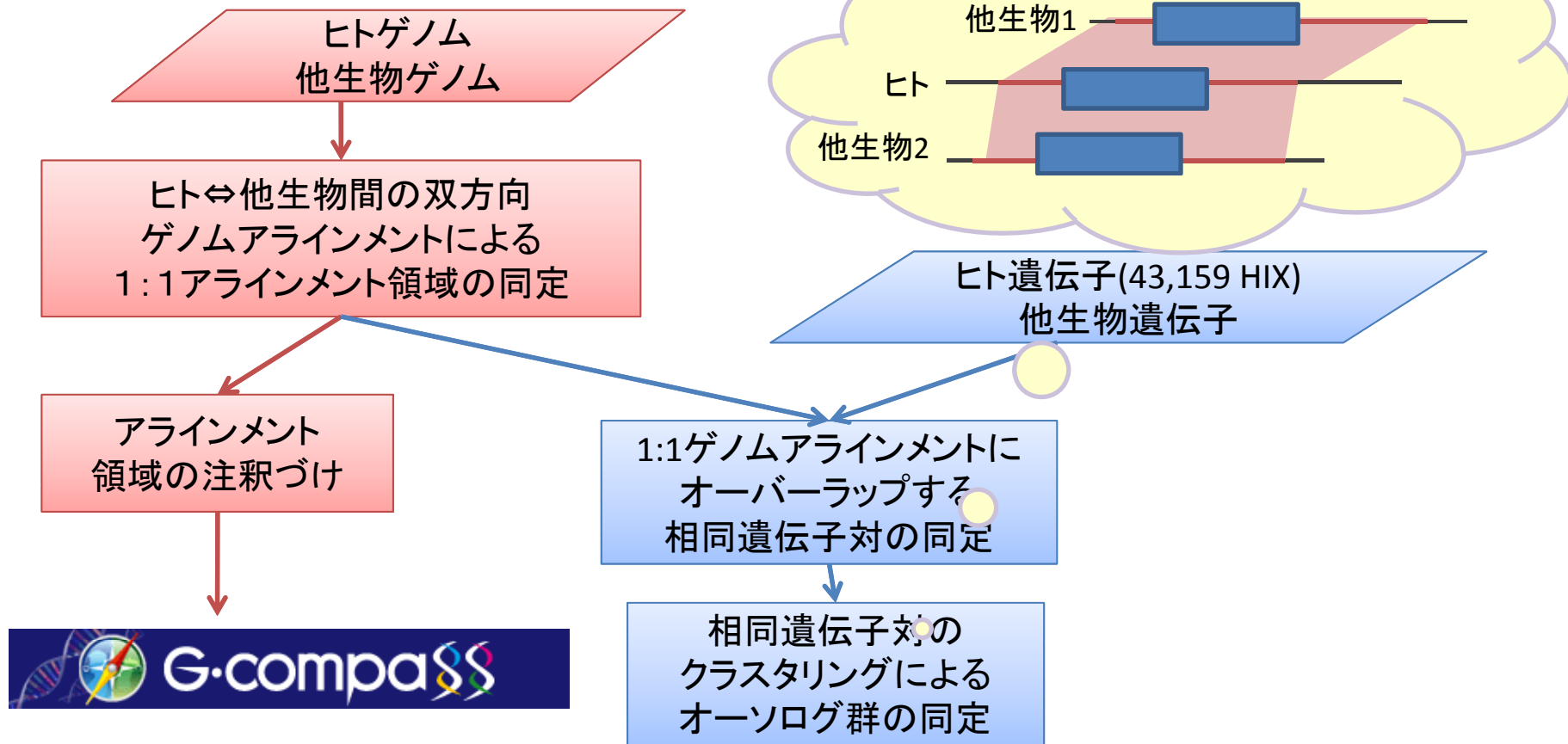
ヒトゲノム
他生物ゲノム

ヒト⇔他生物間の双方向
ゲノムアラインメントによる
1:1アラインメント領域の同定

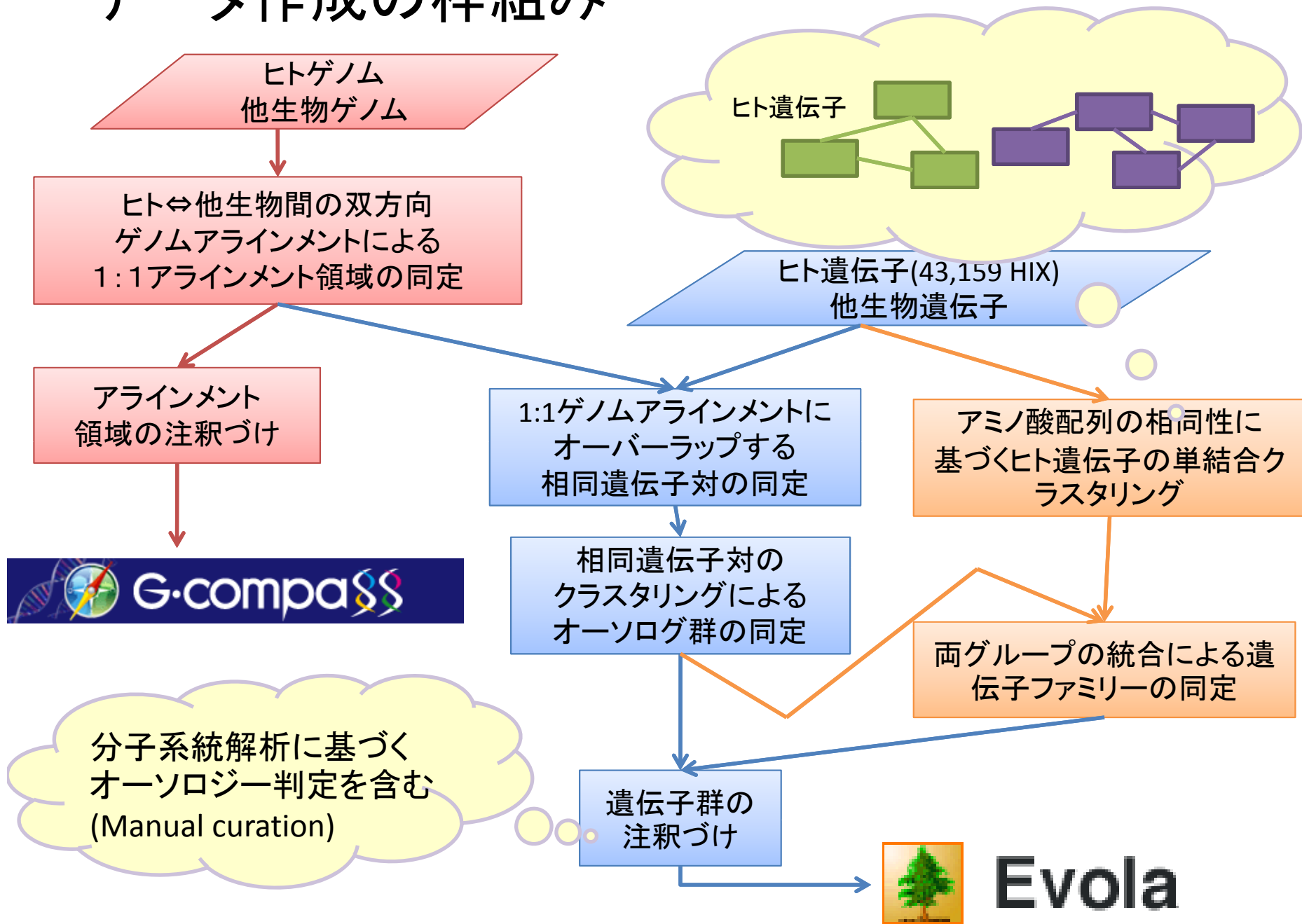
アラインメント
領域の注釈づけ



データ作成の枠組み



データ作成の枠組み

















G-compass 統計データ(抜粋)

生物種のペア (ヒト 対 他生物種)	ヒトゲノム上の アライン メント長 (Gb)	アライン メント ブロック数	アライン メント長の 中央値 (bp)	全長がアライ メントに含 まれる遺伝子 数	配列の相違度	
					ギャップ無	ギャップ有
ヒト-チンパンジー	2.36	188,292	6,546	20,196	0.0144	0.0213
ヒト-マカクザル	2.22	420,684	3,473	16,278	0.0631	0.0864
ヒト-マウス	1.06	828,478	776	9,749	0.3158	0.4228
ヒト-イヌ	1.63	760,305	1,317	12,604	0.2512	0.3714
ヒトゲノム	3.11	-	-	35,303*	-	-

*based on H-InvDB 6.2

Evola 統計データ

Evola 6.0 統計情報								
Species	Genes*	Human genes*	Species	Genes*	Human genes*	Species	Genes*	Human genes*
 Human hg18	22,496	-	 Dog canFam2	16,588	17,006	 Zebrafish danRer5	14,165	13,262
 Chimpanzee panTro2	19,442	19,663	 Horse equCab1	15,998	16,359	 Medaka oryLat1	12,683	12,359
 Macaque rheMac2	17,507	17,739	 Cow bosTau4	15,348	15,642	 Tetraodon tetNig1	11,807	11,575
 Mouse mm9	17,540	17,831	 Opossum monDom4	12,083	12,342	 Fugu fr2	12,710	12,455
 Rat m4	16,110	16,453	 Chicken galGal3	11,567	12,112			

*ヒトと他生物種のオーソログの関係は、「1遺伝子対1遺伝子」ではなく、「n対m」(n≥1, m≥1)として定義されるため、「Human genes」と「(他生物)Genes」の数字は一致しない

Evola, G-compassをどのように研究に活かすか

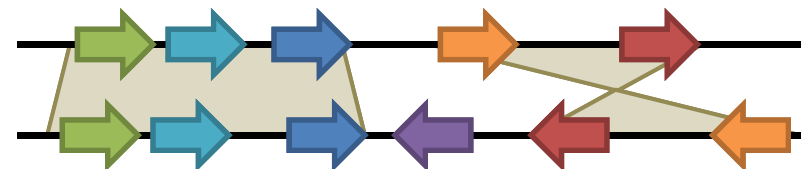
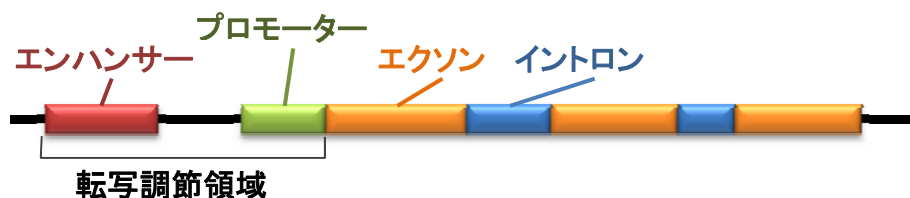
対象とする遺伝子、ゲノム領域を種間(ヒト \leftrightarrow モデル生物)で比較する

Evola

- オースログは存在するか
- パラログはいくつあるか、いつ重複したか
- 相同遺伝子間で、高度に保存された or 種特異的に多様化した領域はあるか

G-copmass

- 相同遺伝子の非コード領域や近傍領域(e.g. 転写調節領域)の配列も保存されているか
- 対象とする保存領域に特徴的なゲノム構造は存在するか
- 遺伝子の並び(シンテニー)も種間で保存されているか



Evola, G-compassを使ってみましょう

アクセス方法

1. H-InvDBのメインページ(http://h-invitational.jp/hinv/ahg-db/tools_ja.jsp)→クイックガイド

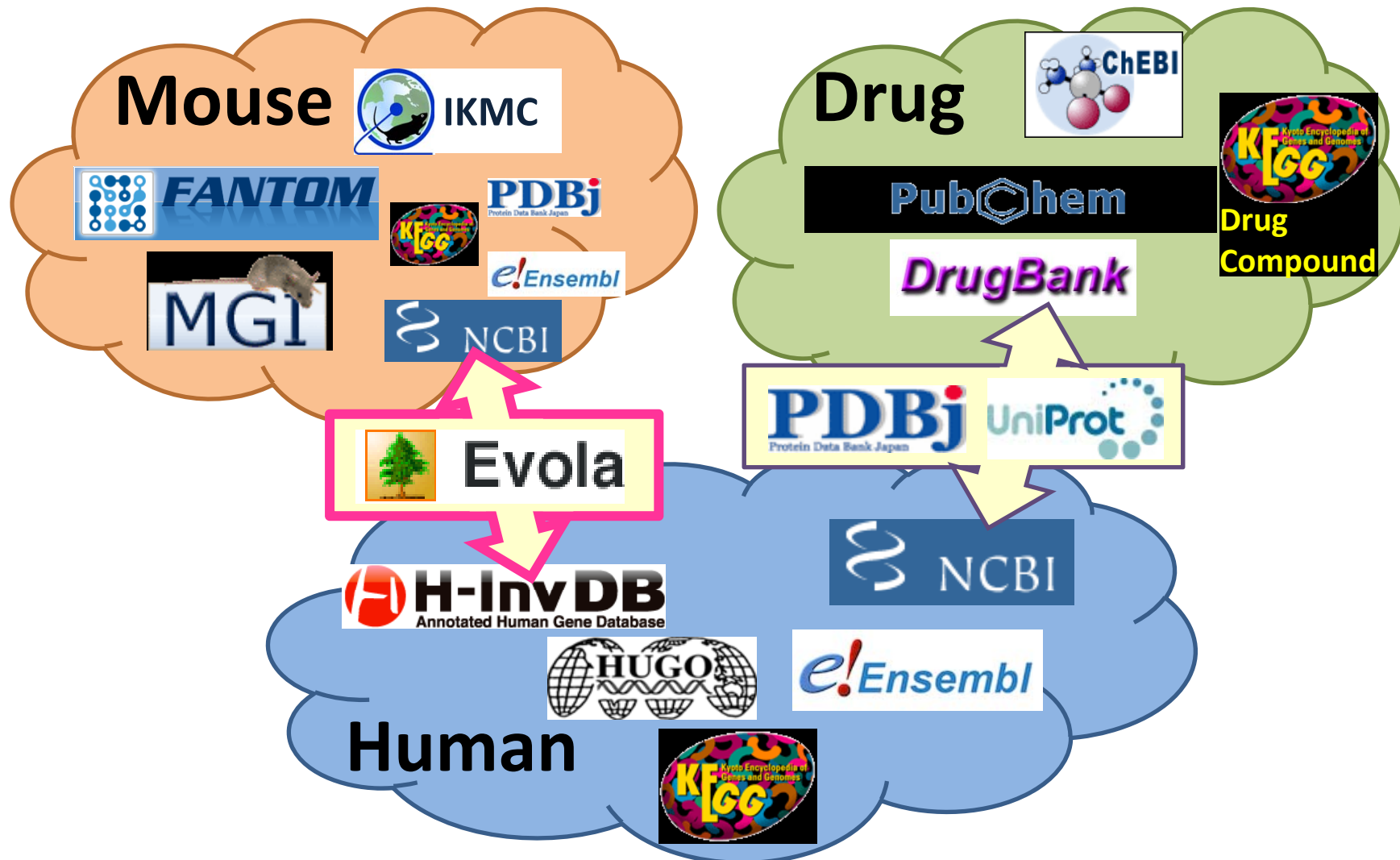


2.

evola	検索
-------	----

g-compass	検索
-----------	----

Evolaデータの活用例: リンク自動管理システムによる種を超えたID変換



Evola 7.5

10月上旬に公開予定!!

H-InvDB 7.5に準拠

参考文献

Evola: Ortholog database of all human genes in H-InvDB with manual curation of phylogenetic trees.

Matsuya A, Sakate R, Kawahara Y, *et al.*

Nucleic Acids Res. D787-792 (2008)

A web tool for comparative genomics: G-compass.

Fujii Y, Itoh T, Sakate R, *et al.*

Gene 364, 45-52 (2005)

G-compass: A web-based comparative genome browser between human and other vertebrate genomes.

Kawahara Y, Sakate R, Matsuya A, *et al.*

Bioinformatics 25, 3321-2 (2009)